Lecture Notes in Networks and Systems 311

G. Ranganathan
Xavier Fernando
Fuqian Shi *Editors*

Inventive Communication and Computational Technologies

Proceedings of ICICCT 2021

Springer

Lecture Notes in Networks and Systems

Volume 311

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at https://link.springer.com/bookseries/15179

G. Ranganathan · Xavier Fernando · Fuqian Shi
Editors

Inventive Communication and Computational Technologies

Proceedings of ICICCT 2021

Springer

Editors
G. Ranganathan
Department of Electronics
and Communication Engineering
Gnanamani College of Technology
Namakkal, Tamil Nadu, India

Fuqian Shi
Rutgers University
New Brunswick, NJ, USA

Xavier Fernando
Department of Electrical and Computer
Engineering
Ryerson Communications Lab
Toronto, ON, Canada

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-16-5528-9 ISBN 978-981-16-5529-6 (eBook)
https://doi.org/10.1007/978-981-16-5529-6

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

With a great deal of dedication, 5th ICICCT 2021 is dedicated to all the authors, great scientists, academicians, young researchers, conference delegates and students, who have research expertise on ICTs from all over the world. Nevertheless, this proceedings is dedicated to cover a wide spectrum of themes related to intelligent computing and communication innovations and developments.

Preface

The 2021 International Conference on Inventive Communication and Computational Technologies (ICICCT 2021) was held at Gnanamani College of Technology, Namakkal, India, during June 25–26, 2021. ICICCT 2021 aims to cover the recent advancement and trends in the area of communication and computational technologies to facilitate knowledge sharing and networking interactions on emerging trends and new challenges.

ICICCT 2021 tends to collect the latest research results and applications on data communication and computer networking, software engineering, wireless communication, VLSI design and automation, networking, Internet of things, cloud and big data. It includes a selection of 74 papers from 328 papers submitted to the conference from universities and industries all over the world. All the accepted papers were subjected to strict peer-reviewing by 2–4 expert referees. The papers have been selected for this volume because of quality and the relevance to the conference.

We would like to thank the guest editors Dr. Xavier Fernando, Professor and Director, Ryerson Communications Lab, Department of Electrical, Computer and Biomedical Engineering, and Dr. Fuqian Shi, Professor, Rutgers University, New Jersey, USA, for their valuable guidance and technical support for the article selection process.

ICICCT 2021 would like to express our sincere appreciation to all the authors for their contributions to this book. We would like to extend our thanks to all the referees for their constructive comments on all papers and our keynote speaker Dr. Abul Bashar, College of Computer Engineering and Sciences, Prince Mohammad Bin Fahd University, Al-Khobar, Kingdom of Saudi Arabia; especially, we would like to thank the organizing committee for their hard work. Finally, we would like to thank the Springer publications for producing this volume.

Namakkal, India — Dr. G. Ranganathan
Toronto, Canada — Xavier Fernando
New Brunswick, USA — Fuqian Shi

Acknowledgements ICICCT 2021 would like to acknowledge the contribution made to the organization by its many volunteers. Members contribute their time, energy and knowledge at a local, regional and international level.

We also thank all the chair persons and conference committee members for their support.

Contents

Editors and Contributors

About the Editors

Dr. G. Ranganathan working as a professor and the head in Gnanamani College of Engineering and Technology, Namakal, India. He has done his Ph.D. in the Faculty of Information and Communication Engineering from Anna University, Chennai in the year 2013. His research thesis was in the area of bio-medical signal processing. He has total of 29+ years of experience in industry, teaching, and research. He has guided several project works for many UG and PG students in the areas of bio-medical signal processing. He has published more than 35 research papers in International and National Journals and Conferences. He has also co-authored many books in electrical and electronics subjects. He has served as a referee for many reputed international journals published by Elsevier, Springer, Taylor and Francis, etc. He has membership in various professional bodies like ISTE, IAENG etc. and has actively involved himself in organizing various international and national level conferences, symposiums, seminars, etc.

Dr. Xavier Fernando is the director of Ryerson Communications Lab that has received total research funding of $3,185,000.00 since 2008 from industry and government (including joint grants). He has (co-)authored over 200 research articles, three books (one translated to Mandarin), and several book chapters and holds few patents. The present and past members and affiliates of this lab can be found at this LinkedIn group and Facebook page. He was an IEEE Communications Society Distinguished Lecturer and has delivered over 50 invited talks and keynote presentations all over the world. He was a member in the IEEE Communications Society (COMSOC) Education Board Working Group on Wireless Communications. He was the chair of IEEE Canada Humanitarian Initiatives Committee 2017–2018. He was also the chair of the IEEE Toronto Section and IEEE Canada Central Area.

Dr. Fuqian Shi was graduated from College of Computer Science and Technology, Zhejiang University, and got his Ph.D. on Engineering and was a visiting associate professor at the Department of Industrial Engineering and Management System, University of Central Florida, USA, from 2012 to 2014. Dr. Shi is currently serving for Rutgers Cancer Institute of New Jersey, USA. Dr. Shi is a senior member of IEEE, Membership of ACM, and served as over 30 committee board membership of international conferences. Dr. Shi also served as an associate editors of *International Journal of Ambient Computing and Intelligence* (IJACI), *International Journal of Rough Sets and Data Analysis* (IJRSDA), and special issue editors of *IEEE Access*. He published over 100 journal papers and conference proceedings; his research interests include fuzzy inference system, artificial neuro-networks, and bio-mechanical engineering.

Contributors

D. Abhiram TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

CH.N.S. Abhishek TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

M. Aditya Reddy Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

V. Agalya Department of EEE, CMR Institute of Technology, Bengaluru, India

Vanita Agarwal Department of Electronics and Telecommunication, Cusrow Wadia Institute of Technology, Pune, India;
Department of Electronics and Telecommunication, College of Engineering, Pune, India

I. S. Akila Department of ECE, Coimbatore Institute of Technology, Coimbatore, India

B. Aksa QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

Fat'hi Salim Said AL-Ghafri Middle East College, Muscat, Sultanate of Oman

Suhad A. Ali Department of Computer Science, College of Science for Women, Babylon University, Hillah, Iraq

Fahiem Altaf Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, India

K. Ambujam AKT Memorial College of Engieenring and Technology, Kallakurichi, Tamilnadu, India

P. P. Amritha TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

K. Amrithesh Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

P. Mohan Anand Department of Computer science and Engineering, Indian Institute of Technology Kanpur, Kanpur, India

R. Anand Department of Computer Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka, India

J. Anand Babu Department of Information Science and Engineering, Malnad College of Engineering and Technology, Hassan, India

A. Peter Soosai Anandaraj Computer Science and Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

T. S. Angel Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

G. N. Anil Department of Computer Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka, India

J. P. Anita Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

M. Anupama Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

K. Anuraj Department of Electronics and Communication Engineering, Amrita Viswa Vidyapeetham, Amritapuri, India

K. Archana Department of EEE, Cambridge Institute of Technology, Bengaluru, India

R. Arjun Sri Ramakrishna Engineering College, Coimbatore, India

Sachin Ashok Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

Manish Assudani G. H. Raisoni University, Amravati, Maharashtra, India

J. Aswin Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

B. Sekhar Babu Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

M. Balaji Bharatwaj Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

N. J. Basil Department of Electronics and Communication Engineering, Amrita Viswa Vidyapeetham, Amritapuri, India

Shikhar Bhardwaj Punjab Engineering College (Deemed to be University), Chandigarh, India

Vasujit Bhattacharjee Computer Science and Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

P. Bhuvaneshwari MVJ College of Engineering, Bangalore, India

Ashok Chandak Department of Electronics and Telecommunication, Cusrow Wadia Institute of Technology, Pune, India

Shalini Chandra Department of Computer Science, BBA University, (A Central University), Lucknow, India

G. Chandra Sekhar Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

Hari Akhilesh Chandrasekar SRM Institute of Science and Technology, SRM Nagar Kattankulathur, Kanchipuram, Chennai, TN, India

Sunil Chaudhari Fr.Conceicao Rodrigues College of Engineering, Mumbai, India

Ajay Chunduri Microchip Technology Pvt. Ltd., Hyderabad, India

Jayshil Dave Computer Science and Engineering, SRM Institute of Science and Technology Chennai, Chennai, India

Shiela David SRM Institute of Science and Technology, Ramapuram, Chennai, TamilNadu, India

V. Deepa Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

M. V. Deepak Chand Reddy Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Kalpana Deorukhkar Father Conceicao Rodrigues College of Engineering, Mumbai, Maharashtra, India

C. Devika Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

N. M. Dhanya Amrita Vishwa Vidyapeetham, Amritanagar, Ettimadai, Tamil Nadu, India

S. Nithya Dharshni Vel tech High tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India

S. Dilipkumar Department of CSE, Bharathidasan University, Trichy, India

P. Dinesh Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

M. Divya Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

M. Durairaj Department of CSE, Bharathidasan University, Trichy, India

C. Elakkiya Department of ECE, Coimbatore Institute of Technology, Coimbatore, India

H. Abdul Gaffar Vellore Institute of Technology, Vellore, India

G. Gajenthiran Sri Ramakrishna Engineering College, Coimbatore, India

A. Gandhimathinathan Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Sanjay Ganorkar Department Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Pune, India

M. Gayathri Computer Science and Engineering, SRM Institute of Science and Technology Chennai, SRM Nagar Kattankulathur, Kanchipuram, Chennai, Tamil Nadu, India

V. Geetha Department of Information Technology, Pondicherry Engineering College, Puducherry, India

R. Gogul Sriman Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

Subbarao Gogulamudi Department of CSE, Annamalai University, Cuddalore, Tamilnadu, India

A. Gokul Department of Electronics and Communication Engineering, Amrita Vishwa Vidhyapeetham, Amritapuri, India

Vellaturi Gopikrishna Department of Computer Science and Engineering, PACE Institute of Technology and Sciences, Ongole, Andrapradesh, India

Riya Gupta Fr.Conceicao Rodrigues College of Engineering, Mumbai, India

Sudheer Kumar Gupta Computer Science and Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

Sheimaa A. Hadi Department of Computer Science, College of Science for Women, Babylon University, Hillah, Iraq

Ch. Harika QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

C. Harikrishnan Amrita Vishwa Vidyapeetham, Amritanagar, Ettimadai, Tamil Nadu, India

R. Harish TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Sandeep Harit Punjab Engineering College (Deemed to be University), Chandigarh, India

Konda Harsha Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

T. Hemanth Kumar Reddy Computer Science Engineering, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamilnadu, India

P. Hima Varshini Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

V. Hrishitha Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

S. Hrushikesava Raju Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

M. Ilango Hindusthan College of Arts and Science, Coimbatore, India

Babu Illuri Department of ECE, Vardhaman College of Engineering, Hyderabad, India

B. Indra Kiran Reddy Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Aminul Islam Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, India

Md. Mahfujul Islam Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh

Muhmmad Nazrul Islam Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh

Tejas Jambhale Vellore Institute of Technology, Vellore, India

Majid Jabbar Jawad Department of Computer Science, College of Science for Women, Babylon University, Hillah, Iraq

M. Jayakumar Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

M. Jogendra Kumar Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Deepa Jose KCG College of Technology, Chennai, TamilNadu, India

Lakshmy K.V. TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Kevlyn Kadamala Father Conceicao Rodrigues College of Engineering, Mumbai, Maharashtra, India

Gadde Karthikeya Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India

Jeevaa Katiravan Department of Computer Science and Engineering, Velammal Engineering College, Chennai, India

Manjot Kaur Department of Computer Science, Punjabi University, Patiala, India

M. Kavitha Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur, India

S. Kavitha Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur, India

T. G. Keerthan Kumar Department of Information Science and Engineering, Siddaganaga Institute of Technology, Tumakuru, India

M. Keerthana Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

B. Keerthi QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

P. Keerthi QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

Shubhangee Kishan Varma Department of Electronics and Telecommunication, Cusrow Wadia Institute of Technology, Pune, India

Kottilingam Kottursamy School of Computing, SRM Institute of Science and Technology, Chennai, Tamilnadu, India

Ashok Koujalagi Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

G. Kranthi Kumar Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

Karthik Krishna Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

G. Krishna Kishore Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

Binod Kumar Department of MCA, Rajarshi Shahu College of Engineering, Pimpri-Chinchwad, India

Muddamsetty Tanuj Kumar Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India

Lubna Mohammed Kunhi Department of CSE, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Karkala, India

M. Lakshmi Akhila Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

R. Lavanya Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Mulaka Madhava Reddy Department of Information Technology, PACE Institute of Technology and Sciences, Ongole, Andrapradesh, India

V. Madhava Reddy QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

T. Madhu Babu Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

V. Mahalakshmi Department of CSE, Annamalai University, Cuddalore, Tamilnadu, India

Rajnish Mahaseth MVJ College of Engineering, Bangalore, India

P. Mahesh QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

Soumyadev Maity Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, India

C. Malathy SRM Institute of Science and Technology, SRM Nagar Kattankulathur, Kanchipuram, Chennai, Tamil Nadu, India

Pattisapu Manikanta Manohar Department of CSE, Raghu Engineering College (A), Visakhapatnam, Andhra Pradesh, India

Maya Manish Kumar Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

T. N. Manjunath Department of Information Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka, India

Najmuddin M. Maroof Department of Electronics and Communication Engineering, Khaja BandaNawaz College of Engineering, Gulbarga, Karnataka, India

Senthilkumar Mathi Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Prabhakaran Mathialagan SRM Institute of Science and Technology, SRM Nagar Kattankulathur, Kanchipuram, Chennai, Tamil Nadu, India

Elita Menezes Father Conceicao Rodrigues College of Engineering, Mumbai, Maharashtra, India

Chindamani Meyyappan Sri Ramakrishna Engineering College, Coimbatore, India

M. Mohana Priya Department of ECE, Coimbatore Institute of Technology, Coimbatore, India

M. Mohit Department of Electronics and Communication Engineering, Amrita Vishwa Vidhyapeetham, Amritapuri, India

Akshada Muneshwar National Institute of Technology, Warangal, Telangana, India

M. S. Muneshwara Department of Computer Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka, India

Dharmalingam Muthusamy Government Arts and Science College—Modakkurichi, Erode, Tamil Nadu, India

M. Naga Sivaram Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

M. Nagarjuan Department of ECE, Vardhaman College of Engineering, Hyderabad, India

Aishvarya Nair TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Amrutha Muralidharan Nair Department of CSE, Karpagam Academy of Higher Education, Coimbatore, India

Aswathy K. Nair Department of Electronics and Communication Engineering, Amrita Vishwa Vidhyapeetham, Amritapuri, India

Neethu B. Nair Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

T. Namratha Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

P. Narendran Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

Faizan Nasir Department of Computer Science, Aligarh Muslim University, Aligarh, India

R. Neeraj Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

M. Niranjan Department of Electronics and Communication Engineering, Amrita Vishwa Vidhyapeetham, Amritapuri, India

V. Nirmala Information and Communication Technology, School of Computing SASTRA, Deemed to be University, Thanjavur, Tamil Nadu, India

B. Nithya Department of Computer Science, VISTAS, Chennai, Tamil Nadu, India

B. Nivedha Department of ECE, Coimbatore Institute of Technology, Coimbatore, India

D. Nivedha Vel tech High tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India

M. Nivedha Vel tech High tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India

Dishank Oza Fr.Conceicao Rodrigues College of Engineering, Mumbai, India

A. Pavan Kumar Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

V. Pavan Kumar Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Sindhe Phani Kumar Department of Computer Science and Engineering, PACE Institute of Technology and Sciences, Ongole, Andrapradesh, India

S. S. Poorna Department of Electronics and Communication Engineering, Amrita Viswa Vidyapeetham, Amritapuri, India

U. Prabu Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India

V. R. Prakash Department of Electronics and Communication Engineering, Hindustan Institute of Technology & Science, Chennai, Tamilnadu, India

Bhaskaruni Sai Praneetha Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

K. Praveen TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

G. D. Praveenkumar Government Arts and Science College—Modakkurichi, Erode, Tamil Nadu, India

J. Premaladha Information and Communication Technology, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamilnadu, India

G. Prethija Department of Information Technology, Velammal Engineering College, Chennai, India

Radhika Priyavardhini Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

Nawsheen Tarannum Promy Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh

S. K. Pushpa Department of Information Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka, India

B. M. Ragavendra Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

N. Raghavendra Sai Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Akshita Rai Computer Science and Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

Sanju Rajan School of Computing Sciences, Hindustan Institute of Technology & Science, Chennai, Tamilnadu, India

C. B. Rajesh Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

V. Rajmohan Department of Electronics and Communication Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, India

R. Ramaguru TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

K. Ramesh Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India

S. R. Ramesh Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Santhosh Veeraraghavan Ramesh Electronics and Communication Engineering, Sri Sairam Engineering College, Chennai, Tamilnadu, India

G. R. Ramya Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Rajneesh Randhawa Department of Computer Science, Punjabi University, Patiala, India

R. Ranjitha Department of EEE, CMR Institute of Technology, Bengaluru, India

Tirandasu Ravi Kumar Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

V. Ravi Kumar Pandi Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

Proteeti Prova Rawshan Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh

Avuthu Yuvaraja Reddy Department of Electronics and Communication Engineering, Amrita Viswa Vidyapeetham, Amritapuri, Coimbatore, India

Kallam Praneeth Sai Kumar Reddy Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India

Vyshnavi Reddy Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

P. N. Renjith Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India

Y. Risheet Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

Md. Roshan Tanveer Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

J. RubyDinakar VTU Research Scholar, Faculty of Computer Science and Engineering, PES University, Bangalore, India

Banupriya Sadayapillai School of Computing, SRM Institute of Science and Technology, Chennai, Tamilnadu, India

Neeraj Sahu G. H. Raisoni University, Amravati, Maharashtra, India

P. Sai Kiran Reddy Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

G. Sai Pravallika Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

H. Saiganesh Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

K. S. Saileswar Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

Sreelekshmi Saju Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

Abdus Samad Department of Computer Engineering, Aligarh Muslim University, Aligarh, India

S. Sandeep Kumar Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

R. Santhanakrishnan Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

M. V. B. T. Santhi Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

R. Santhosh Department of CSE, Karpagam Academy of Higher Education, Coimbatore, India

J. N. Sarath Department of Electronics and Communication Engineering, Amrita Vishwa Vidhyapeetham, Amritapuri, India

T. V. Sarath Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

V. Sarma TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

K. Kottayil Sasi Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

Siva Satya Sri Ganesh Seeram Department of Electronics and Communication Engineering, Amrita Viswa Vidyapeetham, Amritapuri, India

P. Seetha Rama Krishna Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

A. V. Senthil Kumar Department of MCA, Hindusthan College of Arts and Science, Coimbatore, India

Thangavel Senthil Kumar Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Yegappan Sethu Sri Ramakrishna Engineering College, Coimbatore, India

M Sethumadhavan TIFAC-CORE in Cyber Security, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

R. Sharan Prasanna Sri Ramakrishna Engineering College, Coimbatore, India

S. Sharanyaa Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India

Vivek Sharma Department of Information Science and Engineering, Nagarjuna College of Engineering and Technology, Bengaluru, India

Jyothi Shetty Department of CSE, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Karkala, India

T. Shivakumara Department of Master of Computer Application, BMS Institute of Technology and Management, Bengaluru, Karnataka, India

Nitin Shivsharan Computer Engineering, SSPM's College of Engineering Kankavali, Sindhudurga, India

M. Shyamala Devi Computer Science and Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

Jamshed Siddiqui Department of Computer Science, Aligarh Muslim University, Aligarh, India

P. Sivraj Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

Indraneel Sreeram Department of CSE, St. Ann's College of Engineering and Technology, Chirala, Andhra Pradesh, India

Y. Sri Sai Charan Computer Science Engineering, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamilnadu, India

K. Sri Thanvi Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

P. V. Sridevi AUCE(A), Andhra University, Visakhapatnam, India

M. S. Srikanth Department of Information Science and Engineering, Nagarjuna College of Engineering and Technology, Bengaluru, India

Chungath Srinivasan TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

B. K. N. Srinivasarao National Institute of Technology, Warangal, Telangana, India

Singaraju Srinivasulu Department of Information Technology, PACE Institute of Technology and Sciences, Ongole, Andrapradesh, India

P. Sripriya Department of Computer Application, VISTAS, Chennai, Tamil Nadu, India

K. Stella Vel tech High tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India

G. Subba Rao Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

P. Sudheesh Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Akella Viswa Sai Suman Department of Electronics and Communication Engineering, Amrita Viswa Vidyapeetham, Amritapuri, India

Regidi Suneetha AUCE(A), Andhra University, Visakhapatnam, India

M. Surendra Reddy Computer Science Engineering, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamilnadu, India

K. Sureshkumar Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

L. U. Suriya Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

J. N. Swaminathan QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

M. S. Swetha Department of Computer Science and Engineering, Department of Information Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka, India

Aju Mathew Thomas TIFAC-CORE in Cyber Security, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

Sudi Sai Thrilok Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India

Praveen Tumuluru Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

N. Umeshraja Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

S. Vagdevi Professor, Department of Computer Science and Engineering, City Engineering College, Bangalore, India

Sulakshan Vajipayajula Chief Architect - Security Analysis, STFM, IBM security, Bangalore, India

Dusi Venkata Divakara Rao Department of CSE, Raghu Engineering College (A), Visakhapatnam, Andhra Pradesh, India

Richa Verma Department of Computer Science, BBA University, (A Central University), Lucknow, India

Lavanya Vidhya Department of E.C.E, Middle East College, Muscat, Sultanate of Oman

S. S. Vidhya Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

M. Vignesh Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

G. Vijaya Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

J. Vishnuprakash Sri Ramakrishna Engineering College, Coimbatore, India

T. Viswak Sena Sri Ramakrishna Engineering College, Coimbatore, India

Mohammed Abdul Waheed Department of Computer Science Engineering, V.T.U.P.G. Centre, Kalaburagi, Karnataka, India

S. Yadhaarshini Department of ECE, Coimbatore Institute of Technology, Coimbatore, India

Dimensionality Reduction Based Component Discriminant Factor Implication for Mushroom Edibility Classification Using Machine Learning

M. Shyamala Devi, A. Peter Soosai Anandaraj, Sudheer Kumar Gupta, Vasujit Bhattacharjee, Akshita Rai, and Santhosh Veeraraghavan Ramesh

Abstract Despite the fact that technology is progressing, people continue to consume deadly wild mushrooms due to their inability in classifying the different mushroom categories. Currently, no specific traits are defined to accurately predict the edibility of mushrooms based on their attributes. To overcome this challenge, Machine Learning [ML] can be used for identifying the poisonous mushroom based on the feature appearance. By considering the above, the Mushroom dataset extracted from UCI data warehouse are used for predicting the mushroom edibility level. The division of mushroom edible classes are achieved in four different ways. Firstly, the dataset is preprocessed with feature scaling and missing values. Secondly, raw data set is fitted to all the classifier with and without the presence of feature scaling. Thirdly, raw data is applied with the principal component analysis with 8, 10 and 12 components and PCA reduced dataset is fitted to all the classifier with and without the presence of feature scaling. Fourth, the raw data is applied with the Linear discriminant analysis and LDA reduced dataset is fitted to all the classifier with and without the presence of feature scaling. Fifth, raw data is applied with the Factor analysis with 8, 10 and 12 components and FA reduced dataset is fitted to all the classifier with and without the presence of feature scaling. Sixth, the performance of raw data set, PCA reduced data set, LDA reduced dataset and FA reduced dataset are compared by analyzing the performance metrics like Precision, Recall, Accuracy

M. Shyamala Devi (✉) · A. P. S. Anandaraj · S. K. Gupta · V. Bhattacharjee · A. Rai
Computer Science and Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India
e-mail: shyamaladevim@veltech.edu.in

S. K. Gupta
e-mail: vtu10449@veltechuniv.edu.in

V. Bhattacharjee
e-mail: vtu10453@veltechuniv.edu.in

A. Rai
e-mail: vtu10335@veltechuniv.edu.in

S. V. Ramesh
Electronics and Communication Engineering, Sri Sairam Engineering College, Chennai, Tamilnadu, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_1

and F-score. The implementation is done by using python language under Spyder platform with Anaconda Navigator. Experimental results show that, the kernel SVM, KNN and Adaboost classifier for the FA reduced dataset tends to retain the accuracy with 100% before and after feature scaling. The KNN classifier with LDA reduced dataset tends to retain the 96% accuracy before and after feature scaling. Kernel SVM, KNN and Adaboost classifier for the FA reduced dataset tends to retain the accuracy with 99% before and after feature scaling. From the above analysis, KNN classifier ismore efficient based on its accuracy with all PCA, LDA and FA reduced dataset.

Keywords Machine learning · Classification · Precision · Accuracy · PCA · LDA and factor analysis

1 Introduction

Data mining method is used for performing an empirical study on mushroom diseases. The mushroom recognition with machine learning algorithms such as classification can be done for predicting the edibility of mushrooms. After implementing various learning classification models on our dataset to evaluate mushroom poisoning, it has been observed that while all the learning models are relatively good, some exceptionally good, the consistent and effective default KNN wins out. It is already known that, the KDD is a procedure for data mining that portrays step-by-step processes for extracting the data models that are used for prediction and cluster analysis.

2 Literature Review

This paper develops a model for mushroom analysis using data mining. Here, the algorithms are directly applied to the dataset. WEKA implements the data preprocessing algorithm, category cluster and rules of association. The results of the experiment show the benefit of background images, especially when the KNN algorithm is used. The results of the experiment show an advantage for background images [1]. In this project, the authors have performed an observational review on the different mushroom diseases. Here, the input data is translated into a suitable file format. The file created is then fed into the Evaluator. After that, the highly influential factors are recognized. At last, classification algorithms are applied and the performance has also been compared [2]. In this paper, a methodology has been developed for edible mushroom identification with the help of data mining. At first, the mushroom dataset is divided and then the K-fold test dataset is used for testing purposes. From the recultsm it has been observed that, the dataset is evenly distributed as 51.8% edible and 48.2% poisonous [3].

The first step is image segmentation. The first extracted feature after segmentation is considered as the colour of the cap. Then, the ratio of the image is obtained from the segmented image after extracting the colour of the cap. After that the segmented image is processed again for its white blob detection. As a result, the application has obtained an accuracy result of 92% after performing a series of tests. wherein the 8%error rate is due to the image quality. Hence, the authors successfully developed a neural network model to recongize the mushroom [4]. In this paper, the performance of Mushroom dataset is analysed with the usage of decision tree algorithms. The proposed method classifies the edible mushrooms and poisonous mushrooms from the risk factor detection dataset of mushrooms. It then compares the output of three different decision tree classification techniques. In the future, the Hoeffding tree algorithm will be used to classify ID3 and CART methods. If the number of instances exceeds 500,000, the two traditional algorithms will crash, but the instance will be processed and offer a better result by enhancing the Hoeffding tree algorithm [5]. In this paper, a classification model is used for classifying the edible and poisonous mushrooms by using different machine learning algorithms. Algorithms such as neural networks, K-NN and decision trees are used. From the results, it has been observed for Eigen with real dimensions. The results of the experiment showcases the advantage of background images, especially when the KNN algorithm is used [6].

In this paper, ANN and ANFIS algorithms are used to classify mushrooms. The artificial mushroom was collected from the repository of UCI Machine Learning. As a result, the accuracy increases with an increase in the size of the dataset and a maximum of 70% of the entire data set is used as the training set [7]. In this paper, Artificial Neural Networks are employed for classifying the mushrooms [8]. In this paper, a classification model has been implemented for performing mushroom classification by incorporating the Supervised Learning Models. Liner Regression, Gaussian Naive-Bayes and Support Vector Machine are the methods utilized in this project [9]. In this project, a mushroom classification based behavioral model has been constructed. In the resuts, the dataset numbers show that the edible mushroom has a 0.518% absolute count of 4,208, while the poisonous mushroom has a 0.482% absolute count of 3,916. The analysis indicate that, the Decision tree utilizes the 23 leaves generated by the J48 classification model and the tree size is 28 [10]. In this paper, a system has been developed to identify the edible and poisonous mushrooms. The project uses feature selection and decision tree approaches to identify edible and poisonous mushrooms. Here, the tenfold Cross validation methodology has tested the efficacy of the model. As a result, ID3, C4.5 and Random Forest have formed three decision trees. Study results have shown that, the best result has been obtained by the model developed by Information Gain and Random Forest with 94.19% accuracy [11]. In this paper, convolution neural networks are used to analyze and classify different mushroom types. The authors have then divided the mushroom dataset into 45 types; 35 for edible types and 10 for poisonous mushroom types, based on the mushroom name [12].

In this paper, edible mushrooms are characterized, recognized and evaluated. At first, the size of the fruit body of mushrooms will be collected and calculated.

The various portions of fruit bodies such as cap, stalk, gills, volva, annulus, shape and color are reported as present and absent. After that, all the species identified were reported on the grounds of various characteristics displayed by the mushrooms [13]. In this paper, the principle of deep learning is used to identify the palatable mushrooms. First of all the classification model is subjected to evaluate image dataset, which comproses both edible and poisonous mushrooms to determine their class index. In the results we can see that, high FPR is a dangerous indication of the model because it classifies poisonous mushrooms as edible [14]. In this paper, the efficiency of algorithms involved in the identification of poisonous mushrooms using confusion matrix analysis has been evaluated [15].

3 Overall Proposed Architecture

The mushroom dataset from UCI machine repository is used here. The dataset used in this work has 22 independent variables and 1 target mushroom edibility target class. The dataset has 8124 mushroom information for all the 23 variables. The overall workflow is shown in Fig. 1.

The paper contributions is given below.

(i) Firstly, the dataset is preprocessed with feature scaling and missing values.

(ii) Secondly, the raw dataset is fitted to all the classifiers with and without the presence of feature scaling.

(iii) Thirdly, raw data is applied to the principal component analysis with 8, 10 and 12 components and PCA reduced dataset is then fitted to all the classifier with and without the presence of feature scaling.

(iv) Fourth, the raw data is applied with the linear discriminant analysis [LDA] and LDA reduced dataset, which is fitted to all the classifier with and without the presence of feature scaling.

(v) Fifth, raw data is applied with the Factor analysis with 8, 10 and 12 components and FA reduced dataset is fitted to all the classifier with and without the presence of feature scaling.

(vi) Sixth, the performance comparison of raw data set, PCA reduced data set, LDA reduced dataset and FA reduced dataset is done by analyzing the performance metrics like precision, recall, accuracy and F1-score.

4 Results and Discussion

4.1 Qualitative Classifier Analysis of the Dataset

The mushroom dataset obtained from the UCI database repository is used for execution. The code is drafted with python under Anaconda Navigator and Spyder IDE.

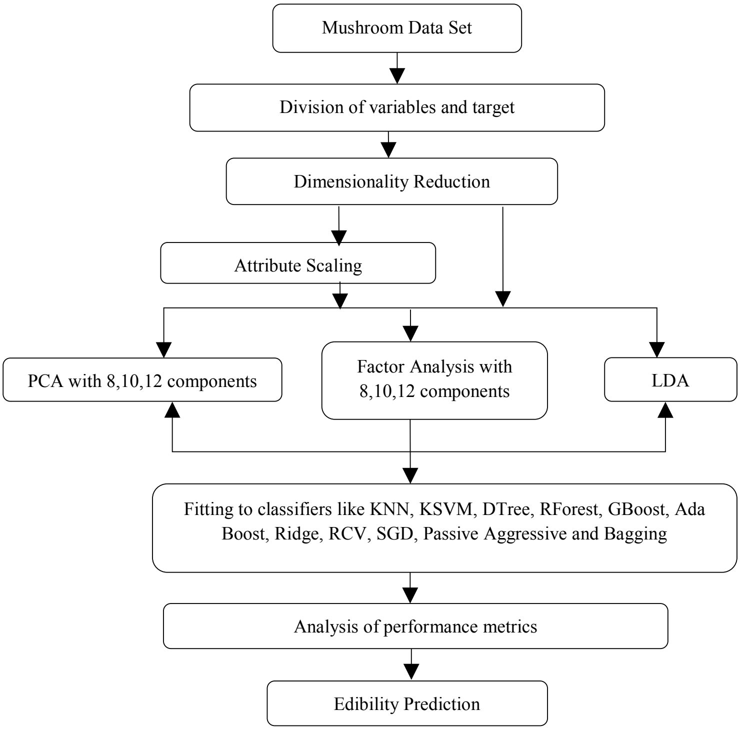

Mushroom Data Set
Division of variables and target
Dimensionality Reduction
Attribute Scaling
PCA with 8,10,12 components
Factor Analysis with 8,10,12 components
LDA
Fitting to classifiers like KNN, KSVM, DTree, RForest, GBoost, Ada Boost, Ridge, RCV, SGD, Passive Aggressive and Bagging
Analysis of performance metrics
Edibility Prediction

Fig. 1 System architecture flow

The Correlation of all the attributes towards the final dependent feature is done and feature discrimination is applied. The data set is splitted in such a way that the training has 80% of the dataset and the testing has 20% of the dataset. Attribute reduction is generally applied to minimize the values of varaibles in the dataset. It vastly affects the exploration of the target prediction with the independent variables. The raw data set is fitted to all the classifiers with and without presence of feature scaling. The performance analysis is done as shown in Tables 1 and 2 (Fig. 2).

4.2 *Quantitative Analysis with PCA*

The raw dataset is applied with the principal component analysis with 8, 10 and 12 components and further the PCA reduced dataset is fitted to all the above mentioned classifiers with and without the presence of feature scaling. The PCA reduced dataset is fitted to all the classifiers with and without the presence of feature scaling. The performance analysis for the PCA reduced dataset with 8 component is done and is shown in Tables 3 and 4.

Table 1 Performance metrics of raw data set before feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.955108682	0.955076923	0.95506341	0.955076923
KNeighbors	0.996307692	0.996307692	0.996307692	0.996307692
Kernel SVM	0.990254097	0.990153846	0.990149762	0.990153846
Gaussian Naive Bayes	0.916076133	0.915692308	0.915731245	0.915692308
Decision tree	1	1	1	1
ETree	1	1	1	1
RForest	1	1	1	1
GBoosting	1	1	1	1
ABoost classifier	1	1	1	1
Ridge classifier	0.950447707	0.950153846	0.950109386	0.950153846
Ridge classifierCV	0.950447707	0.950153846	0.950109386	0.950153846
SGD classifier	0.862293332	0.809230769	0.804112685	0.809230769
PAggressive	0.898304706	0.896	0.896049397	0.896
Bagging classifier	1	1	1	1

Table 2 Performance metrics of raw data set after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.956323313	0.956307692	0.956297704	0.956307692
KNeighbors	1	1	1	1
Kernel SVM	1	1	1	1
Gaussian Naive Bayes	0.916076133	0.915692308	0.915731245	0.915692308
Decision Tree	1	1	1	1
ETree	1	1	1	1
RForest	1	1	1	1
GBoosting	1	1	1	1
ABoost classifier	1	1	1	1
Ridge classifier	0.950447707	0.950153846	0.950109386	0.950153846
Ridge classifierCV	0.950447707	0.950153846	0.950109386	0.950153846
SGD classifier	0.985347351	0.985230769	0.985235274	0.985230769
PAggressive	0.939098065	0.939076923	0.939058597	0.939076923
BaggingClassifier	1	1	1	1

The performance analysis for the PCA reduced dataset with 10 component is done and is shown in Tables 5 and 6.

The performance analysis for the PCA reduced dataset with 12 component is done and is shown in Tables 7 and 8.

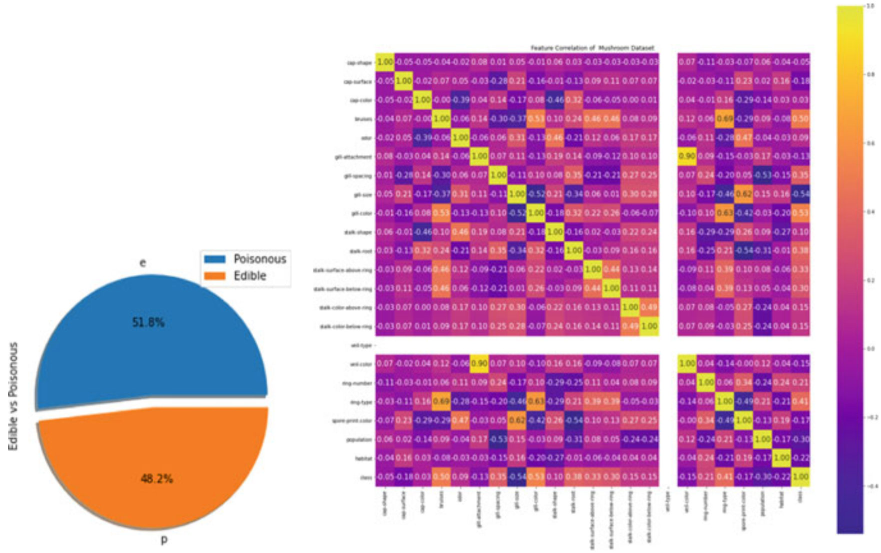


Fig. 2 Target class and correalation analysis

Table 3 Metrics of PCA 8 component dataset before feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.809789	0.809846	0.809803	0.809846
KNeighbors	0.974154	0.974154	0.974152	0.974154
Kernel SVM	0.986463	0.986462	0.986461	0.986462
Gaussian Naive Bayes	0.840425	0.839385	0.838933	0.839385
Decision tree	0.991464	0.991385	0.991387	0.991385
ETree	0.972543	0.972308	0.972319	0.972308
RForest	0.995077	0.995077	0.995077	0.995077
GBoosting	0.964803	0.964308	0.964267	0.964308
ABoost classifier	0.909104	0.908923	0.908846	0.908923
Ridge classifier	0.807957	0.808	0.807855	0.808
Ridge classifierCV	0.807957	0.808	0.807855	0.808
SGD classifier	0.792529	0.792615	0.792504	0.792615
PAggressive	0.78189	0.775385	0.77276	0.775385
BaggingClassifier	0.995106	0.995077	0.995078	0.995077

4.3 Quantitative Analysis with LDA

The raw dataset is applied with the Linear Discriminant Analysis [LDA] and LDA reduced dataset is fitted to all the above mentioned classifiers with and without the

Table 4 Performance metrics of PCA 8 component dataset after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.897935	0.896615	0.896347	0.896615
KNeighbors	0.996334	0.996308	0.996307	0.996308
Kernel SVM	0.99266	0.992615	0.992613	0.992615
Gaussian Naive Bayes	0.918854	0.918769	0.918722	0.918769
Decision tree	0.995693	0.995692	0.995692	0.995692
ETree	0.988927	0.988923	0.988924	0.988923
RForest	0.999385	0.999385	0.999385	0.999385
GBoosting	0.990785	0.990769	0.990768	0.990769
ABoost classifier	0.969888	0.969846	0.969837	0.969846
Ridge Classifier	0.903854	0.899692	0.899138	0.899692
Ridge classifierCV	0.903854	0.899692	0.899138	0.899692
SGD classifier	0.898498	0.897231	0.896971	0.897231
PAggressive	0.881059	0.870154	0.868563	0.870154
BaggingClassifier	0.993848	0.993846	0.993846	0.993846

Table 5 Metrics of PCA 10 component dataset before feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.809169	0.809231	0.80918	0.809231
KNeighbors	0.974872	0.974769	0.974758	0.974769
Kernel SVM	0.993247	0.993231	0.99323	0.993231
Gaussian Naive Bayes	0.843955	0.841846	0.841176	0.841846
Decision tree	0.992629	0.992615	0.992616	0.992615
ETree	0.988405	0.988308	0.988311	0.988308
RForest	0.996924	0.996923	0.996923	0.996923
GBoosting	0.961989	0.961846	0.961824	0.961846
ABoost classifier	0.930545	0.929846	0.929736	0.929846
Ridge classifier	0.803617	0.803692	0.803603	0.803692
Ridge classifierCV	0.803617	0.803692	0.803603	0.803692
SGD classifier	0.809206	0.808	0.808113	0.808
PAggressive	0.774454	0.771692	0.77015	0.771692
BaggingClassifier	0.99385	0.993846	0.993847	0.993846

presence of feature scaling. The PCA reduced dataset is fitted to all the classifiers with and without the presence of feature scaling. The performance analysis for the LDA reduced dataset is done and is shown in Table 9 and 10.

Table 6 Metrics of PCA 10 component dataset after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.906952	0.905846	0.905634	0.905846
KNeighbors	0.998772	0.998769	0.998769	0.998769
Kernel SVM	0.998155	0.998154	0.998154	0.998154
Gaussian Naive Bayes	0.935702	0.935385	0.935321	0.935385
Decision tree	0.994462	0.994462	0.994461	0.994462
ETree	0.995088	0.995077	0.995076	0.995077
RForest	0.999385	0.999385	0.999385	0.999385
GBoosting	0.991385	0.991385	0.991385	0.991385
ABoost classifier	0.981558	0.981538	0.981535	0.981538
Ridge classifier	0.909842	0.906462	0.906024	0.906462
Ridge classifierCV	0.909842	0.906462	0.906024	0.906462
SGD classifier	0.899443	0.899077	0.898951	0.899077
PAggressive	0.874805	0.874462	0.874292	0.874462
BaggingClassifier	0.99631	0.996308	0.996307	0.996308

Table 7 Metrics of PCA 12 component dataset before feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.886127	0.885538	0.885345	0.885538
KNeighbors	0.991408	0.991385	0.991383	0.991385
Kernel SVM	0.9957	0.995692	0.995693	0.995692
Gaussian Naive Bayes	0.886526	0.880615	0.879717	0.880615
Decision tree	0.99509	0.995077	0.995077	0.995077
ETree	0.986477	0.986462	0.986463	0.986462
RForest	0.998772	0.998769	0.998769	0.998769
GBoosting	0.979538	0.979077	0.979055	0.979077
ABoost classifier	0.945277	0.945231	0.94521	0.945231
Ridge classifier	0.895295	0.891077	0.890454	0.891077
Ridge classifierCV	0.895295	0.891077	0.890454	0.891077
SGD classifier	0.896139	0.893538	0.893105	0.893538
PAggressive	0.873927	0.873846	0.87387	0.873846
BaggingClassifier	0.995088	0.995077	0.995076	0.995077

4.4 Factor Analysis Before and After Feature Scaling

The raw dataset is applied with the principal component analysis with 8, 10 and 12 components and FA (Factor Analysis) reduced dataset is fitted to all the above mentioned classifiers with and without the presence of feature scaling. The FA

Table 8 Metrics of PCA 12 component dataset after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.913038	0.912	0.911813	0.912
KNeighbors	0.999385	0.999385	0.999385	0.999385
Kernel SVM	0.997538	0.997538	0.997538	0.997538
Gaussian Naive Bayes	0.932919	0.932923	0.932912	0.932923
Decision tree	0.994467	0.994462	0.994461	0.994462
ETree	0.996929	0.996923	0.996923	0.996923
RForest	0.999385	0.999385	0.999385	0.999385
GBoosting	0.996308	0.996308	0.996308	0.996308
ABoost classifier	0.991394	0.991385	0.991384	0.991385
Ridge classifier	0.914647	0.911385	0.910986	0.911385
Ridge classifierCV	0.914647	0.911385	0.910986	0.911385
SGD classifier	0.905512	0.904615	0.904425	0.904615
PAggressive	0.867397	0.863385	0.862562	0.863385
BaggingClassifier	0.999385	0.999385	0.999385	0.999385

Table 9 Metrics of LDA reduced dataset before Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.950985	0.950769	0.950732	0.950769
KNeighbors	0.961858	0.961846	0.961839	0.961846
Kernel SVM	0.953239	0.953231	0.953222	0.953231
Gaussian Naive Bayes	0.950448	0.950154	0.950109	0.950154
Decision tree	0.93419	0.934154	0.934163	0.934154
ETree	0.931856	0.931692	0.931715	0.931692
RForest	0.937863	0.937846	0.937852	0.937846
GBoosting	0.96123	0.961231	0.961227	0.961231
ABoost classifier	0.962464	0.962462	0.962463	0.962462
Ridge classifier	0.951036	0.950769	0.950728	0.950769
Ridge classifierCV	0.951036	0.950769	0.950728	0.950769
SGD classifier	0.949601	0.949538	0.949518	0.949538
PAggressive	0.8883	0.864615	0.863434	0.864615
BaggingClassifier	0.93661	0.936615	0.936609	0.936615

reduced data set is fitted to all the classifiers with and without the presence of feature scaling. The performance analysis for the FA reduced dataset with 8 component is done and is shown in Tables 11 and 12.

The performance analysis for the FA reduced dataset with 10 component is done and is shown in Tables 13 and 14.

Table 10 Performance metrics of LDA reduced dataset after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.950985	0.950769	0.950732	0.950769
KNeighbors	0.961858	0.961846	0.961839	0.961846
Kernel SVM	0.953239	0.953231	0.953222	0.953231
Gaussian Naive Bayes	0.950448	0.950154	0.950109	0.950154
Decision tree	0.93419	0.934154	0.934163	0.934154
ETree	0.931856	0.931692	0.931715	0.931692
RForest	0.937863	0.937846	0.937852	0.937846
GBoosting	0.96123	0.961231	0.961227	0.961231
ABoost classifier	0.962464	0.962462	0.962463	0.962462
Ridge classifier	0.951036	0.950769	0.950728	0.950769
Ridge classifierCV	0.951036	0.950769	0.950728	0.950769
SGD classifier	0.950808	0.950769	0.950753	0.950769
PAggressive	0.949081	0.947692	0.947723	0.947692
BaggingClassifier	0.937844	0.937846	0.937836	0.937846

Table 11 Metrics of FA 8 component dataset before feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.923581	0.921231	0.920956	0.921231
KNeighbors	1	1	1	1
Kernel SVM	1	1	1	1
Gaussian Naive Bayes	0.861178	0.860308	0.860388	0.860308
Decision tree	0.999385	0.999385	0.999385	0.999385
ETree	0.999385	0.999385	0.999385	0.999385
RForest	1	1	1	1
GBoosting	0.998769	0.998769	0.998769	0.998769
ABoost classifier	1	1	1	1
Ridge classifier	0.916607	0.913846	0.913503	0.913846
Ridge classifierCV	0.916607	0.913846	0.913503	0.913846
SGD classifier	0.932919	0.931077	0.930878	0.931077
PAggressive	0.944051	0.943385	0.943302	0.943385
BaggingClassifier	0.999385	0.999385	0.999385	0.999385

The performance analysis for the FA reduced dataset with 12 component is done and is shown in Tables 15 and 16.

Table 12 Metrics of FA 8 component dataset after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.923581	0.921231	0.920956	0.921231
KNeighbors	1	1	1	1
Kernel SVM	1	1	1	1
Gaussian Naive Bayes	0.880615	0.880615	0.880615	0.880615
Decision tree	0.998154	0.998154	0.998154	0.998154
ETree	1	1	1	1
RForest	1	1	1	1
GBoosting	0.998769	0.998769	0.998769	0.998769
ABoost classifier	1	1	1	1
Ridge classifier	0.916607	0.913846	0.913503	0.913846
Ridge classifierCV	0.916607	0.913846	0.913503	0.913846
SGD classifier	0.952327	0.952	0.951955	0.952
PAggressive	0.752266	0.635077	0.597167	0.635077
BaggingClassifier	0.998769	0.998769	0.998769	0.998769

Table 13 Metrics of FA 10 component dataset before feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.915851	0.914462	0.914245	0.914462
KNeighbors	1	1	1	1
Kernel SVM	1	1	1	1
Gaussian Naive Bayes	0.895396	0.895385	0.895341	0.895385
Decision tree	0.999385	0.999385	0.999385	0.999385
ETree	0.999385	0.999385	0.999385	0.999385
RForest	1	1	1	1
GBoosting	1	1	1	1
ABoost classifier	0.999385	0.999385	0.999385	0.999385
Ridge classifier	0.907609	0.903385	0.902851	0.903385
Ridge classifierCV	0.907609	0.903385	0.902851	0.903385
SGD classifier	0.957623	0.957538	0.957519	0.957538
PAggressive	0.768713	0.768615	0.768654	0.768615
BaggingClassifier	0.99385	0.993846	0.993847	0.993846

5 Conclusion

The proposed research work performs the variable analysis of the mushroom data by exploring the correlation between the features. The features are reduced with PCA, LDA and factor analysis and further the performance of classifiers are analyzed to

Table 14 Metrics of FA 10 component dataset after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.940263	0.939692	0.939611	0.939692
KNeighbors	1	1	1	1
Kernel SVM	1	1	1	1
Gaussian Naive Bayes	0.839126	0.835692	0.83572	0.835692
Decision tree	1	1	1	1
ETree	1	1	1	1
RForest	1	1	1	1
GBoosting	1	1	1	1
ABoost classifier	1	1	1	1
Ridge classifier	0.92094	0.918769	0.918498	0.918769
Ridge classifierCV	0.92094	0.918769	0.918498	0.918769
SGD classifier	0.941006	0.940308	0.940217	0.940308
PAggressive	0.801918	0.723692	0.699363	0.723692
BaggingClassifier	1	1	1	1

Table 15 Metrics of FA 12 component dataset before feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.946747	0.946462	0.946414	0.946462
KNeighbors	1	1	1	1
Kernel SVM	1	1	1	1
Gaussian Naive Bayes	0.891368	0.890462	0.890525	0.890462
Decision tree	0.999385	0.999385	0.999385	0.999385
ETree	1	1	1	1
RForest	1	1	1	1
GBoosting	1	1	1	1
ABoost classifier	1	1	1	1
Ridge classifier	0.938623	0.937846	0.937745	0.937846
Ridge classifierCV	0.938623	0.937846	0.937745	0.937846
SGD classifier	0.969894	0.969846	0.969852	0.969846
PAggressive	0.910839	0.901538	0.90056	0.901538
BaggingClassifier	0.999385	0.999385	0.999385	0.999385

find the best fit classifier for predicting the edibility. Experimental results shows that the kernel SVM, KNN and Adaboost classifier for the FA reduced dataset tends to retain the accuracy with 100% before and after feature scaling. The KNN classifier with LDA reduced dataset tends to retain the 96% accuracy before and after feature scaling. kernel SVM, KNN and Adaboost classifier for the FA reduced dataset tends

Table 16 Metrics of FA 12 component dataset after feature scaling

Classifier	Precision	Recall	FScore	Accuracy
LogisticRegression	0.940921	0.940308	0.940224	0.940308
KNeighbors	1	1	1	1
Kernel SVM	1	1	1	1
Gaussian Naive Bayes	0.900656	0.899077	0.898786	0.899077
Decision tree	1	1	1	1
ETree	0.999385	0.999385	0.999385	0.999385
RForest	1	1	1	1
GBoosting	1	1	1	1
ABoost classifier	1	1	1	1
Ridge classifier	0.925339	0.923692	0.923483	0.923692
Ridge classifierCV	0.925339	0.923692	0.923483	0.923692
SGD classifier	0.947337	0.947077	0.947032	0.947077
PAggressive	0.937282	0.936615	0.93665	0.936615
BaggingClassifier	0.999385	0.999385	0.999385	0.999385

to retain the accuracy with 99% before and after feature scaling. From the above analysis KNN classifier is found to be more efficient in its accuracy with all PCA, LDA and FA reduced dataset.

References

1. V. Maniraj, J. Nithya, Integrating ontology with datamining with a case of mushroom analysis. COMPUSOFT. Int. J. Adv. Comput. Technol. **IV** (2015)
2. D.R. Chowdhury, S. Ojha, An emperical study on mushroom disease diagnosis: a data mining approach. Int. Res. J. Eng. Technol. (IRJET) **4**(101) (2017)
3. M. Hussaini, A data mining based on ensemble classifier classification approach for edible mushroom classifiacation. Int. Res. J. Eng. Technol. (IRJET) **5**(7) (2018)
4. U.J. Lidasan, P.M. Tagacay, Mushroom recognition using neural network. Int. J. Comput. Sci. Issues (IJCSI) **15**(05) (2018)
5. B. Lavanya, G.R. Preethi, Performance analysis of decision tree algorithms on mushroom dataset. Int. J. Res. Appl. Sci. Eng. Technol. (IJTESET) **5**(11) (2017)
6. M.A. Ottom, N.A. Alawad, M.O.K. Nahar, Classification of mushroom fungi using machine learning techniques. Int. J. Adv. Trends Comput. Sci. Eng. (IJATCSE) **8** (2019)
7. S.K. Verma, M. Dutta, Mushroom classificaiton using ANN and ANFIS algorithm. IOSR J. Eng. (IOSRJEN) **8** (2018)
8. E.S. Alkronz, A.K. Moghayer, M. Meimeh, M. Gazzaz, Classification of mushroom using artificial nueral network. Int. J. Acad. Appl. Res. (IJAAR) **3**(2) (2019)
9. J.B. Chelliah, S. Kalaiarasi, A. Anand, G. Janakiram, B. Rathi, K.N. Warrier, Classification of mushrooms using supervised learning models. Int. J. Emerg. Technol. Eng. Res. (IJETER) **6**(4) (2018)
10. S. Ismail, A. Mustapha, Behavioral features for mushroom classification, in *IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE)*, April (2018)

11. S. Nuanmeesri, W. Sriurai, Development of edible and poisonous mushrooms classification model by using the feature selection and the decision tree techniques. Int. J. Eng. Adv. Technol. (IJEAT) **9**(2) (2019)
12. O. Chaowalit, P. Visutsak, Image analysis of mushroom types classification by convolution neural networks. Research Gate, December (2019)
13. M. Yadav, R. Chandra, S.K. Yadav, P.K. Dhakad, Morphological characterization, ıdentification and edibility test of edible mushrooms from vindhya forest of Northern India. Research Gate, March (2017)
14. S. Kavitha, R.L. Suganthi, J. Jose, Ensemble deep learning for prediction of palatable mushrooms. Int. J. Eng. Sci. Res. (IJESR) **6**(1) (2018)
15. J. Heland, C.J. Ortega, C.A. Langman, Q.L.R. Natividad, T.E. Bantung, R.M. Resureccion, Q.L.J. Manalo, Analysis of performance of classification algorithms in mushroom poisonous detection using confusion matrix analysis. Int. J. Adv. Trends Comput. Sci. Eng. (IJATCSE) **9**(1.3) (2020)

Flip-Flop-Based Approach for Logic Encryption Technique

M. Anupama and S. R. Ramesh

Abstract Logical encryption is the key method used for the protection of hardware by preventing overproduction of llegal integrated circuits and piracy of IP. Through this technique, additional key input has been introduced into the given circuit. In this work, the design is modeled using a technique called flip-flop encryption through inserting multiplexers in random locations of the circuit. The logic is implemented in various ISCAS'89 benchmark circuits, and the overheads are compared in smaller and larger benchmark circuits. In addition, the proposed logic has been implemented in the critical path of the circuit. The power variation is reported from 1.6% to 1.9% and area is below 2.06% compared to random location of key gates. This technique has been implemented in the field of hardware security by preventing Trojans and reduces IC overproduction.

Keywords Logic encryption · Key gates · Integrated circuits · Encrypt flip-flop · Hardware security

1 Introduction

Manufactured ICs are threatened by number of untrusted third parties. The attackers can reverse engineer these ICs. These are the main issue of intellectual property piracy raised by the fabrication industry. Through this method, they can state the ownership of the IP [1–3]. Therefore, from these designs, the ICs are over produced and sold illegally by the untrusted foundries. The hardware IPs are protected by encrypting the functionality of the circuits. The overproduction of circuits also can be protected by encryption which allows only the authorized use of circuits. The functionality of the circuits can be hidden by the method of inserting key gates controlled by key bits. This design for test [DFT] technique used for encrypting

M. Anupama (✉) · S. R. Ramesh
Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India

S. R. Ramesh
e-mail: sr_ramesh@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_2

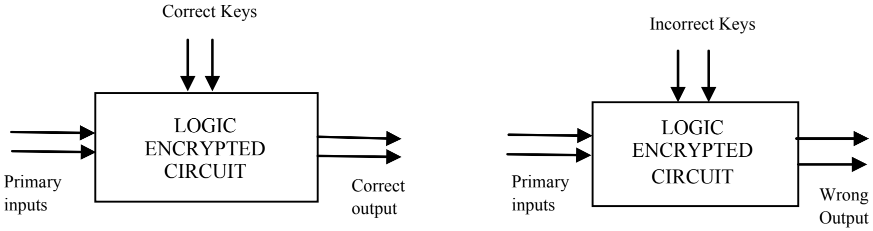


Fig. 1 Block diagram of logic encryption

the circuit functionality is called logic encryption which allows authorized users to operate the chips [4].

Logic encryption is an area of research that has introduced to prevent the ICs from the risk activities such as insertion of hardware Trojans, piracy and overproduction. Logic encryption method increases the security of an IC [5]. The insertion of key gates into original netlist is the most existing logic encryption method. But these techniques result in high area, power and performance overheads. The reduction in overhead allows the logic encryption technique to be used in various applications [6].

Using logic encryption technique when key input is added into the combinational circuit, the encrypted circuit will operate same as the original circuit for certain key values. The wrong output will be produced for other incorrect key values as shown in Fig. 1. The correct keys are required to activate the IP. Therefore, the secret keys are protected from hackers by storing in the memory of chips. This technique prevents the overproduction ICs which cannot be sold until correct functionality of the circuit is exhibited with the correct keys.

1.1 Methods of Logic Encryption

The type of logical encryption includes sequential and combinational. In serial logical encryption, upon application of an incorrect input sequence, the additional logical states called as black state modifies the state transition graph. This limits the use of chip. Valid states are only achieved after applying the correct input sequence [7]. To protect the functionality, the logic encryption is introduced by inserting the extra key inputs to the combinational device. Any additional logic elements which include basic gates, universal gates, multiplexers and look up table are used as key gate.

In XOR/XNOR based encryption, the gates are inserted randomly to the module. The one input of the key gates is connected internally and other act as key input. While applying correct keys, they are configured as buffers else the lines which are inverted gives incorrect output. The XOR or LUT-based implementations have higher area, power and performance overhead which increases security. But the large per gate overhead limits the performance. The encryption using this gate is more complex. In

AND/OR-based encryption, the key gates used are AND/OR. When true or false keys are given to the encrypted circuit, the gate produces correct result which confuses the attacker.

Figure 2a shows unencrypted circuit with AND, XOR and OR elements. The insertion of different key gates leads to have some hidden functions. In MUX-based encryption technique as shown in Fig. 2b where 2X1 mux are inserted into the circuit. For 2X1 mux, both input wires chosen as true and false whereas the select line as key bit. If the value of key gate is correct, then MUX propagates true value from the correct input else it takes false input. If XOR is replaced with XNOR element, the attacker cannot guess the correct key or differentiate that inverter belongs to the key gate.

According to the existing logic encryption technique where XOR/XNOR key gates are inserted random at the circuit leads to high power and area overhead. The gate replacement technique used in the existing work results in the reduction of overheads without encryption quality degradation. But this technique can only be implemented for XOR/XNOR-based logic encryption [9].

This paper aims to reduce the area and power overheads by using MUX as the key gates. This work demonstrates flip-flop encryption by inserting multiplexers in two methods and compare the overheads.

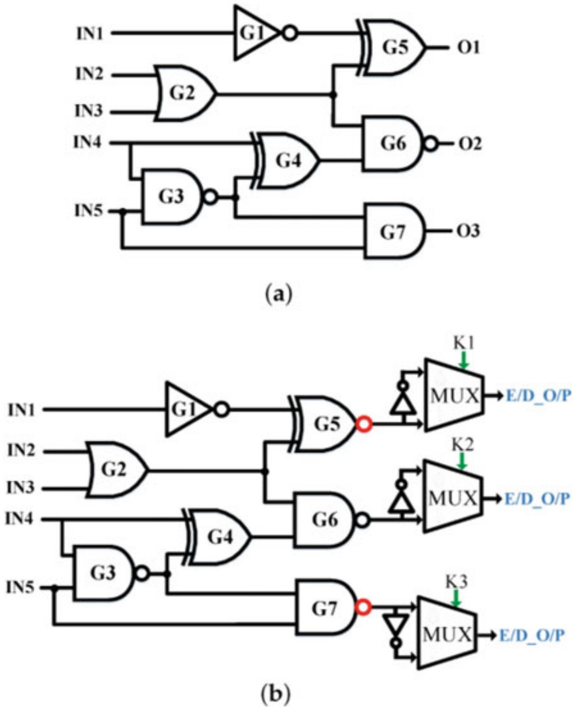


Fig. 2 **a** Original circuit; **b** encrypted circuit using MUX key gate [8]

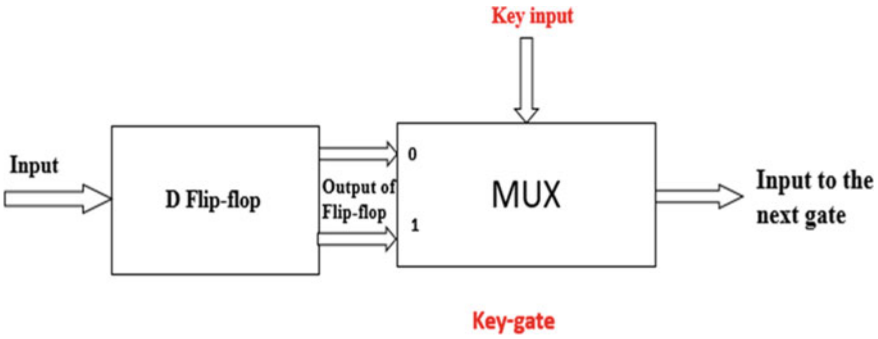


Fig. 3 Basic flip-flop encryption

2 Proposed Work

In this work, logic encryption technique termed as encrypt flip-flop is considered where responses of the flip-flop are encrypted. Consider the flip-flop producing two outputs as shown in Fig. 3. With the insertion of MUXes in front of the outputs of flip-flop, the circuit will be encrypted. The output lines of flip-flop are connected to both inputs of the MUX. The output line of MUX is linked to next logic level. MUX is used as the key gate and the select line as key shown in Fig. 2. Depends on the value of select line, either one of the output line produces output to the next logic gate [10]. Figure 3 depicts basic encrypt flip-flop strategy used in this work.

Key input of the key gate will have value as 0 or 1. The wrong input of key causes the inverted input pass to next level, which results in an incorrect function of the circuit. In MUX-based encryption, the true and false lines are the two inputs of key gate which carry either 0 or 1 for certain input combination. The propagation of correct value for a wrong key to the next logic level can be avoided by considering output of the flip-flop as inputs of key gate.

In the proposed work, the location of key gate is chosen at critical path of the benchmark circuit. Several two input MUXes are inserted at the output of the flip-flop located at the critical path of the circuit. The flip-flop encryption is performed by choosing the critical path of the ISCAS’89 benchmark circuits [11]. Synthesis has been done and evaluated the power and area utilization for smaller and larger benchmark circuits using Xilinx Vivado tool. Simulation of benchmark circuit has been done in modelsim and the encrypted circuit for all the input combinations of the circuit with and without keys, respectively.

3 Implementation and Experimental Results

The encrypt flip-flop technique is applied on ISCAS’89 [12, 12] benchmarks of different sizes, which is encrypted using a MUX. In this MUX-based encryption, several two input-based MUX are inserted into the design. The flip-flops are selected in random and encrypted them by inserting a MUX in front of each of them. Figures 4 and 5 show the implementation of encrypt flip-flop technique on ISCAS'89 benchmark circuit where key gates inserted at random locations in the circuit. This encryption technique results in remarkable area, power and performance overheads. Figures 4 and 5 show the comparison results of area and power utilization for the actual and encrypted circuit.

Experimental results in Figs. 4 and 5 shows that flip-flop encryption leads to relatively low or no area overhead and power overhead for large benchmark circuits. It is observed that for benchmark circuits with less number of gates, the area and power overhead is more when flip-flop encryption technique is used. The variation in power is from 3.31% to 13.2%, and area is below 1.09%.

Figure 6 illustrates the simulation output of the S298 benchmark circuit by random insertion of key gates using flip-flop encryption technique. In S298 benchmark circuit, the G0, G1, G2 are the inputs to the circuit as shown in the figure. After encryption,

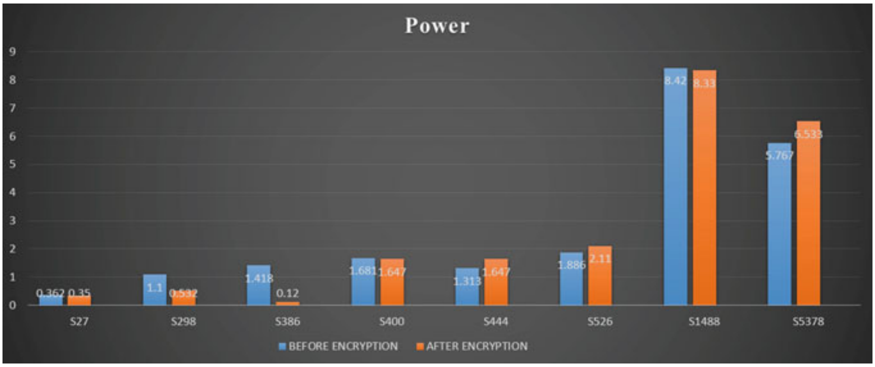


Fig. 4 Power analysis of benchmark circuit by random insertion of key gates

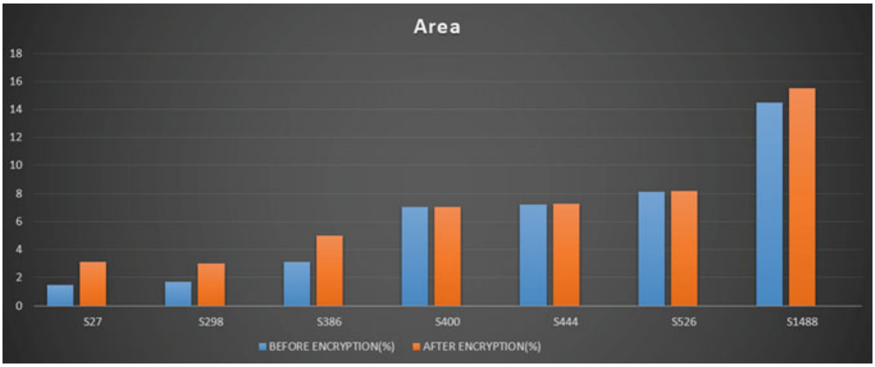
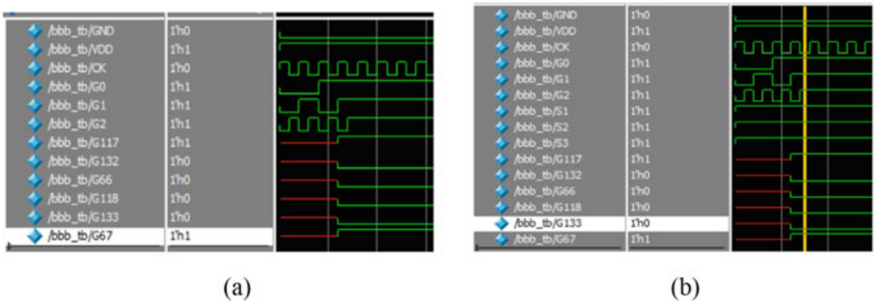


Fig. 5 Area analysis of benchmark circuit by random insertion of key gates



```
/bbb_tb/GND   1'h0
/bbb_tb/VDD   1'h1
/bbb_tb/CK    1'h0
/bbb_tb/G0    1'h1
/bbb_tb/G1    1'h1
/bbb_tb/G2    1'h1
/bbb_tb/G117  1'h1
/bbb_tb/G132  1'h0
/bbb_tb/G66   1'h0
/bbb_tb/G118  1'h0
/bbb_tb/G133  1'h0
/bbb_tb/G67   1'h1
```

(a)

```
/bbb_tb/GND   1'h0
/bbb_tb/VDD   1'h1
/bbb_tb/CK    1'h0
/bbb_tb/G0    1'h1
/bbb_tb/G1    1'h1
/bbb_tb/G2    1'h1
/bbb_tb/S1    1'h1
/bbb_tb/S2    1'h1
/bbb_tb/S3    1'h1
/bbb_tb/G117  1'h1
/bbb_tb/G132  1'h0
/bbb_tb/G66   1'h0
/bbb_tb/G118  1'h0
/bbb_tb/G133  1'h0
/bbb_tb/G67   1'h1
```

(b)

Fig. 6 S298 benchmark circuit simulation output using model sim tool. **a** Before encryption, **b** after encryption

the inputs are G0, G1, G2, S1, S2, S3 where S1, S2, S3 are the keys for the circuit. One of the correct key is 111.

Consider S27 benchmark circuit, before encryption, the inputs to the circuit are G0, G1, G2 and G3. After encryption, the inputs are G0, G1, G2, G3, S1, S2 and S3, respectively, where S1, S2 and S3 are the keys for the circuit. The correct keys are 000 and 100. Similarly, correct keys of different benchmark circuits can be found out.

Therefore, the logic encryption technique protects the circuit from attackers as they need to find the correct keys. The advantage of this method is impossible to find the correct key when it is used in larger benchmark circuits. This results in wrong output for the incorrect key inputs. The encryption quality cannot be guaranteed when key gates are inserted at random positions.

Experimental results in Table1 show that flip-flop encryption with insertion of key gates at critical path leads to relatively low area and power overhead for all the benchmark circuits. The variation in power is 1.6–1.9%, and area is below 2.06% compared to random location of key gates.

4 Conclusion and Future Work

This paper describes the details of logical encryption technique for flip-flop encryption. The key gates chosen is MUX for more complex obfuscation. Using encrypt flip-flop technique, the outputs of selected flip-flops are encrypted by inserting a MUX. The comparison of overheads is performed on various ISCAS'89 benchmark circuits by inserting key gates in random and in critical path of the circuit. When MUX is used in flip-flop encryption as key gate, the overhead of the circuit is reduced. In this work by the insertion of key gates at the random locations in the benchmark circuits, the logic encryption technique is demonstrated. The insertion of key gates at the critical path results in low design overhead for all benchmark circuits compared to random insertion method. This encryption methods can be performed in larger

Table 1 Overhead results of insertion of key gate at critical path

Circuits	Parameters	Before encryption	After encryption
S27	Power (W)	0.362	0.355
	Area utilization (%)	1.5	1.51
S298	Power (W)	1.1	1.12
	Area utilization (%)	1.7	1.9
S386	Power (W)	1.418	1.886
	Area utilization (%)	3.12	3.08
S400	Power (W)	1.681	1.775
	Area utilization (%)	7.07	7.07
S444	Power (W)	1.313	1.605
	Area utilization (%)	7.23	7.2
S526	Power (W)	1.886	3.978
	Area utilization (%)	8.12	8.2
S1488	Power (W)	8.42	8.33
	Area utilization (%)	14.5	14.2

circuits where it is impossible for the attacker to find out keys to unlock the operation of the circuit. The encryption quality cannot be guaranteed using these methods. The future work can be done to improve the encryption quality.

References

1. K. Pritika, M. Vinodhini, Logic encryption of combinational circuits, in *3rd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)* (2019)
2. S. Dupuis, P.-S. Ba, G. Di Natale, M.-L. Flottes, B. Rouzeyre, A novel hardware logic encryption technique for thwarting illegal overproduction and hardware trojans. IOLTS 49–54 (2014)
3. G.S. Nandith, S.R. Ramesh, A novel approach for statistical parameter estimation and test pattern generation, in *4th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India (2020), pp. 545–550. https://doi.org/10.1109/ICOEI48184.2020.9142902
4. S.R. Ramesh, R. Jayaparvathy, Artificial neural network model for arrival time computation in gate level circuits. Automatika **60**(3), 360–367. https://doi.org/10.1080/00051144.2019.1631568
5. K. Juretus, I. Savidis, Reducing logic encryption overhead through gate level key insertion, in *IEEE International Symposium on Circuits and Systems (ISCAS)* (2016)

6. N.V. Teja, E. Prabhu, Test pattern generation using NLFSR for detecting single stuck-at faults, in *International Conference on Communication and Signal Processing (ICCSP),* Chennai, India (2019), pp. 0716–0720. https://doi.org/10.1109/ICCSP.2019.8697949
7. M.A. Kiryakina, S.A. Kuzmicheva, M.A. Ivanov, Encrypted PRNG by logic encryption, *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (2020)
8. Q. Alasad, J.-S. Yuan, Y. Bi, E2LEMI: energy-efficient logic encryption using multiplexer insertion. Electronics **6**(1) (2017). https://doi.org/10.3390/electronics6010016
9. X. Chen, Q. Liu, Y. Wang, Q. Xu, H. Yang, Low overhead implementation of logic encryption using gate replacement techniques, in *18th International Symposium on Quality Electronic Design* (2017)
10. R. Karmakar, S. Chattopadhyay, R. Kapur, Encrypt flip-flop: a novel logic encryption technique for sequential circuits. 1–14 (2018)
11. V.S. Rathor, B. Garg, G.K. Sharma, A novel low complexity logic encryption technique for design-for-Trust. IEEE Trans. Emerg. Topics Comput. (2018)
12. M. Hemachand, E. Prabhu, Secured netlist generation using obfuscation technique. J. Crit. Rev. **7**(4) (2020)
13. https://filebox.ece.vt.edu/~mhsiao/iscas89.html

Analysis and Implementation of Less Delay and Low Area DT Multiplier

J. N. Swaminathan, K. Ambujam, V. Madhava Reddy, P. Mahesh, Ch. Harika, P. Keerthi, B. Keerthi, and B. Aksa

Abstract In modern day multimedia processing applications and instruments, the multiplier plays an important role. There are three major approvals to boost the efficiency of multipliers. They are Power, Area, and Delay. To increase the efficiency of multipliers is to reduce the range and delay. The results are executed and fetched through delay and area. We describe three multipliers in this article. The Array multiplier is one of the three multipliers, is the simplest method and high efficiency, but due to the large number of partial products, it suffers from high propagation delay. The Wallace tree multiplier technique which is proposed to solve the Array multiplier problem. It has less delay and high performance because, compared to the Array multiplier, it reduces the amount of partial products but needs a wide region. The Dadda tree multiplier is the fastest multiplier and is used to solve the Wallace tree multiplier problem, its output is also high, and in early stages it decreases the area partial products

Keywords Wallace tree multiplier · Array multiplier · Delay · Area · Dadda tree multiplier · Verilog · Xilinx-14-7 Version

1 Introduction

Most of the major digital circuits and digital signal processing systems rely with multiplier execution, and it is necessary to increase the speed of any digital system [1]. At present, the primary factor in deciding the instruction cycle time of digital signal processing is still multiplication time [2]. The process of simple multiplications is Partial product generation and then Partial product addition. After the earlier two operation Final addition will be executed [3–6]. To build logic, we use Verilog here. To construct the code in a very simple and effective manner, we use Verilog. To

J. N. Swaminathan (✉) · V. Madhava Reddy · P. Mahesh · Ch. Harika · P. Keerthi · B. Keerthi · B. Aksa
QIS College of Engineering and Technology, Ongole 523272, Andhra Pradesh, India

K. Ambujam
AKT Memorial College of Engieenring and Technology, Kallakurichi 606202, Tamilnadu, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_3

design the multipliers, we use logic-gates, half and full adder. In Half Adder and Full Adder, gates such as XOR, AND and OR Gates are used for design. For instance, the multiplicand (A) and multiplier (B) [7–10]. For each multiplication, we perform multiplication with shifting and adding operations for each piece. Let $A = 1011$ $B = 1001$ and 1011 × 1001 (Multiplicand and Multiplier) 1011, 0000, 0000, 1011 are Partial Product and 10,011 is Final Result. Multiplicand is A and Multiplier is B. In order to minimize the delay time, multiplication is often used in the output of certain instructions in a fast or effective way.

1.1 Array Multiplier

As a consequence of its normal structure, Array Multipliers are well known. These multipliers are fully focused on the operations of adding and moving. In each multiplier, the key thing organized in a particular way by the use of adder and shifter is partial product. Partial product is multiplying multiplier to the each of the multiplicands. The number which is to undergo multiplication is that "Multiplicand", and therefore number which it is undergone multiplication is "multiplier". Usually, the multiplier is placed first and therefore the multiplicand is placed. second, however, the primary factor is that the multiplicand and the second is the multiplier [11, 12]. In Array Multiplication, to reduce the delay in both 4-bit and 8-bit multiplication, we use both half and complete adders (Fig. 1).

The value delay is 2.63 ns while we are using full Full Adders in the 4-bit multiplication. In 8-bit multipliers, gates, half and full adders is larger compared to 4-bit multipliers. The 8-bit partial product structure is depicted below. Here w_1, w_2, w_{64} are the Partial Products of the Multiplication. With each multiplicand, we are the multiplication (AND Gate) of any multiplier. Here I am from one to 64 (Fig. 2).

Where w_1, w_2, soon are the Partial Product of the multiplication. p_0, p_2, p_3, p_{15} are the Final Result. The computation delay of 8-bit multipliers for all complete adders is 10.786 ns. We have a delay of 8.760 ns for both full additives and half adder circuits. The total number of LUTs present for Array type Multiplier is 101. The area will be greater depending on the number of LUTs.

1.2 Wallance Multiplier

The Wallace Multiplier was proposed by Chris Wallace, an Australian computer scientist, in 1964 [13–15]. We use a row reduction technique to maximize the output to minimize the delay time relative to the Array Multiplier. An easy multiplier is the Wallace-Tree type Multiplier. We use half and full adders to compress row numbers in the partial product. For Wallace multiplier, there are primarily three steps. Partial products using logic gates are created, the partial product number is reduced by half

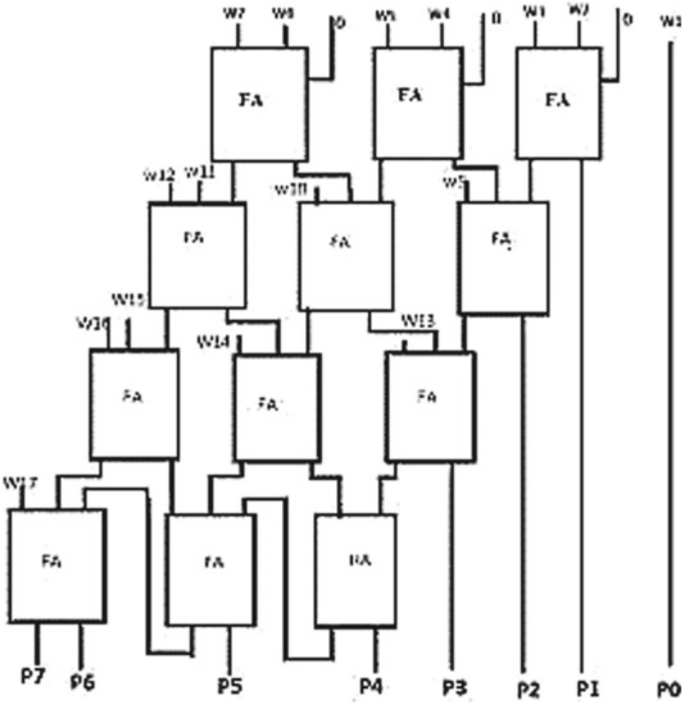


Fig. 1 4-Bit array multiplier

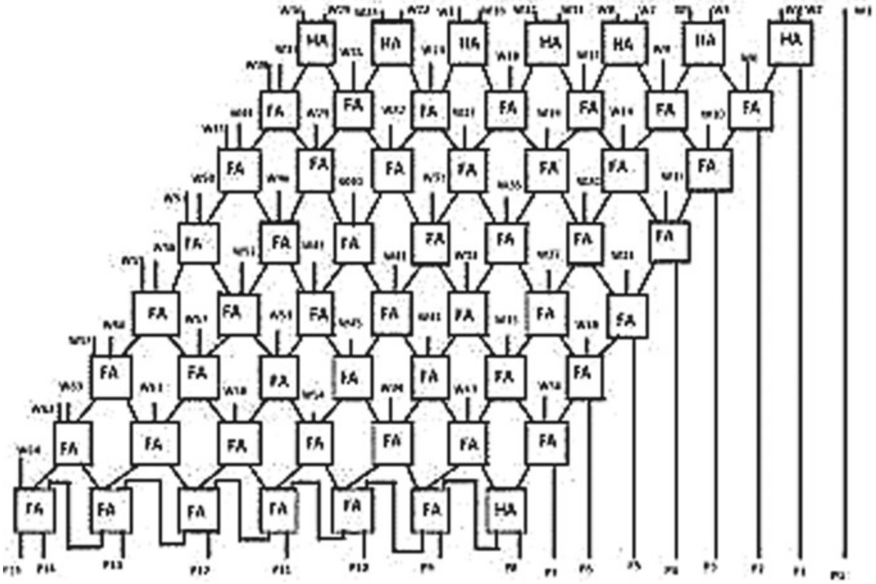


Fig. 2 8-bit array multiplier

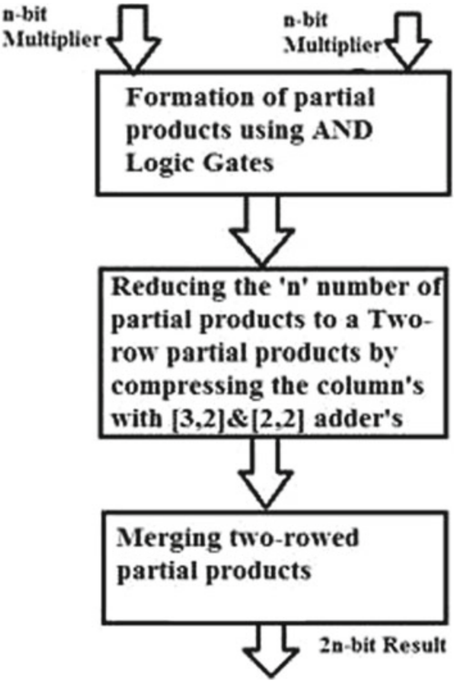

n-bit Multiplier
n-bit Multiplier
Formation of partial products using AND Logic Gates
Reducing the 'n' number of partial products to a Two-row partial products by compressing the column's with [3,2]&[2,2] adder's
Merging two-rowed partial products
2n-bit Result

Fig. 3 Flow chart of three stages in wallace multiplier

and full adders, so two rows of partial products are finally merged with propagate adder carriers (Fig. 3).

The Wallace-tree multiplier three steps:

1. Multiply (that is-AND) each bit of one of the arguments, yielding results of n^2, by each bit of the other. The wires bear different weights, depending on the location of the multiplicated bits.
2. Lower the partial product numbers by layers of complete and half adders to two.
3. Group wires into two and connect traditional adder to them.

The delay of the 4-Bit multiplier is 2.25 ns. The Structure of 8-bit Wallace tree Multiplier is additionally followed three stages of the Wallace tree multiplier. Using this logic, the delay is given as 2.97 ns. The multiplier delay is reduced to more than 60% relative to the collection. In Wallace, the number of LUTs present is 80. So, as compared to the collection, the area decreases. present in Wallace is 80. the area reduces when compared to Array [16, 17]. We are going for the Dadda tree multiplier to cut the delay of Wallace-tree multiplier.

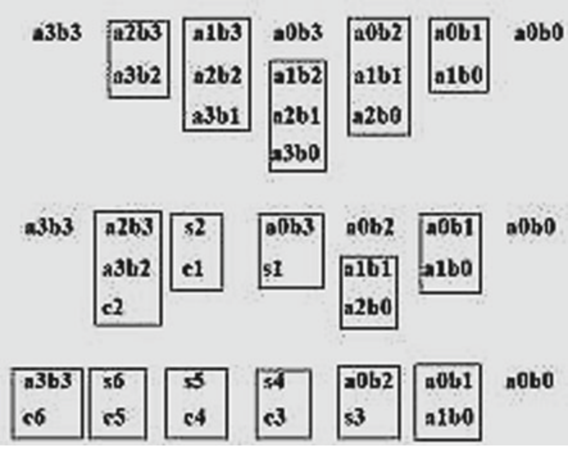


Fig. 4 Structure of 4-bit DT multiplier

2 Proposed DT Multiplier

The Dadda-Multiplier is the concept of a multiplier hardware devised in 1965 by Luigi Dadda. The tree-like structure of the Dadda-Multiplier is similar to Wallace-Tree Multiplier. There are also three major components of the Dadda-Tree Multiplier is intiated by Partial Product Generation followed with Partial Product Addition then Final Addition. In these multipliers, key part modified partial product is devised. The shape of the multiplier here is the shape of a tree. Due to their high computational velocity, row compression multipliers are more common. Both Wallace and Dadda multipliers have the technique of column compression. Data multipliers have less delays compared to Wallace Tree multipliers. The reduction is accomplished by using counters that are [3:2] counters (full adders) and [2:2] counters to decrease number of rows (Half Adders). We are conducting this operation in three phases here. The three phases that use AND logic Gates to form partial products, diminish the partial products count by using half and full adders, and combine two rows of partial products with propagation adder carriers (Fig. 4).

The way we execute partial products in the dadda multiplier varies from the Wallace tree multiplier. So, the delay in the multiplication of the Dadda tree is reduced. The multiplier's delay is 1.30 ns.

3 Result and Implementation

1. We organize the Partial Product into the Tree-like system in Dadda Multiplier.
2. Here is the multiplicand and multiplier, '*a*' and '*b*'. By using 3 [2, 3] counters and 3 [2] counters, we reduce these rows.

Fig. 5 LUT usage of DT versus Wallance versus Array

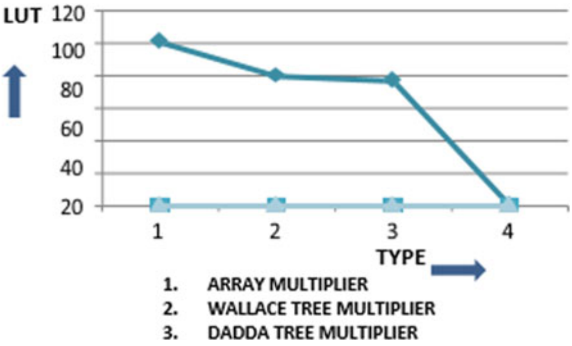
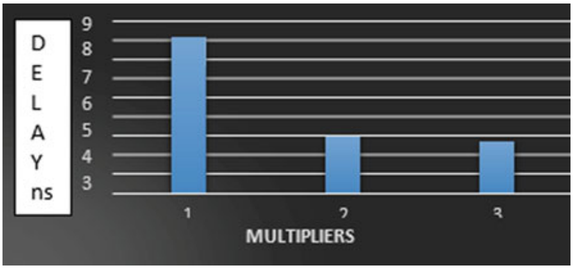


Fig. 6 LUT usage of DT versus Wallance versus array



3. From step one, we're reduced to six rows. Here s_1, s_2, … and c_1, c_2, … are the results of [2, 3] counter and [2] counter that's half adders and full adders respectively.
4. Here we compress the rows to four rows by using full and half adders.
5. Form fourth step we reduce row count to three by using nine [3:2] counters and one [two, 2] counter
6. This is the outcome of the partial products of two rows (Fig. 6).

Following this, by using carry propagation, we are reducing step 6 and got a delay of 2.70 ns by applying this logic. The number of LUTs in Dadda's tree multiplier is 77. The delay of the Dadda tree multiplier is realized in less area, which is compared with both Array & Wallace Multipliers. the Dadda tree Multiplier is more accurate which is proved area of the logic-gates and computation delay (Fig. 6).

4 Conclusion

In this paper, discussed above three multipliers are compared based on their delay and area. Array is a simple method and it is very complex. Wallace tree multiplier is high area complex and less complex. Therefore, to improve the performance of multipliers dadda tree multiplier is more effective. Therefore this paper proves that the Dadda

Table 1 Delay and LUT analysis of array versus Wallance versus DT

Multiplier	Delay (ns)	LUT'S
Conventional array	14.146	131
Proposed array	8.136	101
Conventional wallace	3.268	106
Proposed wallace	2.97	80
Conventional dadda	2.97	80
Proposed dadda	2.70	77

tree-multiplier is the faster multiplier than array and Wallace-tree multipliers (Table 1).

Acknowledgements The Authors want to acknowledge Dr. N. S. Kalyan Chakravarthy, Chairman and Correspondent, QIS Group of Institutions, Ongole, Andhra Pradesh.

References

1. K. Vikas et al., Comparative analysis of proposed parallel digital multiplier with Dadda and other popular multipliers. Int. J. **4**, 237–240 (2017)
2. S. Vaidya, D. Dandekar, Delay-power performance comparison of multipliers in VLSI circuit design. Int. J. Comput. Netw. Commun. (IJCNC) **2**(4) (2010)
3. K.B. Jaiswal, V.N. Kumar, P. Seshadri, G. Lakshminarayanan, Low power Wallace tree multiplier using modified full adder, in *3rd International Conference on Signal Processing, Communication and Networking (ICSCN)* (2015)
4. M. Naresh, B. Suneetha, Design of low power full adder based Wallace tree multiplier using cadence 180 nm technology. Int. J. Innov. Res. Sci. Eng. Technol. **6**(5) (2017)
5. S. Venkateswara Reddy, Design and implementation of 32 bit multiplier using Vedic mathematics. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **2**(8)
6. A.C. Swathi, T. Yuvraj, J. Praveen, A. Raghavendra Rao, A proposed wallace tree multiplier using full adder and half adder. Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng. **4**(5) (2016)
7. C. Senthilpari, A.K. Singh, K. Diwadkar, Low power and high speed 8×8 bit multiplier using non-clocked pass transistor logic. 1-4244-1355-9/07, 2007, IEEE
8. C.N. Marimuthu, P. Thiangaraj, Low power high performance multiplier, in *ICGST-PDCS*, vol. 8 (2008)
9. T. Sharma, Prof. B.P. Sing, K.G. Sharma, N. Arora, High-speed array multipliers based on 1-bit full adders. Recent Trends Eng. Technol. **4**(4), 26–28 (2010)
10. S. Vaidya, D. Dandekar, Delay-power performance comparision of multipliers in VLSI circuit design. Int. J. Comput. Netw. Commun. (IJCNS) **2**(4) (2010)
11. J.D. Lee, Y.J. Yoony, K.H. Leez, B.-G. Park, Application of dynamic pass transistor logic to 8-bit multiplier. J. Korean Phys. Soc. **38**(3), 220–223 (2001)
12. S. Neeta, R. Sindal, Modified booth multiplier using wallace structure and efficient carry select adder. Int. J. Comput. Appl. **68**(13) (2013)
13. D. Lugi, Some schemes for the parallel multipliers. Alta frequenz (May 1965)
14. J.-Y. Kang, J.-L. Gaudiot, A fast and well-structured multiplier, in *proceedings of the EUROMICRO systems on the digital system design (DSI'04)*, September 2004, pp. 508–515

15. C.R. Daugh, D.A. Wooley, Two's complement parallel array multiplication algorithm. IEEE Trans. Comput. **C-22**(12), 1045–1047 (1973)
16. C.W. Wallace, Suggestion for the fast multiplier. IEEE Trans. Electron. Comput. **EC-13**(1), 14–17 (1964)
17. N.S. Kalyan Chakravarthy, O. Vignesh, J.N. Swaminathan, High speed and low power buffer based parallel multiplier for computer arithmetic, in *Inventive Communication and Computational Technologies. Lecture Notes in Networks and Systems* ed. by G. Ranganathan, J. Chen, Á. Rocha, vol. 145. (Springer, Singapore, 2021). https://doi.org/10.1007/978-981-15-7345-3_34

An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease

S. Sharanyaa, P. N. Renjith, and K. Ramesh

Abstract Parkinson's disease is a neurodegenerative disorder, and it is a progressive disease that affects the human lifestyle. The survey described here paves the way to understand how the voice features are helpful in classification and prediction of Parkinson's disease among several motor and non-motor symptoms used for diagnosis. This article focuses on dysphonia, a voice disorder that leads to alteration in voice quality of the people. Since there are limited work carried out on voice pathology detection in Parkinson's disease, this research aims to provide key contribution on comparative analysis study in this field. Also this aims to investigate and present the performance of various machine learning and deep learning techniques that are very much useful in voice pathology detection for discriminating healthy voice from pathological voice in people who suffer from Parkinson's disease. The purpose of this analysis is to focus on various machine learning methodologies and techniques used in feature selection and extraction and how acoustic signals are processed and classified. Also, this works aims to provide information about various sources of datasets available for voice disorder along with the summary of reports given by many researchers for the prediction of Parkinson's disease using voice measures.

Keywords Dysphonia · UPDRS (Unified Parkinson's Disease Rating Scale) · Sequential minimal optimization · Principal component analysis · Mel frequency cepstrum coefficient · Minimum redundancy maximum relevance · Mini batch gradient descent

1 Introduction

In the current decade with so many advancements in clinical research, most of the research works lack in focusing on aging diseases. One such aging disease is Parkinson's disease which is a neurodegenerative disorder of the brain. There are

S. Sharanyaa (✉) · P. N. Renjith · K. Ramesh
Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_4

no significant tests to diagnose the disease. Only clinical criteria is to be considered [1, 15]. This is the second common neurological disorder after Alzheimer's disease in the world [16]. Loss or falling of dopamine levels (dopaminergic neurons) in the substantia nigra located in the midbrain causes lack of control and coordination in muscle movements [6] shown in Fig. 1. Nearly one million Americans were affected by the disease [17]. The occurrence of PD is expected to have a drastic rise in the next 20 years. Age is considered to be a potential risk factor in PD with average age between 50 and 60 years with other risks such as family history and environmental toxins. Parkinson's disease is associated with both motor symptoms and non-motor symptoms. Almost 90% of the patients will experience non-motor symptoms that include anxiety, depression, cognitive dysfunction, insomnia (sleep problems), and so on [18]. The classic motor symptoms include resting tremor, bradykinesia (slow movements), rigidity, shuffling gait, and postural instability. Apart from this, people may also suffer from other motor symptoms like abnormal facial expression (hypomania), blurred vision, speech disorder, and inability to move (freezing) [19].

Compared to other PD symptoms, several researches have indicated that vocal impairments are very common and an early identification for Parkinson's disease since 90% of the PWP are being affected with vocal disorders as shown in Fig. 2 [21, 42]. This abnormal functioning of voice is represented by dysphonia. Voice quality is affected, and there exists hoarseness in voice along with other symptoms like roughness and breathiness, reduced energy, and vocal tremor.

Vocal disorders do not appear suddenly. Various vocal tests are performed on PD patients like sustained phonation [22] running speech tests to assess the vocal impairments in people. Various acoustic features are obtained and evaluated to find the healthy state of voice. Traditional voice attributes such as F_0 (fundamental frequency), jitter (difference in frequency), shimmer (difference in amplitude), noise-to-harmonic ratio (amplitude of noise) are considered as some of the essential features [1, 22]. But the accuracy of these attributes in the detection of voice disorders is most likely related to the selection of algorithms. This is the reason why many researchers

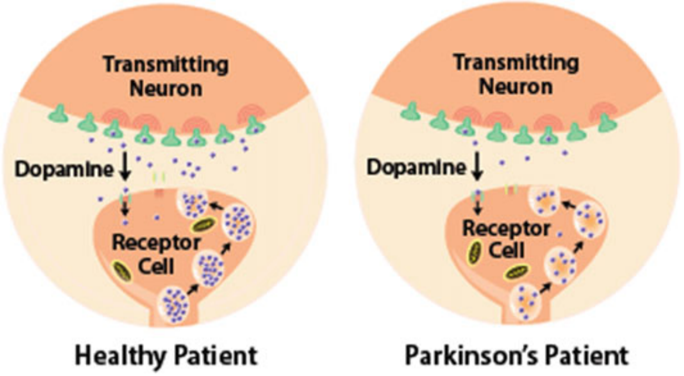


Fig. 1 Dopamine secretion of healthy control and Parkinson's patient

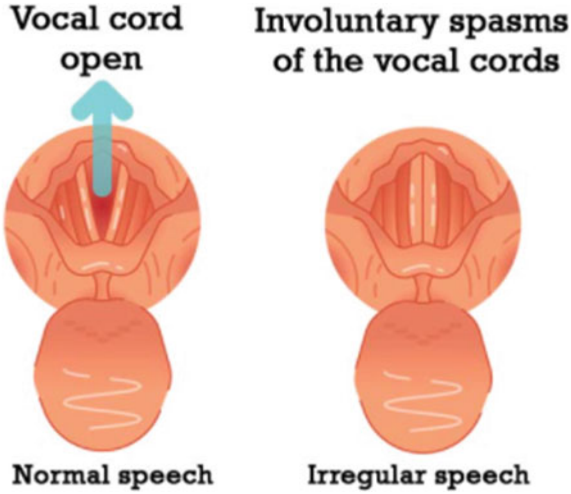


Fig. 2 Regularity and irregularity in speech tough vocal cord [42]

focus on the study of acoustic parameters and is most likely related to the selection of algorithms. This is the reason why many researchers focus on the study of acoustic parameters and applying various classification techniques to get high accuracy in discriminating things [5]. They are the result of slow degradation whose beginning stage cannot be easily noticed [6]. On considering several recent researches toward neurodegenerative disease, machine learning plays a vital role in early diagnosis of diseases. Depending on the dataset types and nature of the problem, various ML algorithms can be used that improves accuracy and lowers the computation time in various predictor models [7].

1.1 Data Collection

Data collection is an important step and root source for data analysis which can be used in detecting Parkinson's disease using voice feature. Voice recordings of patients are collected by clinicians and speech therapists at their clinical or pathological speech centers and the data are made available to public as benchmarking dataset for research purpose or kept private otherwise with confidentiality. Authors can use data—both public dataset and generate private datasets. Some of the publicly available datasets are from UCI Machine Learning Repository, Public Dataset Saarbrucken Voice Database, Mpower Dataset, and Patient Voice Analysis Dataset, and so on. Some of the dataset sources are given below (Table 1).

Table 1 Voice features dataset and its available sources

Source title	Available source	Merits	Demerits
Tsanas, Athanasios, Max Little, Patrick McSharry, and Lorraine Ramig. "Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests." Nature Precedings (2009): 1–1	https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring	More number of voice samples; also provides UPDRS score of each sample	The recordings were automatically captured in the patient's homes. so noisy data
Little, Max, Patrick McSharry, Stephen Roberts, Declan Costello, and Irene Moroz. "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection." Nature Precedings (2007): 1–1	https://archive.ics.uci.edu/ml/datasets/Parkinsons	Traditional acoustic features were extracted	Very few voice samples (195 instances)
Sakar, Betul Erdogdu, M. Erdem Isenkul, C. Okan Sakar, Ahmet Sertbas, Fikret Gurgen, Sakir Delil, Hulya Apaydin, and Olcay Kursun. "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings." IEEE Journal of Biomedical and Health Informatics 17, no. 4 (2013): 828–834	https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings	Variety of voice samples like vowels, numbers, short sentences are available	Less number of samples
Barry, B. Saarbruecken Voice Database. Institute of Phonetics; Saarland University. Available online: http://stimmdb.coli.uni-saarland.de/ (accessed on 23 February 2017)	http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4	Raw voice samples are available and easy access	Fine tuning of Parkinson's voice data is difficult

1.2 Attributes (Voice Features)

According to Sarkar et al. [7], the fundamental attributes for detecting Parkinson's disease based on voice pathology are given in Table 2, and these are the common voice features considered for analyzing and classifying Parkinson's disease.

2 Literature Survey

In the current decade with so many advancements in clinical research, most of the research works lack in focusing on aging diseases.

Little et al. [1] used existing traditional voice measures together with nonstandard measures to classify healthy people from people with Parkinson's disease by detecting dysphonia voice in them. Dataset contains 195 recordings of 31 subjects, in which 23 were patients with Parkinson's disease, and 8 belong to healthy subjects. On an average, six phonations for each subject were taken. Sustained phonation sounds were recorded wherein traditional attributes such as F_0 (fundamental frequency), jitter, shimmer, and noise-to-harmonic ratio were used for analysis. Also, the authors have introduced a new feature called pitch period entropy (PPE)—a robust measure used for diagnosing the disease. Correlation dimension (D2), recurrence period density entropy (RPDE) and de-trended fluctuation analysis [23] were used for calculating nonstandard measures. Linear discriminant analysis, a dimensionality reduction technique, is used to reduce the set of measures. The authors have done research on combining the traditional voice measures with nontraditional measures, henceforth, providing better classifier accuracy of 91.4% using kernel support vector machines. Also, it is found that harmonics-to-noise ratio (HNR) combined with nonstandard measures improves the performance specifically when SVM classifier

Table 2 Voice features of dysphonia table

Frequency-Jitter	Amplitude-Shimmer	Voice tone and others
Jitter (%)	Shimmer	Noise-to-harmonics ratio (NHR)
Jitter (Abs)	Shimmer (dB)	Harmonics-to-noise ratio
Jitter: RAP	Shimmer: APQ3	Recurrence period density entropy (RPDE)
Jitter: PPQ5-	Shimmer: APQ5	De-trended fluctuation analysis (DFA)
Jitter: DDP -	Shimmer: APQ11	Pitch period entropy (PPE)

is applied to combinations of selected voice measures such as harmonics-to-noise ratio (HNR), RPDE, DFA, and PPE.

Sakar et al. [2] collected and analyzed voice samples of people affected with Parkinson's disease because of increased interest in voice pattern analysis among researchers. These voice recordings have samples of sustained vowels, words, and sentences. The authors found that sustained vowels provide more discriminating capacity of PD patients from healthy ones using various machine learning techniques. Data collected for this study has 26 recordings of every subject which belong to 20 PWP (14 male, 6 female), and 20 healthy ones (10 male, 10 female) who visited the neurology department in the Cerrahpasa faculty of medicine in Istanbul university. Also, an independent testing dataset of 168 voice recordings which belong to 28 PD patients were collected and used for validation. With this, clinicians also examined the patients individually and determined their UPDRS and Hoehn and Yarn Scale. Praat Acoustic Software is used [29] for extracting 26 linear and time frequency-based features. Classification is done with two methods. First method is leave one subject out (LOSO) validation where voice recordings of one subject are left out and used for validation. The second method is the summarized leave one out (SLOO) method, in which features are summarized using six metrics based on central tendency and dispersion metrics. SVM and K-NN classifiers were used for PD diagnosis. The authors found that SVM classifier provides higher accuracy of 77.5% using SLOO method compared to K-NN classifier which is 65.0%.

According to Yang et al. [3], to analyze the voice attributes of sustained vowels, the authors have used statistical pattern analysis, in which two techniques—sequential forward selection (SFS) and kernel principal component analysis are used as feature selection and dimensionality reduction techniques. Dataset is taken from [23] which consists of 195 voice recordings of sustained vowels of 31 subjects available in UCI Machine Learning Repository [35]. Fisher's linear discrimination analysis [FLDA] is a nonlinear classification technique used for classifying healthy subjects from patients with Parkinson's disease. Maximum a posterior (MAP) decision tree and support vector machine have been used had nonlinear classifier techniques. MAP classifier with KPCA feature set gives 91.8% accuracy on voice records with sensitivity rate and specificity rate of 0.986 and 0.708. The authors also have concluded that, on detection of dysphonia, gender is not considered as a sensitive criteria on voice recordings.

In the article of Rachel et al. [4], the voice features of dysphonia and healthy subjects are extracted using PRAAT software [29] and MATLAB software. Dataset contains speech therapy samples of 25 dysphonic patients and 25 healthy people in the speech pathology department at SRM Hospital and Research Center, India. An analysis is made on both feature extraction methods of speech signal and found MATLAB feature extraction gives better significance value. Thirteen features were extracted using PRAAT software, and the methods extracted from the speech signal and 20 features were obtained using MATLAB software. Thus, the features extracted from the MATLAB software provided better significance between the normal and dysphonic patients for most of the acoustic features compared to PRAAT software and as a result, significance value increases in dysphonic patients compared to healthy subjects. Analysis on data is done using SPSS software version 19.0.

Verde et al. [5] have used mobile health systems for detecting voice pathology. The authors analyze and compare various machine learning algorithms that are used for detecting the voice pathology in people. Datasets used for all analysis were taken from Saarbrucken Voice Database [24]. This dataset contains totally 2041 voice recordings of vowels /a/, /o/, /u/ and some sentences. But the authors have selected and used only balanced records of 1370 vocal subjects in which 685 pathological voices (257 male and 428 female) and 685 healthy voices (257 male and 428 female) for their experiments. The results are evaluated based on accuracy, sensitivity, specificity, and ROC area. The authors found SVM and decision tree provides better accuracy compared to Bayesian classification (BC), logistic model tree (LMT), K-nearest neighbor (K-NN) which are based on features selected using opportune feature selection methods. The analysis is made using Weka tool. Features considered are fundamental frequency F0, jitter, shimmer, HNR, MFCC (from 1 to 13), and first derivative and second derivative cepstral coefficient. All these parameters were calculated on each recording pronouncing the vowel /a/; feature selection methods used are Info Gain Attribute Evaluation algorithm, principal components analysis (PCA), and correlation method feature selection. Comparison is made with features extracted using all three methods individually with each classifier model and found SVM classifier with sequential minimal optimization (SMO) optimization algorithm gives higher accuracy of 85.77%. When feature selection is taken into consideration, the features selected with the Info Gain Attribute Evaluation method and PCA gives higher classification accuracy using the SMO technique (84.16 and 71.75%). On considering correlation method, decision tree algorithm provides better accuracy (84%).

Sakar et al. [7] have used two-step approaches in analyzing the vocal features to diagnose the Parkinson's disease in the early stage. Machine learning algorithms have been used for evaluation. In the first step, using the UPDRS Score, patients are grouped based on having high severity of speech pathology [26]. For this, supervised and unsupervised machine learning techniques are used. In the second step, the authors used binary classification method to obtain the optimal threshold value (obtained values is 15) and subjects whose UPDRS Score is greater than the threshold are considered severely affected and are omitted. Subjects with lesser threshold values are compared with healthy subjects for early identification of Parkinson's disease. Different datasets were used for analysis. For first step, dataset is taken from [27] and available in UCI machine learning Repository [37] which contains speech samples of 42 patients along with UPDRS score. Sixteen features were extracted from this data. For the second step, dataset for healthy groups were taken from Little et al. [1] contains 48 samples of 8 healthy subjects for comparison. Principal Component analysis and spectral clustering is used to validate the threshold value. SVM, ELM, and K-NN Classifiers are used for analysis. SVM performs better compared to K-NN and extreme learning algorithm, which provides highest accuracy of 96.4% with Mathews correlation coefficient values as 0.77.

The paper of Xu, Zhijing et al. [8] focusses on diagnosing Parkinson's disease based on voice print features of the Parkinson's patients. The classification is done by distinguishing healthy controls from people affected with Parkinson's disease. The authors propose a method that combines deep neural network (DNN) classifier with mini batch gradient descent (MBGD) optimization algorithm which is the combination of gradient descent algorithm and stochastic gradient descent algorithm. In DNN, for training the parameters, the back propagation algorithm is used to find the gradient (slope) of every layer. Feature extraction being an important factor, weighted mel-frequency cepstrum coefficient (WMFCC) is used on the voice dataset. In the audio retrieval process, WMMFC is a traditional feature where the frequency is equally placed based on Mel Scale. Generally, effective feature selection is affected due to smaller coefficient values of features. Hence, WMFCC uses entropy methods to calculate the weights of independent variables by taking the mean value [28]. This overcomes the problems of having a lesser cepstrum coefficient. Majority of this work is focused on the feature extraction process. The dataset taken from [2] contains the recordings of sustained vowels /a/, /o/ and /u/ from 20 healthy controls and 20 PD, and totally 120 voice samples were used for training the model. Using the same recording device, voice recordings of 28 PD patients were collected for testing the model. Compared to support vector machine (SVM) classifier, the usage of DNN classifier gives better classification accuracy (89.5%) when used with WMFCC method that solves the problem of smaller values of higher order spectrum coefficients. Also, MBGD provides minimal computational loads that results in an increase of classifier training speed because MBGD splits the training dataset into small subsets which are used to compute the model error. Finally, model evaluation factors such as accuracy, sensitivity, specificity, Matthews correlation coefficients, and probability excess were computed on each speech samples /a/, /o/, /u/. Though the proposed method is better than traditional classification algorithms in providing higher accuracy, the relationship between samples feature selection, classification, and dataset size can be improved.

Zuzuna et al. [9] had focused their research toward dysphonia pathological speech using various machine learning classification techniques such as K-nearest neighbor, support vector machine, and random forest. Voice recordings of people pronouncing vowels /a/, /i/, /u/ were taken from Saarbrucken database which contains 194 voice samples of which 94 samples of dysphonia and 100 are healthy subjects [30]. 300 features were selected using simple filter FS from original 1560 features. For dimensionality reduction, principal component analysis (PCA)—a statistical procedure is used to build new feature matrix [26]. The highest accuracy of 91.3% is obtained to identify pathological speech. It is found that, on considering the classification performance of all features, reduced number of features after feature selection shows improvement in obtaining accuracy. The authors have also made analysis on comparing male samples and female samples separately.

The article of Wroge et al. [10] finds out the effectiveness of using supervised machine learning classifier models including deep neural networks to diagnose Parkinson's disease accurately. mPower Voice dataset is used for the analysis [31]. This dataset contains voice recording of patients saying /aa/ phonation for about

10 s [14]. Voice samples are preprocessed using Python's Pyaudio analysis library which has given 11 unique features. Voice box's voice activation algorithm is used to remove background noise present in raw voice data. This cleaned data is fed into two feature extraction methods. One is AVEC2013 (Audio Visual Emotion Recognition Challenge) which applies mRmR (Minimum Redundancy maximum Relevance) technique to extracts 1200 features [32]. The other one is GeMaps (Geneva minimality acoustic parameter set) algorithm in which Open smile toolkit is used that extracts 62 features [33]. Various classifiers such as standard decision tree, gradient boosted decision tree, random forest, support vector machine, and deep artificial neural networks are used. Both features' sets are given as input to all the classifier models and performance of each model is analyzed. It is found that the AVEC feature set performs better compared to GeMaps feature set. The highest accuracy of 86% is obtained using SVM and gradient boosting decision Tree on AVEC feature set and GeMaps feature set provides comparatively lesser performance.

Most recently, Diogo et al. [11] proposed a methodology for early detection of Parkinson's disease using voice features with signal processing techniques and machine learning techniques. For better optimized results, grid search methods and learning curve are used to evaluate the learning performance of the model. The authors have classified the voice samples that are recorded in a natural environment where noise may exist. Datasets from three different speech databases were used for experimenting the results. Two datasets contain recording samples in PCM WAV format in which the first dataset contains 1002 speech lines of 22 PD totally, and the second dataset contains 30 healthy subjects having 785 speech lines for training purpose. For testing, UCI Machine Learning Repository collected by Sakar et al. [2] has been used from 28 PD. A total of 19 features were extracted using Praat Software. The three classification models random forest, support vector machine, and neural network are optimized using learning curves and statistical analysis, and the best optimal classifier is chosen for classification. Leave-one-out cross-validation technique is used to obtain the accuracy of each classifier, and random forest (2 trees) provides higher accuracy compared to other classifiers like SVM, perceptron, and neural networks.

Most recently, in Nilashi et al. [12] article, since UPDRS Score is commonly used by clinicians to understand the effect of Parkinson's disease, the authors have developed a new hybrid technique to predict total and motor UPDRS. Dataset is taken from UCI Machine learning Repository [34] which contains 200 voices of each subject from 28 men and 14 women [27]. Dataset is grouped based on similarity using self-organizing maps (SOM) and expectation maximization (EM). Cluster size is between 2 and 9. Singular value decomposition is used to reduce the feature dimensions from each cluster. Ensemble of adaptive neuro-fuzzy inference systems model (ANFIS) is used which uses fuzzy logic and neural network techniques to learn and to analyze the performance where data that belong to each cluster is fed into the model. The authors have concluded that the use of SVD (single value decomposition) reduces the computation time of predicting Parkinson's disease.

According to Berus et al. [13], the authors focus on prediction of Parkinson's disease using 26 different voice recordings of each individual using multiple feed forward ANN classifiers. Dataset is taken from UCI Machine Learning Repository with 20 healthy subjects and 20 PD subjects [2, 35]. Pearson's correlation coefficient, Kendall's correlation coefficient are the feature selection techniques used to measure the correlation between variables. Along with this, principal component analysis, a standard measure of dimensionality reduction with self-organizing maps of single-layer visualization tools are used as feature selection methods. Seventeen features were extracted. Result validation is done using leaving one subject out (cross-validation scheme). Highest accuracy is obtained using feature selection based on Kendall's correlation coefficient scheme. But multiple ANN is considered to be the best classifier to diagnose PD using raw voice samples without using feature selection. Multiple ANN classifier provides 86.47% test accuracy via LOSO cross-validation method. The authors concluded that sustained vowel phonation provides lesser information to diagnose Parkinson's disease than more different sound recordings like sentences, numbers, and so on. Other language speech recordings can also be taken for research to diagnose the disease more accurately. Comparative analysis of all these techniques is given in Table 3.

3 Discussion

Due to increase in interest toward voice processing and speech analysis applied to Parkinson's disease diagnosis, studies are made on wide collection of articles which shows the importance of the application in the field of disease diagnosis. From the study, it is understood that gender is not a sensitive thing to be given importance for voice and phonation disorder, but smallest functional disability should be considered. From the comparison study given it is found that, there is no improved trade-off between computation time of processing the data and its performance. Since many of the works have used machine learning algorithms, processing large data is also considered as a challenge.

3.1 Unified Parkinson's Disease Rating Scale

UPDRS is the most commonly used comprehensive tool for tracking the severity and symptom progression in Parkinson's disease [20]. The MDS-UPDRS has four parts. Part I is the non-motor symptoms of daily living, Part II comprises motor symptoms of daily living, whereas Part III considers motor examination, and Part IV represents motor complications [7, 20]. The movement disorder society also uses Hoehn and Yahr scale (HY), another performance scaling factor to measure the disease progression with 5-point scaling score (1–5). It is found that, many articles have not used Unified Parkinson's Disease Rating Scale (UPDRS) score as an important comprehensive tool for prediction. This work provides us the direction to focus on improving the classification and prediction accuracy for voice data.

Table 3 Comparative analysis summary

Ref. No.	Dataset used	Feature extraction and feature selection	Classifier	Performance measure
[1]	Clinical dataset 195 samples of 31 subjects (23 PD and 8 Healthy) [14]	Linear discriminant analysis	Support vector machines	Performance (SVM): 91.4%
[2]	Clinical Dataset: 120 samples (20 PD and 20 healthy) Collected at neurology department at Istanbul University	Praat Software, leave one subject out (LOSO) and summarized leave one out (SLOO)	SVM and K-NN	SVM Provides 77.5% accuracy with SLOO method
[3]	UCI Machine Learning Repository: 195 samples of 31 subjects (23 PD and 8 healthy) [14]	Sequential forward selection Kernel principal component analysis	Maximum a posterior (MAP), support vector machine	Accuracy of MAP with KPCA: 91.8%
[4]	Clinical dataset collected at SRM Hospital and Research Center, India (25 Dysphonic and 25 healthy)	PRAAT software and MATLAB software	SPSS software Version 19.0	Significance value is high using MATLAB
[5]	Public Dataset Saarbrucken Voice Database 685 pathological voices (257 male and 428 female) 685 healthy voices (257 male and 428 female) [16]	InfoGain Attribute Eval algorithm, principal component analysis (PCA), correlation method	Sequential minimal optimization (SMO) optimization algorithm used with SVM Classifier, decision tree, Bayesian classification, logistic model tree, K-nearest neighbor	SVM with SMO provides high accuracy 85.77%
[7]	Public dataset UCI Machine Learning Repository (42 PD subjects—5875 voice recordings.) [34]	Principal component analysis and clustering analysis	Binary classification method, SVM, K-NN and extreme learning method	SVM: 96.4% accuracy and MCC 0.77

(continued)

Table 3 (continued)

Ref. No.	Dataset used	Feature extraction and feature selection	Classifier	Performance measure
[8]	Public dataset UCI Machine learning: 20 samples (20 PD and 20 healthy) [2]	Weighted mel-frequency cepstrum coefficient (WMFCC)	DNN back propagation algorithm and mini batch gradient descent (MBGD) optimization algorithm	89.5%
[9]	Public dataset Saarbrucken database 194 voice samples (94 dysphonia and 100 healthy) [24]	Simple filter FS and PCA	K-Nearest Neighbor, support vector machine and random forest	Accuracy: 91.3%
[10]	mPower Dataset [31]	Voice activation detection algorithm, AVEC (MIMR), and GeMaps	Standard decision tree, gradient boosted decision tree, random forest, support vector machine, and deep artificial neural networks	High accuracy (SVM and gradient boosted decision tree)—86% on AVEC feature set
[11]	Clinical dataset (Portuguese recordings) for training and Public dataset from UCI Machine Learning Repository for Testing [2]	Principal component analysis	SVM, RF, NN, naive Bayes	High accuracy (SVM)—94.7%
[12]	Public dataset UCI Machine Learning Repository (42 PD subjects—5875 voice recordings) [34]	Singular value decomposition, self-organizing map, expectation maximization	Adaptive neuro-fuzzy inference systems model (ANFIS)	EM-SVD-ANFIS computation time (ms) Motor-UPDRS-16,443 Total-UPDRS-15,251
[13]	UCI Machine Learning Repository (20 PD and 20 Healthy) [2]	Pearson's correlation coefficient, Kendall's correlation coefficient, PCA, self-organizing maps	Multiple ANN classifier	Accuracy using ANN: 86.47%

(continued)

Table 3 (continued)

Ref. No.	Dataset used	Feature extraction and feature selection	Classifier	Performance measure
[14]	mPower Dataset (Private) (921 subjects with Depression) [31]	Open smile app is used for feature extraction Feature set: (i) AVEC 2013(32 features) (ii) GeMAPS(62 features) MRMR Algorithm for both	SVM, random forest, feed forward deep neural network models	Overall accuracy 77.0% Severity PD and depression correlation coefficient 0.3936
[39]	UCI Machine Learning Repository- 195 samples of 31 subjects (23 PD and 8 healthy) [14]	Feature importance and recursive feature elimination methods	CART, ANN, SVM	SVM with RFE produces 93.84% accuracy
[40]	UCI Machine Learning Repository—195 samples of 31 subjects (23 PD and 8 healthy) [14]	Multiple feature evaluation approach (MFEA)	Naïve Bayes, neural network, random forests, and support vector machine	Decision tree 10.51%, naïve Bayes 15.22%, neural network 9.19%, random forests 12.75%, and SVM 9.13% improvement
[41]	25 Parkinson, 20 healthy (6 recordings each) [7]	Empirical mode decomposition, intrinsic mode function cepstral coefficient (IMFCC)	SVM	10–20% higher accuracy compared to MFCC

3.2 Machine Learning Techniques

Some of the machine learning models does not provide better classification accuracy because the model does not generate the signal samplings. Differential phonological posterior features improve the performance of the speech reading task and achieve better results. On the other hand, it is restricted due to the limited set of available data. Because of this, in many articles, lack of performance using existing machine learning algorithms which in turn can give better results using deep neural network.

4 Conclusion

The review presented in this article provides variety of voice samples from various datasets both private collection of data and publicly available datasets including vowels, sustained phonation's, sentences, numbers, and so on with various assessment tests. Dominant fundamental features such as jitter, shimmer, HNR, NHR, PPE, and DFA are extracted using multiple voice processing software like Praat, OpenSmile, MDVP, and so on. The experimental results of various authors show that, techniques used in feature extraction and feature selection to select appropriate features have high impact in obtaining good accuracy of classifying the PD disease. Also, choosing the right datasets is a challenging task, and this work has given various sources of dataset that are available. Analysis is made on various state-of-the-art machine learning classifiers and how they are used with feature selection algorithms that provide an extensive work in related area. Future work can be carried out with large number of raw voice recordings of Parkinson's patients. Large number of research work has been carried out with machine learning techniques. Use of deep neural network for raw voice feature extraction and training the classification model in future will be more effective in accuracy and other performance metrics.

References

1. M. Little, P. Mcsharry, E. Hunter, J. Spielman, L. Ramig. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nat. Precedings 1–1 (2008)
2. B. Sakar, et al., Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J. Biomed. Health Inform. **17**(4), 828–8341 (2013)
3. S. Yang, F. Zheng, X. Luo, S. Cai, Y. Wu, K. Liu, M. Wu, J. Chen, S. Krishnan,Effective dysphonia detection using feature dimension reduction and kernel density estimation for patients with Parkinson's disease. PloS One **9**(2) (2014)
4. Rachel, S. Shamila, U. Snekhalatha, D. Balakrishnan, Spectral analysis and feature extraction of speech signal in dysphonia patients. Int. J. Pure Appl. Math. **113**, 151–160 (2017)
5. L. Verde, G. De Pietro, G. Sannino, Voice disorder identification by using machine learning techniques. IEEE Access **6**, 16246–16255 (2018)
6. https://www.parkinsonassociation.org/what-is-parkinsons-disease/
7. E. Sakar, G.S. Betul, C. Okan Sakar, Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. PloS One **12**(8), e0182428 (2017)
8. Z. Xu, et al.,Voiceprint recognition of Parkinson patients based on deep learning. arxiv preprint arxiv:1812.06613 (2018)
9. Z. Dankovičová, Dávidsovák, P. Drotár, L. Vokorokos, Machine learning approach to dysphonia detection. Appl. Sci. **8**(10), 1927 (2018)
10. T.J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D.C. Atkins, R.H. Ghomi,Parkinson's disease diagnosis using machine learning and voice, in *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (IEEE, 2018), pp. 1–7
11. D. Braga, et al., Automatic detection of Parkinson's disease based on acoustic analysis of speech. Eng. Appl. Artif. Intell. **77**, 148–158 (2019) (Elsevier)
12. M. Nilashi, O. Ibrahim, S. Samad, H. Ahmadi, L. Shahmoradi, E. Akbari, An analytical method for measuring the Parkinson's disease progression: a case on a Parkinson's telemonitoring dataset. Measurement **136**, 545–557 (2019)

13. L. Berus, S. Klancnik, M. Brezocnik, M. Ficko, Classifying Parkinson's disease based on acoustic measures using artificial neural networks. Sensors **19**(1), 16 (2019)
14. Y. Ozkanca, et al., Depression screening from voice samples of patients affected by Parkinson's disease. Dig. Biomarkers **3**(2), 72–82 (2019)
15. L.A. Uebelacker, G. Epstein-Lubow, T. Lewis, M.K. Broughton, J.H. Friedman, A survey of Parkinson's disease patients: most bothersome symptoms and coping preferences. J. Parkinsons Dis. **4**(4), 717–723 (2014)
16. B.K. Varghese, D. Amali, K.S. Devi, Prediction of Parkinson's disease using machine learning techniques on speech dataset. Res. J. Pharm. Technol. **12**(2), 644–648 (2019)
17. J.A. Obeso, C. Warren Olanow, J.G. Nutt, Levodopa motor complications in Parkinson's disease. Trends Neurosci. **23**, S2–S7 (2000)
18. S.B. O'Sullivan, T.J. Schmitz, G. Fulk, Physical rehabilitation. FA Davis (2019)
19. J.M. Beitz, Parkinson's disease: a review. Frontiers Biosci. **S6**, 65–74 (2014)
20. S.K. Holden, T. Finseth, S.H. Sillau, B.D. Berman, Progression of MDS-UPDRS scores over five years in de novo Parkinson disease from the Parkinson's progression markers initiative cohort. Movement Disorders Clin. Prac. **5**(1), 47–53 (2018)
21. A.K. Ho, R. Iansek, C. Marigliani, J.L. Bradshaw, S. Gates, Speech impairment in a large sample of patients with Parkinson's disease. Behav. Neurol. **11**(3), 131–137 (1998)
22. R.J. Baken, R.F. Orlikoff, Clinical measurement of speech and voice: Cengage learning. Google Scholar (2000)
23. M. Little, P. McSharry, S. Roberts, D. Costello, I. Moroz, Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Nat. Precedings 1–1 (2007)
24. D. Martínez, E. Lleida, A. Ortega, A. Miguel, J. Villalba, Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit, in *Advances in Speech and Language Technologies for Iberian Languages* (Springer, Berlin, Heidelberg, 2012), pp. 99–109
25. L. JabaSheela, S. Vasudevan, V.R. Yazhini,A hybrid model for detecting linguistic cues in Alzheimer's disease patients. J. Inform. Comput. Sci. **10**(1), 85–90 (2020)
26. B.E. Sakar, C. Okan Sakar, G. Serbes, O. Kursun, Determination of the optimal threshold value that can be discriminated by dysphonia measurements for unified Parkinson's Disease rating scale, in *2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)* (IEEE, 2015), pp. 1–4
27. A. Tsanas, M. Little, P. McSharry, L. Ramig, Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. Nat. Precedings 1–1 (2009)
28. E.S. Wahyuni, Arabic speech recognition using MFCC feature extraction and ANN classification. in *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)* (IEEE, 2017), pp. 22–25
29. P. Boersma, D. Weenink, Praat: doing phonetics by computer (2012) [Online]. Available http://www.praat.org/. Accessed 29 March 2012
30. B. Barry, Saarbruecken voice database. Institute of Phonetics; Saarland University. Available online http://stimmdb.coli.uni-saarland.de/. Accessed on 23 February 2017
31. B.M. Bot, C. Suver, E. Chaibubneto, M. Kellen, A. Klein, C. Bare, M. Doerr, et al., The mpower study, Parkinson disease mobile data collected using researchkit. Sci. Data **3**(1), 1–9 (2016)
32. M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, M. Pantic, AVEC 2013: the continuous audio/visual emotion and depression recognition challenge, in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge* (2013), pp. 3–10
33. F. Eyben, K.R. Scherer, B.W. Schuller, J. Sundberg, E. André, C. Busso, L.Y. Devillers, et al., The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Trans. Affect. Comput. **7**(2), 190–202 (2015)
34. D.J. Newman, S. Hettich, C.L. Blake, C.J. Merz, Repository of machine learning databases (1998)
35. D. Dua, E. Karra Taniskidou, UCI machine learning repository [http://archive.ics.uci.edu/ml]. (University of California, School of Information and Computer Science, Irvine, CA, 2017)

36. C.G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G.T. Stebbins, C. Counsell, N. Giladi, et al., Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations the movement disorder society task force on rating scales for Parkinson's disease. Movement Disorders **19**(9), 1020–1028 (2004)
37. Dataset: http://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring
38. S. Sharanyaa, P.N. Renjith, K. Ramesh,Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques, in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)* (IEEE, 2020)
39. Z.K. Senturk, Early diagnosis of Parkinson's disease using machine learning algorithms. Med. Hypotheses **138** (2020)
40. S.A. Mostafa, A. Mustapha, M.A. Mohammed, R.I. Hamed, N. Arunkumar, M.K. Abd Ghani, M.M. Jaber, S.H. Khaleefah, Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. Cogn. Syst. Res. **54**, 90–99 (2019)
41. B. Karan, S.S. Sahu, K. Mahto, Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybern. Biomed. Eng. **40**(1), 249–264 (2020)
42. https://www.medindia.net/patientinfo/spasmodic-or-laryngeal-dysphonia.html

Intelligent Recommender System Based on Deep Learning for Recognition of COVID-19 from Lungs Radiograms

Manish Assudani and Neeraj Sahu

Abstract A new coronavirus outbreak event emerged as a worldwide public health pandemic. Investigation of huge numbers of people is the current need to terminate the roll out of infection within the society. Polymerase chain reaction is an ordinary medical technique used for infection examination. There is a need of exploring novel tools for examinations due to growing quantity of fake infection outcome. In the COVID-19 infection staging point, X-rays of the bronchus of coronavirus-affected patients were discovered to be a critical replacement predictor. In this scenario, correctness depends on radiographic information. A clinical recommendation process that supports the physician in analyzing the patients' bronchus photographs will reduce the physician medical workload. In particular, convolutionary neural networks (CNN) deep learning strategies have proven effective in classifying scanning in medications. For the detection of coronavirus infection, four distinct frameworks based on convolutionary neural networks were explored which are applied to chest X-ray photographs. On the ImageNet database, these designs were pretrained, thus minimizing the requirement for massive trained samples since these samples are having already trained magnitudes. It has been noticed that models dependent on CNN consist of ability to diagnose coronavirus infection.

Keywords Convolution neural networks · Recommender system · Deep learning · Lungs X-rays · Deep learning · ImageNet database · Polymerase chain reaction · Coronavirus

1 Introduction

Till last week of month of Jan 2020, the coronavirus infection epidemic that started in China in the month of first week of Dec 2019 had spread exponentially across the globe. It was initially pointed to it as 2019 novel coronavirus and announced it as an

M. Assudani (✉) · N. Sahu
G. H. Raisoni University, Amravati, Maharashtra, India

N. Sahu
e-mail: neeraj.sahu@ghru.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_5

epidemic [1]. The actual figure of known occurrences in the world is estimated to be about 7 million, and the world mortality rate is around 3–4%.

As the 2019 novel coronavirus is extremely infectious and is fast propagating, it is an imperative for governments in almost all affected countries to separate contaminated persons quickly. Flu-like manifestation such as temperature, coughing, dyspnea, respiratory issues, and viral pneumonia is the general symptoms of COVID-19 patients. But these symptoms alone are not substantial. There are several instances where patients are asymptomatic, but they were positive for novel coronavirus contamination after examining their chest X-ray photographs [2]. Therefore, to make an accurate diagnosis, affirmative computed tomography (CT) of thorax need to be examined along with manifestation. Performing at runtime polymerase chain reaction is basically a conventional clinical method [2] for pathological research.

Across the world, health systems are trying for broadening COVID-19 testing provisions. Increasingly, testing can contribute for the detection and separation of contaminated individuals, which results in decrease in group propagate [3]. But reliability is not guaranteed by availability. At this point, the main concern for governments is the false negative diagnostic tests; for the infected person, the testing results are negative [3]. Such persons can spread the virus unintentionally to others. Therefore, incorrect test findings have a detrimental impact on the attempts made to stop the transmission of the virus. It is impossible to determine the impact of this problem on the general citizenry's and healthcare workers' protection because there is no consistent or accurate evidence on these sample functionalities. The responsiveness of these tests is currently unclear [3].

Thorax X-ray photographs of coronavirus-infected persons are demonstrated as an accurate and reliable strategy for diagnosing coronavirus-contaminated patients [4]. But the reliability of X-ray diagnostics is highly dependent on radiological knowledge. And when the percentage of patients is large, it becomes a cumbersome assignment. The diagnostic workload of the doctor can be minimized by a recognition functionality that will help the physician to analyze the patients' bronchus photographs.

Approach which utilizes deep learning is ANNs, where each section consists of neurocyte that is similar with human body's neurocyte. CNN is one of the approaches that has demonstrated successful and accurate in classifying clinical images. There have been many reports that utilize CNN, focused on radiography to detect pneumonia and other infections. In [5], a framework supported on CNN was suggested to classify various lung diseases. The thorax X-ray photographs database contains approximately one lakh photographs which was used in [6] as a training dataset for the CNN framework for detection of fourteen infections. CNN was also utilized to forecast pneumonitis [6, 7]. In [8], a framework was suggested to help radiologists distinguish contaminated areas in CT images.

Therefore, CNN is also a sound strategy for detecting effectively coronavirus-contaminated peoples. In the work [9], fifty photographs are utilized to compare various established neural network frameworks. Of the fifty-sample data, twenty-five were positive for coronavirus and twenty-five patients were fatalistic for coronavirus. Different reform frameworks for CNN [7, 10–12] or calibrated ResNet50 [13] have been suggested by some researchers for the issue of chest X-ray classification. But they are also in the process of preprinting.

1.1 Objective and Motivation

Real-time polymerase chain reaction (RTPCR) is one of the effective techniques for the detection of coronavirus. But over the period it generates false results. So, there is a need for new technology enhancement which can detect coronavirus infection from lung radiograms. The characteristics of the coronavirus are it attacks the lungs and gradually Detroit's respiratory system which leads to the death of patients. So, there is a need of efficient COVID-19 detection technique which accurately classifies unknown patient into coronavirus-infected patient or non-coronavirus-infected person.

1.2 Novel Contribution

We have formulated a novel technique based on deep learning CNN models to effectively detect and classify coronavirus infection. Four different models play a crucial role in detecting coronavirus infection from chest X-ray photographs. The main feature of these models is that it does not require large training data which is one of the important steps in the classification process. These models are already trained on ImageNet database. This will reduce the cost of training and labeling the dataset. Our technique also increases the accuracy of detection as well as reduces the efforts of a physician for detecting the coronavirus.

2 Model Building

A well before trained system is one that was created for resolving a specific problem by someone else and can be used by anyone to solve a similar issue. Numerous CNN models of image classification are well before trained on datasets of images. ImageNet is basically big, precise as well as versatile collection of photographs. The pictures in this database are organized according to the hierarchy of WordNet [14]. There are about one lakh phrases in WordNet, each with an average of about one thousand images. We have utilized four models for our research work, which are described below.

2.1 Model 1: MobileNet

In [15], another CNN system is suggested called as MobileNet. Depthwise, this system is organized in this separable convolution and implemented the convolution operation independently on each color channel rather than just taking them overall. This architecture lowers the cost of computing. The design of MobileNet is based on depthwise separable convolution. Depthwise separable convolution utilizes two different layers: one layer is used for filtering and another layer is used for combining. This architecture will immediately reduce the cost of computation and the size of the model. MobileNet model can be effectively utilized for various tasks like object identification, face recognition, etc. MobileNet is a lightweight deep neural network.

2.2 Model 2: ResNet50

The design of residual networks was suggested in [13]. It consists of fifty convolutional steps with skip ties. These steps support the process of increasing the model's learning accuracy. Instead of completely linked layers, it also uses global average pooling, thus reducing the model scale. ResNet50 is a network containing 50 layers, and these layers are trained on the ImageNet database. The bypass technique is implemented between layers in the ResNet50 model. As the ResNet50 model grows, it becomes complex, but the model is not distorted.

2.3 Model 3: Xception

Xception was created in [16]. This model is an improved version of Inception. This modern and precise model scores on speed and precision. The number of parameters of Xception is the same as the number of parameters in Inception V3. When compared with Inception V3, the performance of Xception is slightly better as compared with Inception V3 when applied on ImageNet database and excellent when JFT database. Characteristics of depthwise separable convolution are utilized in the Xception model.

2.4 Model 4: Inception V3

The framework of Inception was suggested in [17]. GoogLeNet was considered the original design. The following variants are called Inception Vn (n is variant count). One of the important characteristics of the Inception V3 framework is label smoothing. 7*7 convolutions are used in the Inception V3 model. To transfer information related to the label to the lowest part of the model, an auxiliary classifier is utilized. An auxiliary classifier is also used in the batch normalization process. An improved version of this model is the Xception model.

3 Materials and Methodology

3.1 Dataset

An overall number containing 6786 bronchus radiogram photographs has been obtained from public databases available in different repositories of GitHub [18]. Of these, 350 were positive for COVID-19 and 6436 were acceptable victims who are not suffered from coronavirus within them. Photographs are distinguished as non-coronavirus infected and coronavirus infected. Figures 1 and 2 depict photographs classified as non-coronavirus infected and coronavirus infected, respectively. By using 80:20 split, 80 percent of dataset is used for training and 20 percent is used for

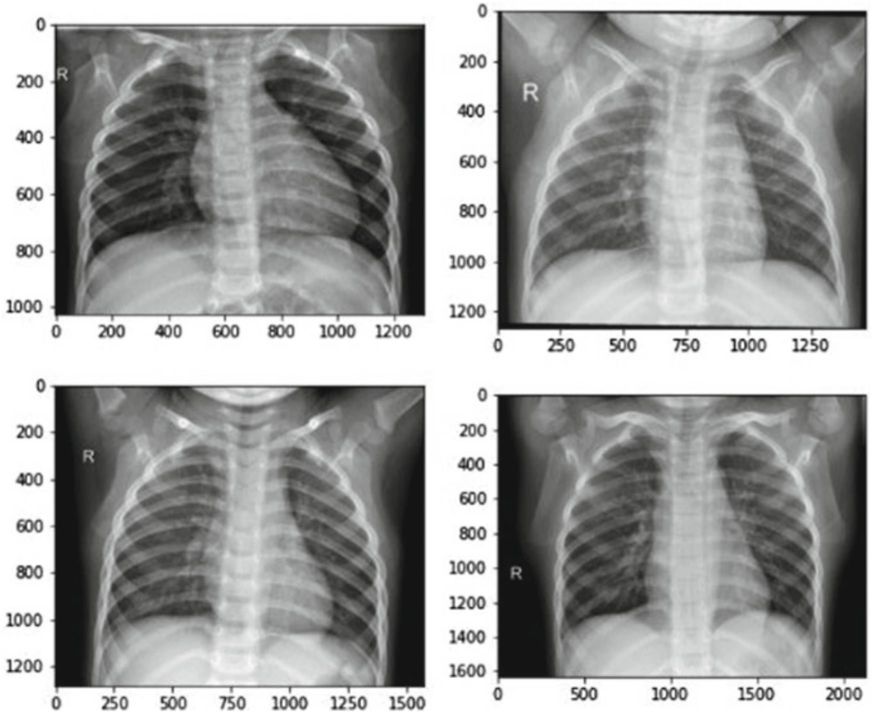


Fig. 1 Specimen photographs for non-COVID-19 bronchus radiogram cases

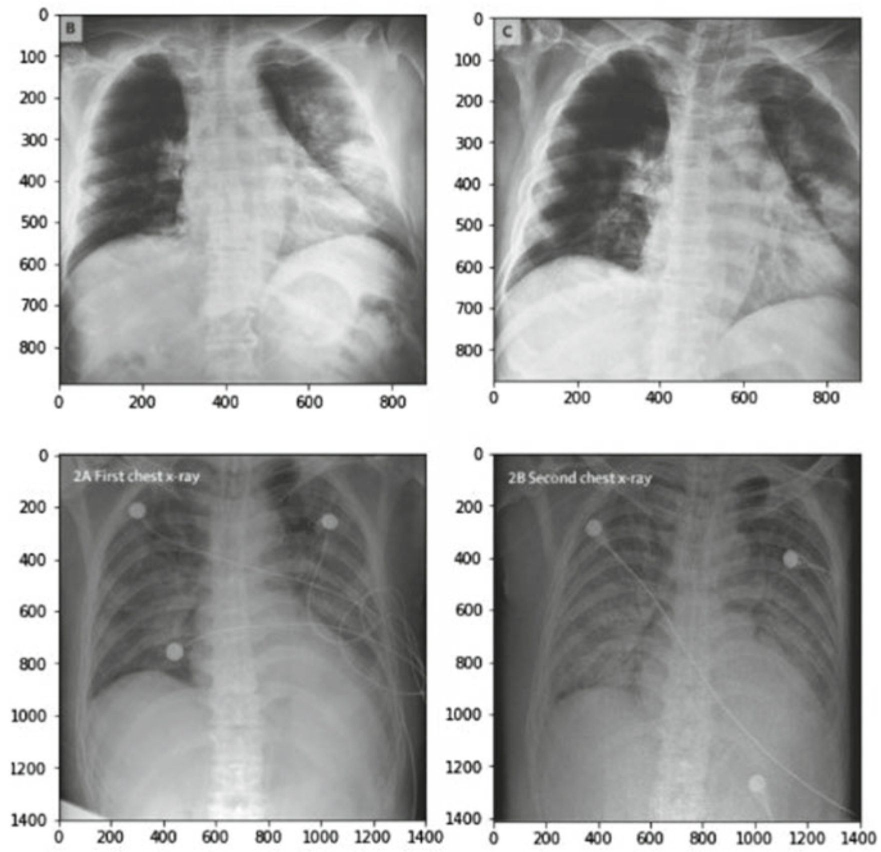


Fig. 2 Specimen photographs for COVID-19 bronchus radiogram cases

testing. 5428 records were used for learning (training), and 1357 records are used for testing purpose. The authenticated collection is kept as 40% of the total training data.

4 Results and Discussions

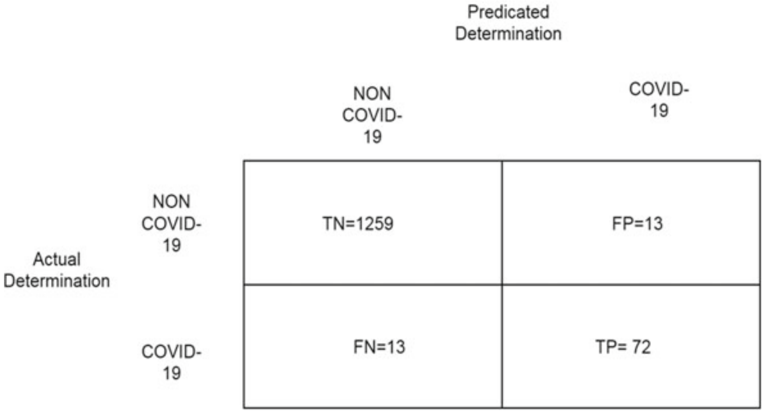
4.1 Experimental Procedure and Tool

The Keras platform with TensorFlow as the runtime was used for the accomplishment of the models mentioned in Sect. 2. Keras introduces well before trained system versions acquired from the ImageNet repository. Although the photographs extracted from the ImageNet repository from which these models are trained are not similar to the photographs acquired for investigation, they will be useful to make the intended task more effective by translating information gained. Pretrained weights often minimize the need for a significant amount of training results.

The Python code was executed by Google Research's online Jupyter Notebook-based Service Collaborative. For quicker processing, which is given by Collaboratory, the Tesla P4 GPU was utilized. To train all the networks, Adadelta Optimizer was utilized, and mean-squared error (MSE) is treated as the forfeiture assignment. The group capacity was set at thirty-two for the training, and the total quantity of period is calibrated at two hundred.

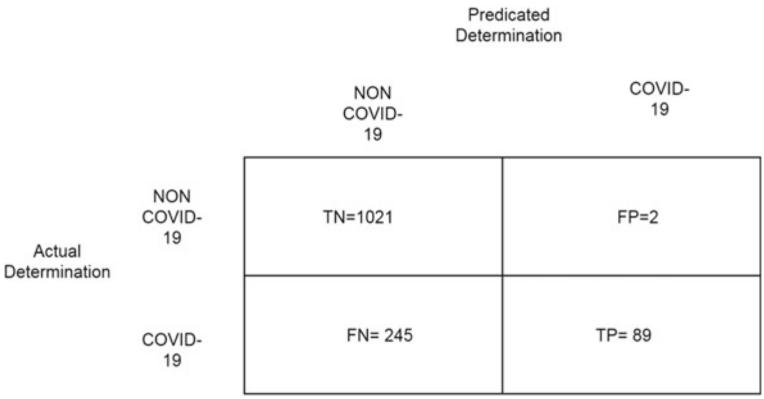
4.2 Discussions

The error matrices are calculated for all four frameworks for testing samples, i.e., MobileNet, ResNet-Version 50, Xception, and InceptionV3 are shown in Figs. 3, 4, 5 and 6. Figure 3 depicts fault grid for MobileNet framework. After the MobileNet framework is applied, observations are true positive (TP): 72, true negative (TN): 1259, false positive (FP): 13, and false negative (FN):13. Figure 4 depicts fault grid for ResNet50 framework. After the ResNet50 framework is applied, observations are true positive (TP): 89, true negative (TN): 1021, false positive (FP): 2, and false negative (FN):245. Figure 5 depicts fault grid for Xception framework. After the Xception framework is applied, observations are true positive (TP): 66, true negative (TN): 1262, false positive (FP): 8, and false negative (FN):21. Figure 6 depicts fault grid for—Inception V3 framework. After the Inception V3 framework is applied, observations are: true positive (TP): 70, true negative (TN): 1258, false positive (FP): 12, and false negative (FN):17.



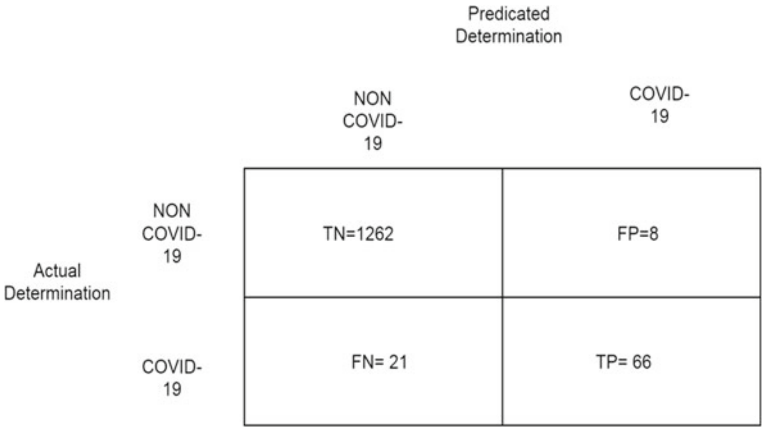
		Predicated Determination	
		NON COVID-19	COVID-19
Actual Determination	NON COVID-19	TN=1259	FP=13
	COVID-19	FN=13	TP= 72

Fig. 3 Fault grid—Mobilenet framework



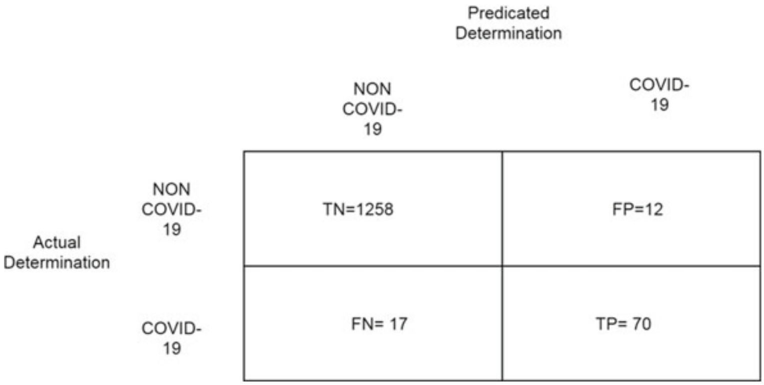
		Predicated Determination	
		NON COVID-19	COVID-19
Actual Determination	NON COVID-19	TN=1021	FP=2
	COVID-19	FN= 245	TP= 89

Fig. 4 Fault grid—ResNet50 framework



		Predicated Determination	
		NON COVID-19	COVID-19
Actual Determination	NON COVID-19	TN=1262	FP=8
	COVID-19	FN= 21	TP= 66

Fig. 5 Fault grid—Xception framework



		Predicated Determination	
		NON COVID-19	COVID-19
Actual Determination	NON COVID-19	TN=1258	FP=12
	COVID-19	FN= 17	TP= 70

Fig. 6 Fault grid—Inception V3 framework

Table 1 Results for four different CNN framework

CNN frameworks	Parameters for evaluation								
	TP	**FP**	**FN**	**TN**	**AC**	**SP**	**PR**	**RC**	**F1 metrics**
MobileNet	**72**	**13**	**13**	**1259**	**0.980**	**0.989**	**0.847**	**0.847**	**0.847**
ResNet50	**89**	**2**	**245**	**1021**	**0.817**	**0.998**	**0.978**	**0.266**	**0.4182**
Xception	**66**	**8**	**21**	**1262**	**0.978**	**0.993**	**0.891**	**0.758**	**0.819**
Inception V3	**70**	**12**	**17**	**1258**	**0.978**	**0.990**	**0.853**	**0.804**	**0.827**

Notes TP: true positive, FP: false positive, FN: false negative, TN: true negative, AC: accuracy, SP: specificity, PR: precision, RC: recall/sensitivity

In Table 1, the effects of these four frameworks have been summarized. Based on different variables such as accuracy, precision, recall, specificity, and F1 metrics, the performance of every model is calculated.

Accuracy is the closeness of the measurements to a specific value. Accuracy (AC) is calculated as follows:

$$\mathrm{AC} = \mathrm{TP} + \mathrm{TN}/\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}(\text{Total number of Data})$$

Specificity (SP) is also called as true negative rate which is a measure of the percentage of the negative dataset that is correctly classified. Specificity (SP) is calculated as follows:

$$\mathrm{SP} = \mathrm{TN}/\mathrm{TN} + \mathrm{FP}$$

Recall or sensitivity (RC) is also called as true positive rate which is a measure of the percentage of the positive dataset that is correctly classified. Recall (RC) is calculated as follows:

$$\mathrm{RC} = \mathrm{TP}/\mathrm{TP} + \mathrm{FN}$$

Precision (PR) is a percentage of retrieved datasets that are relevant (correct). Precision (PR) is calculated as follows:

$$\mathrm{PR} = \mathrm{TP}/\mathrm{TP} + \mathrm{FP}$$

F1 metrics are another measure of accuracy. It is a harmonic mean of PR and RC. It is calculated as follows:

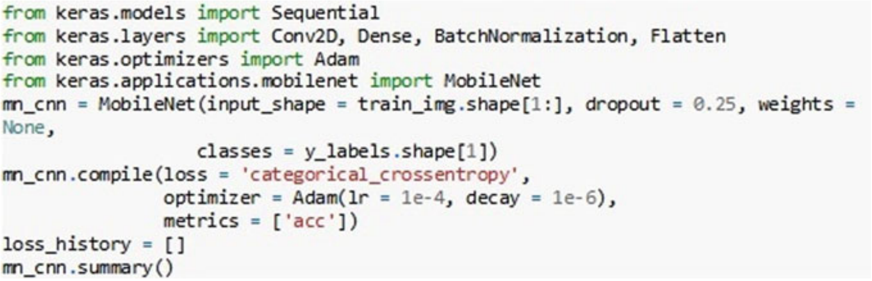
$$\text{F1 Metrics} = 2 * \mathrm{PR} * \mathrm{RC}/\mathrm{PR} + \mathrm{RC}$$

After examining the accuracy (AC) values of four frameworks, it has been suggested that the CNN framework is more reliable for the diagnosis of COVID-19 infection. With the highest F1 ranking, MobileNet turns out to be the best of all four versions. In all four versions, the value of specificity is more than 98%. Specificity is one of the characteristics of the model to eliminate false alarms. But, it is found that recall/sensitivity that determines a model's capacity to spot positive cases is only strong in model MobileNet and Inception Version V3. Even though the Xception framework having equivalence with the MobileNet framework, it cannot be regarded as a stable model due to its low sensitivity of 0.758.

After analyzing F1 metrics of all the four models, it has been found that MobileNet, Xception, and InceptionV3 models are having F1 metrics more than 80%, while ResNet50 is having F1 metrics of 41.82%. This means that performance of MobileNet, Xception, and Inception V3 models is better as compared with the ResNet50 model in the process of classifying any patient as COVID-19 positive or COVID-19 negative.

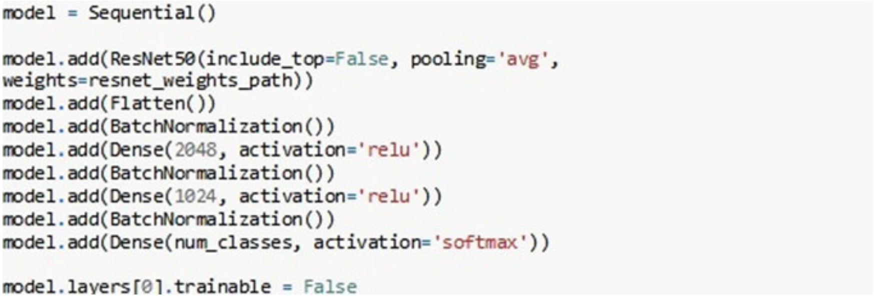
4.3 Pseudocode

Pseudocode of all the four models is shown in Figs. 7, 8, 9 and 10.



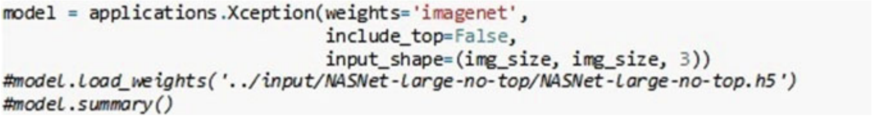
```
from keras.models import Sequential
from keras.layers import Conv2D, Dense, BatchNormalization, Flatten
from keras.optimizers import Adam
from keras.applications.mobilenet import MobileNet
mn_cnn = MobileNet(input_shape = train_img.shape[1:], dropout = 0.25, weights =
None,
                  classes = y_labels.shape[1])
mn_cnn.compile(loss = 'categorical_crossentropy',
               optimizer = Adam(lr = 1e-4, decay = 1e-6),
               metrics = ['acc'])
loss_history = []
mn_cnn.summary()
```

Fig. 7 Pseudocode—Mobilenet framework



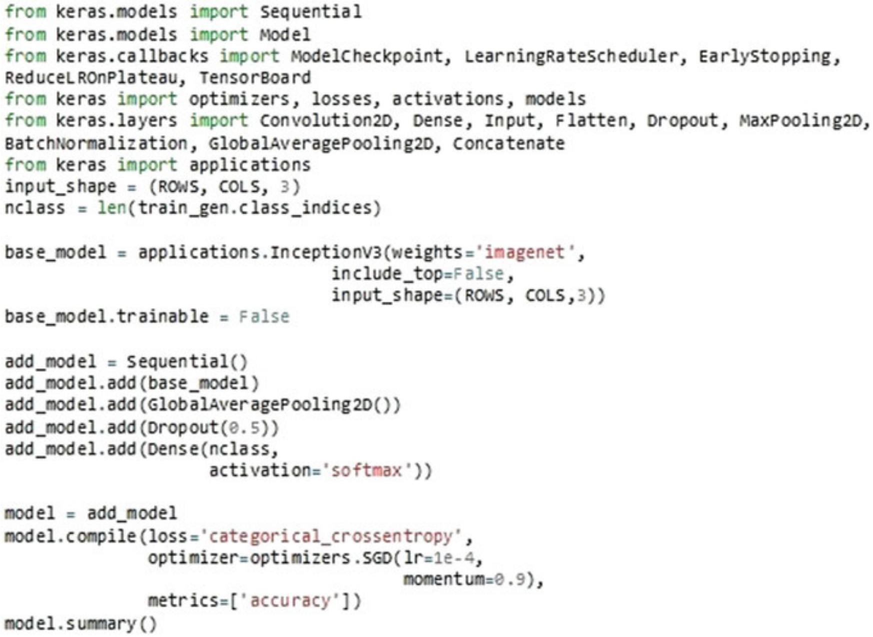
```
model = Sequential()

model.add(ResNet50(include_top=False, pooling='avg',
weights=resnet_weights_path))
model.add(Flatten())
model.add(BatchNormalization())
model.add(Dense(2048, activation='relu'))
model.add(BatchNormalization())
model.add(Dense(1024, activation='relu'))
model.add(BatchNormalization())
model.add(Dense(num_classes, activation='softmax'))

model.layers[0].trainable = False
```

Fig. 8 Pseudocode—ResNet50 framework



```
model = applications.Xception(weights='imagenet',
                              include_top=False,
                              input_shape=(img_size, img_size, 3))
#model.load_weights('../input/NASNet-Large-no-top/NASNet-Large-no-top.h5')
#model.summary()
```

Fig. 9 Pseudocode—Xception framework



```
from keras.models import Sequential
from keras.models import Model
from keras.callbacks import ModelCheckpoint, LearningRateScheduler, EarlyStopping,
ReduceLROnPlateau, TensorBoard
from keras import optimizers, losses, activations, models
from keras.layers import Convolution2D, Dense, Input, Flatten, Dropout, MaxPooling2D,
BatchNormalization, GlobalAveragePooling2D, Concatenate
from keras import applications
input_shape = (ROWS, COLS, 3)
nclass = len(train_gen.class_indices)

base_model = applications.InceptionV3(weights='imagenet',
                                      include_top=False,
                                      input_shape=(ROWS, COLS,3))
base_model.trainable = False

add_model = Sequential()
add_model.add(base_model)
add_model.add(GlobalAveragePooling2D())
add_model.add(Dropout(0.5))
add_model.add(Dense(nclass,
                    activation='softmax'))

model = add_model
model.compile(loss='categorical_crossentropy',
              optimizer=optimizers.SGD(lr=1e-4,
                                       momentum=0.9),
              metrics=['accuracy'])
model.summary()
```

Fig. 10 Pseudocode—Inception V3 framework

5 Conclusion and Future Work

At an unprecedented pace, coronavirus infection is impacting all well-being of the world citizenry. To reduce the transmission of infection, testing vast numbers of patients is important. For the assessment of this condition, polymerase chain reaction is a benchmark for pathological examination. An increasing amount of adverse false results, however, has contributed to the utilization of bronchus radiograms, which is one of the options for the COVID-19 diagnosis. In this case, deep learning-based recommendation systems will be a huge benefit where the intensity ratio of infected persons is inclined to a big quantity and the tomography proficiency required is limited. In this research, for diagnostic guidance of COVID-19 patients, various deep CNN frameworks have been examined on chest X-ray images. Such frameworks, well ahead sequel on the ImageNet network, consist of weights that are well ahead sequel, which will be useful to pass their previous information to the examined dataset.

As per the classification performance of the four deep learning models is concerned, the MobileNet model is evaluated with the highest percentage of the performance of 84.7% followed by Inception V3 (82.7%) and Xception (81.9%). One of the excellent findings of all four models is that all four models are having the capacity to eliminate faulty alarms. The main reason behind this scenario is all the models possess a specificity of more than 98%. ResNet50 possesses the lowest recall percentage of 26.6% among all models. It indicates that the capability to detect COVID-19 positive scenarios is the lowest in ResNet50. The recall percentage of MobileNet is found to be highest with a percentage of 84.7% among all models.

The findings indicate that frameworks based on CNN can diagnose coronavirus infection correctly. Training based on transfer technology performs a crucial part in maximizing notable performance.

Future work is suggested by performing fine-tuning of the above-discussed CNN models. This leads to an increase in performance (F1 metrics) of each model discussed. For the development of a suggested diagnostic method, other pretrained models can also be examined. CNN-based models which are used in this research work can also be utilized for other diseases like tuberculosis (TB), lung cancer, kidney diseases, and liver diseases.

References

1. K. Dhama, S. Khan, R. Tiwari, S. Sircar, S. Bhat, Y.S. Malik, A.J. Rodriguez-Morales, Coronavirus disease 2019–COVID-19. Clin. Microbiol. Rev. **33**(4) (2020). https://doi.org/10.1128/CMR.00028-20
2. J.M. Rhodes, S. Subramanian, E. Laird, G. Griffin, R.A. Kenny, Perspective: Vitamin D deficiency and COVID-19 severity–plausibly linked by latitude, ethnicity, impacts on cytokines, ACE2 and thrombosis. J. Intern. Med. **289**(1), 97–115 (2021). https://doi.org/10.1111/joim.13149
3. C.P. West, V.M. Montori, P. Sampathkumar, COVID-19 testing: the threat of false-negative results, in *Mayo Clinic Proceedings*, vol. 95(6). (Elsevier, 2020), pp. 1127–1129. https://doi.org/10.1016/j.mayocp.2020.04.004
4. F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, B. Cao, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet **395**(10229), 1054–1062 (2020). https://doi.org/10.1016/S0140-6736(20)30566-3
5. S. Huang, F. Lee, R. Miao, Q. Si, C. Lu, Q. Chen, A deep convolutional neural network architecture for interstitial lung disease pattern classification. Med. Biol. Eng. Comput. 1–13 (2020). https://doi.org/10.1007/s11517-019-02111-w
6. J. Ker, L. Wang, J. Rao, T. Lim, Deep learning applications in medical image analysis. IEEE Access **6**, 9375–9389 (2017). https://doi.org/10.1109/ACCESS.2017.2788044
7. D. Das, K.C. Santosh, U. Pal, Truncated inception net: COVID-19 outbreak screening using chest X-rays. Phys. Eng. Sci. Med. **43**(3), 915–925 (2020). https://doi.org/10.1007/s13246-020-00888-x
8. D.P. Fan, T. Zhou, G.P. Ji, Y. Zhou, G. Chen, H. Fu, L. Shao, Inf-net: automatic COVID-19 lung infection segmentation from CT images. IEEE Trans. Med. Imag. **39**(8), 2626–2637 (2020). https://doi.org/10.1101/2020.04.22.20074948
9. T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Med. **121**, 103792 (2020). https://doi.org/10.1016/j.compbiomed.2020.103792

10. U. Özkaya, Ş. Öztürk, M. Barstugan, Coronavirus (COVİD-19) classification using deep features fusion and ranking technique, in *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach* (Springer, Cham, 2020), pp. 281–295. arxiv-2004.03698
11. W. Linda, A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. J. Netw. Comput. Appl. **20**, 1–12 (2020)
12. L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. Comput. Methods Prog. Biomed. **196**, 105608 (2020). https://doi.org/10.1016/j.cmpb.2020.105608
13. A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (COVİD-19) using X-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849 (2020)
14. J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in *2009 IEEE conference on computer vision and pattern recognition* (IEEE, 2009), pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848
15. L. Sifre, S. Mallat, Rigid-motion scattering for texture classification. *arXiv preprint* arXiv:1403.1687 (2014)
16. A.I. Khan, J.L. Shah, M.M. Bhat, CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. Comput. Methods Prog. Biomed. **196**, 105581. https://doi.org/10.1016/j.cmpb.2020.105581
17. C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31(1) (2017, February)
18. https://github.com/ieee8023/covid-chestxray-dataset

Generation of Structured Query Language from Natural Language Using Recurrent Neural Networks

Lubna Mohammed Kunhi and Jyothi Shetty

Abstract With the rise of the digital era, data has become a prominent resource. A humongous amount of data is generated every single day. This data, typically raw, is said to be in an unstructured format and is converted into a structured format to derive meaning and value and often stored in huge databases. The data is then queried upon using querying languages like SQL. This is where the generation of SQL from natural language comes into the picture. Natural language processing is a technology used to help computers understand the natural language of humans. While it is easier for humans to communicate in natural language, it is almost impossible for a computer to master. By providing an efficient method to convert NL to SQL, this paper provides an overview of the various approaches in existence, and further goes on to develop a mechanism for the same using recurrent neural networks. The proposed approach makes use of the Spider dataset and has been differentiated to account for the accuracy for various clauses including JOIN operations in an SQL query.

Keywords Parts of speech tagging · Structured Query Language · Natural language processing · Recurrent neural network

1 Introduction

The amount of data is fast growing as the number of devices that are connected to the Internet is increasing day by day. According to the IBM marketing cloud study, digital devices have become data-pumping factories generating huge portions of data, and thus contributing to 90% of the data on the Internet to have been created since 2016. This data, eventually stored in huge databases most of the time, can

L. M. Kunhi (✉) · J. Shetty
Department of CSE, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Karkala, India
e-mail: 4nm19scs05@nmamit.in

J. Shetty
e-mail: jyothi_shetty@nitte.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_6

be queried by experts in querying languages. As the applications of this data usage are widespread and need to be accessible to almost every entity, there comes a requirement where people can communicate with the dataset using the language known to them, otherwise known as natural language.

Natural language processing is a field under artificial intelligence that deals with how computers can process the natural commands given by people. The concept of NLP began in the early 1950s, which was when natural language evolved as a subject. Even though people are well equipped to understand natural languages, it often poses challenges for computers to understand. Despite the challenges faced, the technology available today makes it efficient to overcome these obstacles. The target would be to incorporate the computer with the functionality of understanding and generating natural language which would leverage users to address the computer as if they were addressing a fellow person.

Data is generated every second in various industries like the health care sector, sales, and marketing, where it is generally stored in a relational database. Given this data, to retrieve required information, one must be aware of querying languages like SQL. Considering a real-life situation where a staff would like to gather the names of employees who belong to a particular department, it would at the least require the use of three basic clauses including SELECT, FROM, and WHERE. The problem becomes more complex when they would want the data from different departments which would require JOIN operations. This is the problem faced across various fields, where the users are technically not equipped with languages like SQL. Training a model with which people can communicate with the language known to them would provide easy access to data.

One of the early methods that were used to convert natural language to Structured Query Language dates back to the 1970s. LUNAR, a system that was used to answer a question regarding rock samples brought back from the moon, was developed by Woods in 1973. This system made use of an augmented pointer network and is based on a graph-theoretic structure. Eventually, LIFER/LADDER was developed by Hendrix in 1978 [1]. It makes use of semantic grammar to parse the questions raised by the user. However, this technique was efficient for a single-table database, hence inefficient with multiple tables. Further, in 1982, CHAT-80 was developed by Warren and Pereira [2], which allowed users to interrogate the system based on geographical knowledge. These early systems, however, had poor retrieval time, less support for language portability, and had complex configurations which contributed to less adaptation of such systems for commercial purposes. Moving forward, various approaches to convert natural language to Structured Query Language have been carried out. Various techniques have evolved ranging from rule-based approaches to techniques that use artificial intelligence.

This paper is partitioned as follows: Sect. 2 provides the taxonomy of different techniques being developed and provides an overview of the same. Section 3 discusses the methodology adopted for the generation of SQL from NL. Section 4 provides the results obtained with discussion. Section 5 provides the conclusions with the future enhancement.

2 Related Work

The techniques can be classified under five major categories: rule-based, sketch-based, machine learning-based, deep learning-based techniques, and reinforcement learning techniques.

One of the initial methods proposed was the rule-based approaches as discussed earlier. One such technique is discussed here as proposed by Tanzim et al. [2]. The process involves four steps which include word check, removal of unnecessary words, and tokenization followed by mapping to CFG rules. A data dictionary is often used to identify valid words from the input. Using the CFG rules, the valid words are mapped onto them to identify the column names and table names from the user input. This technique does not consider the semantic meaning of the words and hence is prone to errors. The results were analyzed by classifying the output query as "good," "fair," and "bad," where "good" being the queries that were correct and produced the rightful results when queried against a database, "fair" being the queries that are syntactically correct but failed to produce rightful results, and "bad" being the queries where no query is being generated for a given input. The results were simulated on a varying number of queries to produce the following results: 47% good, 35% fair, and 18% bad. The accuracy here could have been improved by adding production terms in CFG as well as an increase in the number of entries in the data dictionary.

Another approach that was used in the early research is the sketch-based approach as proposed by Yaghmazadeh et al. [3]. The idea was to build an end-to-end system which would be called SQLizer. The process is three-step fold which involves sketch generation, type-directed sketch completion, and sketch refinement. A query sketch is generated initially rather than a full-fledged query. This enables a translation without requiring database-specific training, hence being data-agnostic. Step 2 involves type-directed program synthesis that completes the sketch from the previous step. Sketch refinement is an optional step used when the initial query sketches generated using semantic parsing may not accurately reflect the structure expected query. This is achieved by fault localization. The results obtained in this approach were made using 455 queries involving three databases, namely Microsoft academic search (MAS), IMDB, and YELP. For a single query, Top 1, the following results were obtained: 80.6% accuracy for MAS, 77.9% accuracy for IMDB, and 75.0% for YELP. This has been a step up from previous methods in terms of accuracy achieved.

Ye et al. [4] combined the dictionary-based approach with statistical learning that resulted in a better accuracy compared to the traditional approaches. The work was focused on the parts of speech (PoS) tagging where traditionally dictionary rules were used to tag the words from the input. The most common PoS tagging based on statistical learning mainly uses conditional random field (CRF), hidden Markov model (HMM), and maximum entropy (ME). Here, the words are extracted from an online encyclopedia to extend the original dictionary. Each word is classified based on the entry tag which is valid for dictionary words. If the word is ambiguous, it is tagged using all possible tags. To find the best tag among the possible tags, maximum entropy is used. This approach attained an accuracy of 95.80% when tested on the

People's Daily January 1989 news dataset. Besides, it also showed an accuracy of 88.0% for unambiguous words.

The use of artificial intelligence has been widely popular in recent researches. One such approach is proposed by Boyan et al. [5], which was named NADAQ. NADAQ consists of mainly three components: data storing module, model management module, and user interface. The first component includes the metadata from the tables. The core interest of this system lies in the management module where the translation takes place. A recurrent neural network is used on top of an encoder–decoder framework along with a query rejection module. Two strategies used here are the MLP and threshold method. MLP methods check for valid questions to query against the database, while the threshold method rejects an input if the entropy exceeds a predetermined threshold value. The results were calculated based on recall and precision of the query results on three different databases, namely IMDB, MAS, and GEO. The F1 score was calculated as follows: 0.839 on the GEO dataset, 0.9 on IMDB, and MAS datasets, which seem to perform better with the MLP method when compared to the threshold method.

A further improvement in this field is the use of the reinforcement learning technique that was proposed by Victor et al. [6]. They proposed a model which would use rewards of query execution on the given database to learn a new policy using cross-entropy loss. This research addresses the limitations of using an augmented pointer network that generates the SQL query on a token-by-token basis and also does not leverage the inherent structure in SQL. Hence, to overcome the problems faced, they developed the SEQ2SQL model. A typical SQL query is considered to have three parts: an aggregation operator, a SELECT component as well "a WHERE" component. To decide the aggregate operator that has to be used, they start with computing a scalar attention score and normalize it to product distribution. The input is simply the sum over the input encodings on which the score is distributed. Then, a multi-layer perceptron is applied to the input. The selection of the SELECT component of the SQL query depends on two factors: the input question and the table columns. The SELECT clause is predicted as a result of a matching problem using a pointer. The WHERE clause is also predicted using a pointer decoder that uses cross-entropy loss. The dataset that was tested upon was the WikiSQL dataset which is annotated with 80,654 tables. This model performed pretty well-compared to the state-of-the-art, providing an F1-score of 69.2%.

3 Proposed Approach

While the existing approaches pose different challenges, the proposed approach aims to tackle the prominent ones. Here, [2] provides the most efficient results for simple queries; however, it does not work well for complex queries requiring the joining of tables. The consideration of generating grammar rules for only three clauses makes it below par when it comes to high complexity. The [3] requires the users to possess some knowledge of SQL, which is demanding when it comes to ease of accessibility.

And, [4] has been enhanced to predict ambiguous words; however, it possesses the problem of incurring high complexity when it requires a huge data dictionary for mapping. Then, [4] provides an enhanced approach over its counterparts by providing a rejection and recommendation component, which would, however, increase the time complexity. The requirement, hence, is to tackle the problem of conversion of natural language questions to Structured Query Language in the most efficient way. The proposed approach involves the consideration of more clauses by including JOIN operations as well as AGGREGATE as well as the use of LSTMs for predicting the PoS tags of the words surrounding the word of interest which would enhance the accuracy. Most research today is performed on the WikiSQL dataset, which is a corpus of 87,726 hand-annotated SQL queries. However, one major drawback of this dataset is that it assumes there is no JOIN operation. In reality, data is distributed over multiple tables and cannot be taken into grant that they appear on a single table; hence, the use of Spider dataset where 40% of the queries contains JOIN operations.

The process involves three major steps: dataset selection, data preprocessing, and conversion from natural language to Structured Query Language.

3.1 Dataset

The first step to start with is the choice of dataset. For implementation, the dataset used is the Spider dataset which is a large-scale dataset consisting of two hundred different databases, and 10,181 questions in natural language, and 5693 unique complex queries. The tables are well defined over a huge area, covering up to 180 domains, hence the name Spider for being cross-domain. It is a large-scale complex dataset annotated manually. There are a variety of unique complex SQL queries and databases that appear on both the train and test sets. More importantly, this dataset assumes the use of JOIN operations which makes it ideal for real-life applications. The training sets are in JSON format and consists of the following content which includes the question in natural language, the tokens present in that question, the id of the database to which the question is directed, the query that corresponds to the given question, the tokens that comprise the final SQL query, and finally, the parsed results of the SQL query. The dataset comprises over 200 tables, which includes the database id with which the table is identified, the original as well as the normalized table name, the column names as well as the normalized column names, the data type of the columns, and the keys which includes the primary key and foreign keys. A sample entry in the training set is shown below:

Database ID: `department_management`

Natural Language question: `How many heads of the departments are older than 56?`

Natural Language question tokens: `"How", "many","heads", "of","the", "departments", "are", "older", "than", "56","?"`

SQL Query: `SELECT count(*) FROM head WHERE age >56`

SQL Query tokens: `"SELECT","count", "(", "*", ")","FROM","head", "WHERE", "age", ">", "56"`

3.2 Data Preprocessing

The model accepts the English language input sentence. Each sentence is preprocessed to understand the meaning and context of the word, eventually tagging each of the relevant words into the corresponding PoS tag. The steps involved in preprocessing are summarized in Fig. 1. The first step involves tokenization. Tokenization is the process where the input sentence is broken down into tokens (or words). This is achieved by importing word_tokenize from the NLTK library. The output is an array of words separated by commas. To prevent the same word from being categorized as two different ones, the words are converted into lower case. This prevents unnecessarily large dimensions of word vectors at later stages. This is implemented with the lower() method provided by Python. An input of the English language may consist of various words that do not provide meaning or prove to be of importance for any later steps. These words are known as stop words and are hence removed. NLTK provides a corpus of stop words to achieve the same. Next is parts of speech tagging, where each word is tagged with one of the PoS tags based on definition and context. This is followed by lemmatization, which is similar to stemming but varies in that the reduced word exists in the dictionary. NLTK has lemmatizers implemented which are based on a rule-based approach.

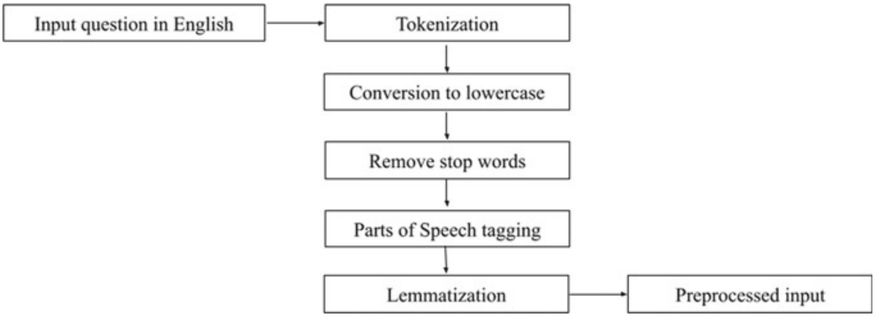


Fig. 1 Processing of input sentence prior to conversion

3.3 Generation of SQL from NL Using RNN

Recurrent neural networks (RNN) are built with internal memory, the feature that has been used for the translation process. It connects previous data to the present task of the neural network. The proposed method is based on the use of long short-term memory (LSTM). LSTM networks are a sort of modified version of RNN, which makes it easier to remember historical data. An LSTM in general has three gates: input gate, output gate, and forget gate. As the name suggests, the input gate coordinates the data input at each step, while the output gate coordinates the data moving on to the next layer. The forget gate controls how much data to lose at each step.

The proposed approach involves the conversion of a sequence of inputs in the form of natural language into a sequence of output in the form of SQL query. The dataset used to train the model is Spider. This particular dataset was chosen because it is a large-scale dataset and can synthesize the dataset without having the table contents. It consists of 200 databases; each having multiple tables and is defined over 180 domains. The model is trained for evaluating five different SQL clauses, namely SELECT, FROM, WHERE, AGGREGATE, and JOIN. The model consists of the encoder–decoder module as explained in the following section.

3.3.1 Encoding and Decoding

Figure 2 demonstrates the working of encoder–decoder for SQL generation from natural. Consider the query in natural language, Q, consisting of “n” tokens as (q_n, q_1, …, q). Each token q_i is mapped to its PoS tag, and hence, its vector representation is $q_i = w$. one_hot(q_i). Here, “w” corresponds to the respective weight, and one-hot encoding is a way of representing words in numerical format. Each vectorized token now acts as an individual input in the RNN at a time “t,” as well as uses the activation value at “$t - 1$,” in addition to the input at a time “t.” To regain the history, LSTMs are used.

The decoder produces a sequence of output tokens sn conditioned on the input query, Q. Each output is embedded in its vector representation which includes the structure tokens and column tokens. In the proposed approach, five clauses (structured tokens) are considered: SELECT, FROM, WHERE, AGGREGATE, and JOIN. Column tokens refer to the column names of the tables which are queried upon. The LSTM updates the hidden state in every layer and produces the output SQL query, S.

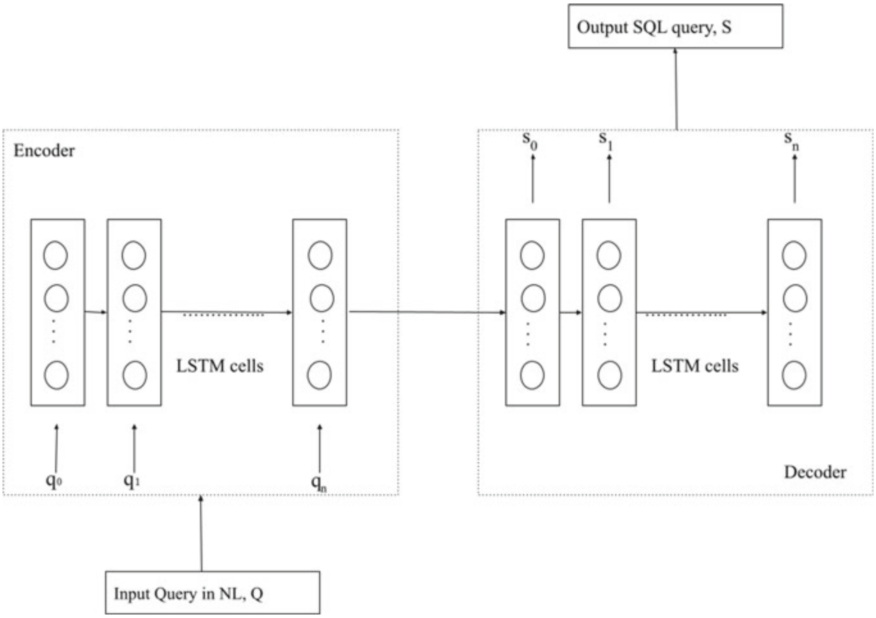


Fig. 2 Working of encoder–decoder for SQL generation

4 Experimental Results and Analysis

The project implementation was done using Python 3.6 using Pycharm as the IDE. The various data preprocessing techniques were performed using natural language toolkit (NLTK) which is a tool often used for working with human language. It consists of over 50 corpora and lexical analysis resources and provides the mechanism for performing various preprocessing steps including tokenization, stemming, tagging, and parsing.

The model to convert NL to SQL employs RNNs with LSTM cells. It is trained with a learning rate of 0.001 using Adam optimizer which is an adaptive learning rate optimization algorithm that leverages the power of adaptive learning to find individual learning rates. The model is trained in batches of 48. It is evaluated using accuracy, precision, and recall. Accuracy is the proportion of correct predictions to the total predictions made by the model, whereas both precision and recall are based on relevance.

The content to follow shows how the model works on a given natural language input queried on a company database.

NL Question: `What are the names of employees under the manager 'Adam'?`

SQL Query:
```
SELECT      employee.employee_name
FROM        employee
```

```
INNER JOIN  manager
ON          manager_id= employee.manager_id
WHERE       manager_name = 'Adam'
```

The evaluation is done for all the five clauses identified. The SELECT clause was found to have an accuracy of 79.54%, and FROM clause with an accuracy of 98.21% which was found to be the highest. WHERE clause was evaluated to have the least accuracy of 70.11%, AGGREGATE with an accuracy of 84.32%, and JOIN with an accuracy of 77.89%. These results are shown in Fig. 3.

The precision and recall of the five different clauses are summarized in Fig. 4. The FROM clause is evaluated better in terms of performance with 0.92 precision and 0.93 recall.

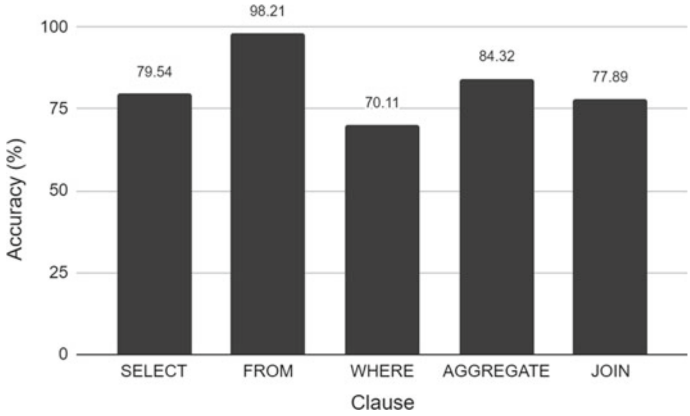


Fig. 3 Accuracy results for different clauses using Spider dataset

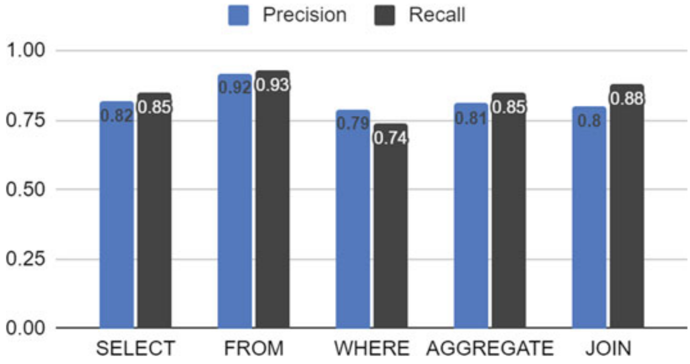


Fig.4 Precision and recall results for different clauses using Spider dataset

5 Conclusions

In this paper, work was carried on generating SQL queries from natural language questions primarily in English. The model architecture uses a recurrent neural network built with LSTM cells for the encoding and decoding process. The model was evaluated on five different clauses, namely SELECT, FROM, WHERE, AGGREGATE, and JOIN, and performance on each clause was identified. The proposed approach provides the highest accuracy for FROM clause with an accuracy of 98.21% followed by AGGREGATE with 84.32%, SELECT with 79.54%, JOIN with 77.89%, and WHERE with 70.11%. The performance related to FROM clause is comparatively better, both in terms of accuracy as well as with precision and recall (0.92 and 0.93, respectively). The results show that this model is also efficient when it comes to JOIN operations, which is seen in most real-life applications. This model can be made more versatile by supporting multiple languages, thus finding its way into various applications with great scope in various fields. The future work may involve consideration of more clauses and integration of speech to text converter.

References

1. P. More, B. Kudale, P. Deshmukh, I.N. Biswas, N.J. More, F.S. Gomes, An approach for generating SQL query using natural language processing. 226–230 (2020). Springer
2. T. Mahmud, K.M. Azharul Hasan, M. Ahmed, T.H.C. Chak, A rule-based approach for NLP based query processing, in *2nd International Conference on Electrical Information and Communication Technologies (EICT)*, Khulna (2015), pp. 78–82
3. N. Yaghmazadeh, Y. Wang, I. Dillig, T. Dillig, SQLizer: query synthesis from natural language, in *Proceedings of the ACM on Programming Languages* (2017)
4. Z. Ye, Z. Jia, J. Huang, H. Yin, Part-of-speech tagging based on dictionary and statistical machine learning, in *35th Chinese Control Conference (CCC)*, Chengdu (2016), pp. 6993–6998
5. B. Xu, R. Cai, Z. Zhang, X. Yang, Z. Hao, Z. Li, Z. Liang, NADAQ: natural language database querying based on deep learning. IEEE Access 7, 35012–35017 (2019)
6. V. Zhong, C. Xiong, R. Socher, SEQ2SQL: generating structured queries from natural language using reinforcement learning (2017)
7. J. Sangeetha, R. Hariprasad, An intelligent automatic query generation interface for relational databases using deep learning technique. (Springer, 2019), pp. 22–30
8. X. Xu, C. Liu, D.X. Song, SQLNet: generating structured queries from natural language without reinforcement learning (2017)
9. S. Kumar, A. Kumar, P. Mitra, G. Sundaram, System and methods for converting speech to SQL. Emerg. Res. Comput. Inform. Commun. Appl. ERCICA 291–298 (2013)
10. P.-S. Huang, C. Wangy, R. Singh, W. Yihz, X. He, Natural language to structured query generation via meta-learning (2018)
11. S.P. Singh, A. Kumar, H. Darbari, Deep neural based name entity recognizer and classifier for English language, in *International Conference on Circuits, Controls, and Communications (CCUBE)*, Bangalore (2017), pp. 242–246
12. J. Deriu, K. Mlynchyk, P. Schlapfer, A. Rodrigo, D. von Grunigen, N. Kaiser, K. Stockinger, E. Agirre, M. Cieliebak, A methodology for creating question answering corpora using inverse data annotation, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020)

13. C. Baik, H.V. Jagadish, Y. Li, Bridging the semantic gap with SQL query logs in natural language interfaces to databases, in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, Macao, China (2019), pp. 374–385. https://doi.org/10.1109/ICDE.2019.00041
14. M. Uma, V. Sneha, G. Sneha, J. Bhuvana, B. Bharathi, Formation of SQL from natural language query using NLP, in *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Chennai, India (2019), pp. 1–5. https://doi.org/10.1109/ICCIDS.2019.8862080
15. S.S. Khan, A. Saeed, Y. Majeed, M. Kamran, *Natural Language Based SQL Query Verification Against Relational Schema*, vol. 932. (Springer, Singapore, 2019)

Machine Learning and Deep Learning Approaches for Intrusion Detection: A Comparative Study

G. Prethija and Jeevaa Katiravan

Abstract Intrusion detection is a major challenge for security experts in the cyber world. Traditional IDS failed to detect complex and unknown cyber-attacks. Machine learning has become a vibrant technology for cybersecurity. There exists several machine learning algorithms to detect intrusion. Most classifiers are well suited to detect the attacks. However, improving accuracy and detecting unknown attacks in existing IDSs is a great challenge. Therefore, the detailed comparative study of various machine learning approaches such as artificial neural networks, support vector machine, decision tree, and hybrid classifiers used by researchers for intrusion detection are done. Deep learning is an emerging approach which suits well for large data. Deep learning techniques find optimal feature set and classify low-frequency attacks better than other techniques. This study also summarizes literatures in deep learning approaches such as deep auto-encoder, Boltzmann machine, recurrent neural networks, convolutional neural networks, and deep neural networks. Moreover, the datasets used in various literatures and the analysis of deep learning approaches based on the performance metrics are also done. Future directions to detect intrusion are also provided. This study in fact will be helpful to develop IDS based on artificial intelligence approaches such as machine learning and deep learning.

Keywords Machine learning · Intrusion detection · Feature selection · Deep learning · Cyber security · Classifier

1 Introduction

Cyber security threats are increasing day by day. So, there is a high demand for intrusion detection. Recently, more cyber threats are reported. Even though the technology grows, the hackers are still increasing, and it is a big challenge for cyber

G. Prethija (✉)
Department of Information Technology, Velammal Engineering College, Chennai, India

J. Katiravan
Department of Computer Science and Engineering, Velammal Engineering College, Chennai, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_7

security experts. The National Cyber Security Center (NCSC) is investigating a large-scale domain name system (DNS) hijacking campaign that has targeted Middle East, Europe, and the US countries which affected government and commercial organizations worldwide in January 2019 [1]. Data breaches, crypto jacking, and Ransomware attacks and threats to connected devices are few cyber security dangers reported. Norton's Cyber security facts and statistics reports that WannaCry Ransomware attack affected nearly tens of thousands of computers across the world. Machine learning plays a predominant role in cyber security field for detecting real threats in an enterprise by security analysts.

An intrusion detection system (IDS) checks for malicious activity among all incoming and outgoing packets. The firewall has major shortcomings such as inability to detect interior attacks, providing reliable security strategy, and it has a single bottleneck spot and invalid spot, etc. An IDS assesses a suspected intrusion and warns the administrator. An IDS also monitors the interior attacks. Host intrusion detection systems (HIDS) check the inward and outward packets only from the devices and warn the administrator or user if any malicious activity is discovered. HIDS cannot monitor the entire network. Network intrusion detection systems (NIDS) monitor all inbound and outbound traffic by placing an IDS within the network. It alerts the administrator once a malicious activity is identified.

In a misuse or signature-based detection approach, current behavior of network is matched against predefined patterns of attacks detected. They are not efficient to recognize unknown attacks. One of the major drawbacks of signature-based IDS is signature database must be frequently updated and preserved. Anomaly-based detection determines the normal behavior of the system and uses it as baseline for detecting anomalies. It can detect unknown attacks. An IDS can be successfully developed using machine learning algorithms.

Machine learning is a complex computation process which infers a learning model from input samples automatically. Learning models use some statistical functions or rules for describing data dependencies. Machine learning algorithms is categorized into unsupervised learning, supervised learning, and semi-supervised learning. In supervised machine learning, all data are labeled. The pair of input and target output is fed to train the given function, and thus, the entire learning model is trained. If an algorithm is used to learn the mapping function $Y = f(X)$ from the input X to output Y, then it is supervised learning. The aim is approximating the mapping function so that the algorithms learn to estimate the output from the input data. Regression and classification problems are the major grouping of supervised learning problems. If the output variable is a categorical value, then it is classification problem. If the output variable is a real value, then it is a regression problem. In unsupervised learning, all data are unlabeled, and no label is provided in sample data. If only input data is available without corresponding output variables, then it is unsupervised learning. Clustering and association problems come under the category of unsupervised learning problems. If inherent grouping in data is done, then it is called clustering problem. If rules that describe large portions of data are discovered, then it is association rule. In semi-supervised machine learning, only some data are labeled, and most of the

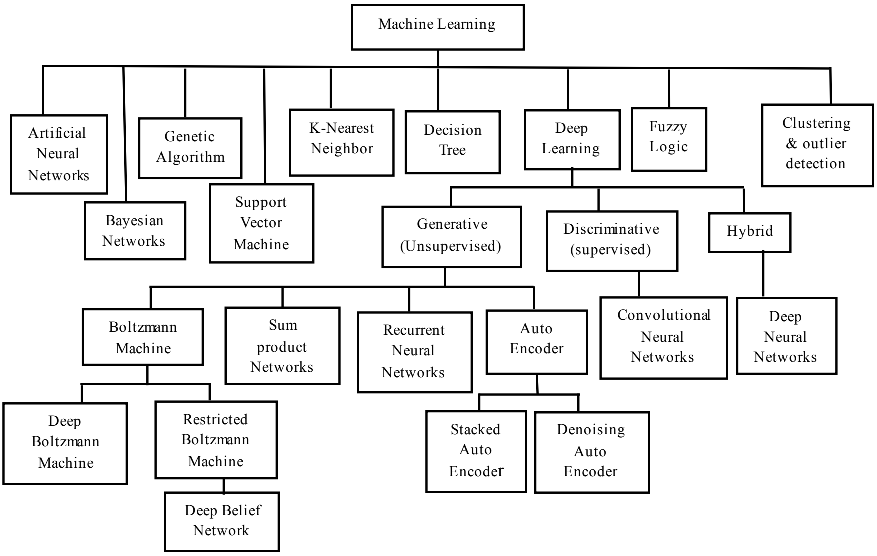


Fig. 1 Machine learning techniques used for IDS

data are unlabeled. It is an amalgamation of supervised and unsupervised techniques. Figure 1 depicts the various machine learning techniques used for IDS.

The paper is organized into sections. In Sect. 2, the survey papers related to machine learning for intrusion detection is discussed. The performance metrics that can be used to evaluate intrusion detection is discussed in Sect. 3. In Sect. 4, the datasets used for intrusion detection using machine learning are explained. In Sects. 5 and 6, recent literatures in machine and deep learning for intrusion detection are focused. In Sect. 7, observations and future directions in intrusion detection using machine learning algorithms are discussed.

2 Related Works

Tsai et al. [2] provided a survey of various machine learning algorithms. They distributed the research articles year-wise based on type classifier design, datasets used, and feature selection algorithm used. However, they did not compare the performance metrics of any machine learning algorithms. Their review is done with research papers published during the period 2000 and 2007.

Buczak et al. [3] surveyed machine learning approaches and data mining techniques that are used for intrusion detection. They categorized the papers based on the machine learning approaches. As well, they have categorized the research papers based on detection methodology either misuse or anomaly. They insisted the importance of datasets. The time complexity of machine learning algorithms is also discussed. However, their discussion is done till 2014.

Mishra et al. [4] provided detailed information about classification of attacks, machine learning approaches, and feature selection algorithm. They compared the performance of machine learning algorithms based on classifier type. The detailed analysis is done on types of attacks for different types of classifier. They carried out performance analysis based on detection rate of various machine learning approaches for all attack types. The tools used for machine learning are also discussed. They focused mostly on low-frequency attacks.

3 Metrics Used to Evaluate Intrusion Detection System

The performance evaluation of any intrusion detection system can be done by the metrics such as: accuracy (ACC), recall (REC), precision (PRE), true negative rate (TNR), false alarm rate (FAR), false negative rate (FNR), F-measure, Mathews correlation coefficient (MCC), ROC graph, and Kappa statistics. The metrics required for evaluation are computed from confusion matrix. A matrix that describes the performance of a given classification model (or "classifier") is called confusion matrix. It denotes true and false classification results. The ways in which confusion is made when a prediction is done by the classification model is depicted by confusion matrix. True positive (TP): It is the number of correctly identified anomaly records. False positive (FP): It represents the number of incorrectly identified usual records that are detected as anomaly. True Negative (TN): It represents the number of correctly detected records. False Negative (FN): It shows the number of incorrectly detected anomaly records.

4 Datasets Used for Intrusion Detection Research

Most researchers used the datasets DARPA, knowledge discovery and data mining (KDD) Cup, and network security laboratory-KDD (NSL-KDD), UNSW-NB15, Kyoto, and AWID for intrusion detection. Figure 2 illustrates the relation between DARPA, KDD, and NSL-KDD datasets.

The datasets used for intrusion detection by researchers have both training data and testing data. The first standard corpus for the evaluation of intrusion detection

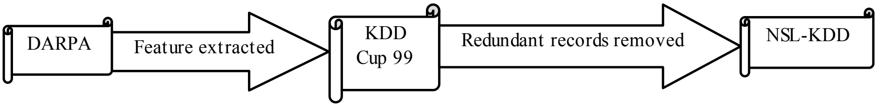


Fig. 2 Relation between DARPA, KDD, and NSL-KDD datasets

Table 1 Data size comparison for different datasets

Dataset	Training size	Testing size
DARPA 99	6.2 GB	3.67 GB
KDD99	4898431bytes	311,029 bytes
NSL-KDD	125,973 bytes	22,444 bytes
UNSW-NB15	175,341 bytes	82,332 bytes
AWID	1,795,575 bytes	575,643 bytes

system was created by MIT Lincoln Laboratory's in 1998 under the sponsorship of DARPA (Kendell 1999).

Tavallaee et al. [5] have analyzed KDD dataset in detail. KDD'99 features can be classified into three groups, namely basic, traffic, and content features. The major problem in this dataset is the enormousness of duplicate records. Tavallaee et al. [5] published the NSL-KDD dataset which eliminates duplicate records in training set thereby overcoming the drawback of classifiers gets biased toward more frequent records. Due to absence of modern attack styles and traffic situations in KDD dataset, a new dataset (UNSW-NB15) was developed by ACCS—an American Cyber security Center. This dataset has a 49-feature set and a total of 2,540,044 records [6]. Kyoto dataset (2009) is created from real environment traffic data collected from honey pot over 3 years. AWID is a dataset that is generated from a wireless network traffic [7]. The traces were produced from a wireless local area network (WLAN) and were secured by the wired equivalent protocol (WEP). Iman Sharafaldin et al. [8] introduced a reliable and real-world dataset, namely CICIDS2017. It contains benign and seven common attack network flows, namely brute force attack, heartbleed attack, botnet, DoS attack, DDoS attack, web attack, and infiltration attack with 80 features.

The dataset size comparison of training and test data for different datasets is shown in Table 1. The datasets used for intrusion detection by researchers have both training data and testing data. The dataset size comparison of training and test data for different datasets is shown in Table 1.

5 Literatures in Machine Learning for Intrusion Detection

Nowadays, researchers used machine learning approaches for intrusion detection. The datasets mostly used for evaluation of the algorithms are KDD Cup 99, NSL-KDD, Kyoto, UNSW-NB15, and AWID. The machine learning approaches are categorized as either single or hybrid based on classifier type used. Feature selection algorithm is also used by few researchers. Table 2 shows the comparison of different machine learning algorithms with classifier type, classification technique, feature selection technique, datasets used, performance metrics, and techniques used for comparison. The performance metric comparison for different machine learning approaches discussed in Table 2 is tabulated in Table 3.

Table 2 Comparison of different machine learning algorithms

Classifier type/baseline	Author	Classification technique	Feature selection algorithm	No. of features and feature no	Dataset used	Merits	Demerits
Single/ANN	Wang et al. [9]	Constrained-Extreme learning machines, Adaptively incremental learning strategy	–	–	KDD-DoS, 10% KDD, UNSW-NB15, NSL-KDD	Better detection rate, Learning speed high	Concept drift
Single/ANN	Chiba et al. [10]	Back Propagation Learning Algorithm	Information Gain Feature Selection Algorithm Modified Kolmogorov–Smirnov Correlation-based Filter Algorithm	12 features (3, 5, 6, 12, 23, 24, 27, 28, 31, 32, 33, 35) 17 features (2, 3, 4, 5, 6, 7, 8, 10, 11, 14, 22, 23, 24, 28, 30, 36, 39) 34 features (1, 5, 6, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41)	KDD CUP 99	higher detection rate and lower false positive rate	Learning Rate and Momentum term is to be added as optimal parameters

(continued)

Table 2 (continued)

Classifier type/baseline	Author	Classification technique	Feature selection algorithm	No. of features and feature no	Dataset used	Merits	Demerits
Single/SVM	Zhao et al. [11]	Multiclass SVM	Redundant Penalty Between Features Mutual Information Algorithm (RPFMI)	DOS 23 features (7, 2, 13, 4, 19, 15, 16, 17, 18, 14, 28, 20, 23, 31, 29, 11, 26, 27, 40, 42, 3, 38, 1) U2R 16 features (8, 9, 6, 21, 22, 23, 10, 4, 27, 39, 15, 11, 14, 18, 12, 19) R2L 5 features (8, 9, 10, 12, 15) Kyoto 6 features (16, 17, 4, 14, 19, 2)	KDD Cup 99 Kyoto 2006+	Feature selection algorithm is optimal Well applied to large and small samples	Anomaly detection with Byzantine fault tolerance is to be done
Single/SVM	Thaseen et al. [12]	SVM	Chi-square feature	31 features (1, 2, , 4, 5, 6, 10, 11, 12, 13, 14, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41)	NSL-KDD	Detects unknown pattern, Parameter tuning is done	Kernel methods can be used for parameter optimization
Single/SVM	Safaldin et al. [13]	SVM	Modified gray wolf optimization	12 features	NSL-KDD	Find optimal feature	Other classifiers can be used

(continued)

Table 2 (continued)

Classifier type/baseline	Author	Classification technique	Feature selection algorithm	No. of features and feature no	Dataset used	Merits	Demerits
Single/Decision tree	Eesa et al. [14]	Decision tree	Cuttlefish optimization algorithm and decision tree	41, 35, 30, 25, 29, 15, 10, 5 features	KDD Cup 99	Produce the optimal subset of features	CFA as a rule generator
Single/Clustering	Zhong et al. [15]	ICLN, SICLN	–		KDD Cup99	Tolerant to missing or delay labels deals with both labeled and unlabeled data	Better estimation method is required to improve efficiency
Hybrid/SVM + K-Means	Al-Yaseen et al. [16]	Multilevel hybrid SVM + ELM modified K-means algorithm	–	–	10% KDD dataset	Detect known and unknown attacks	Cannot classify new attacks efficiently
Hybrid/SVM + Naïve Bayes	Gu et al. [17]	SVM + naïve Bayes	–	–	UNSW-NB15 dataset NSL-KDD Kyoto 2006 +	High-quality data is achieved	Not applicable for multiclass classification
Hybrid	Hajisalem et al. [18]	ABC and AFS algorithms, Fuzzy C-means clustering (FCM)	Correlation-based feature selection (CFS)	NSL-KDD 6 features (6, 4, 5, 12, 29, 27) UNSW-NB15 6 features (46, 45, 23, 24, 47, 43)	NSL-KDD UNSW-NB15	Optimizing rapidity and accuracy	

Table 3 Performance metrics comparison of different machine learning approaches for intrusion detection

Author	Approaches	Dataset used	Performance metrics					
			Accuracy	False alarm/positive rate	True positive/recall/detection rate	Precision	True –ve rate	False –ve rate
Wang et al. [9]	CAI	10% KDD	99.91 ± 0.02	0.10 ± 0.03	± 0.08	99.9 ± 0.01		
	CAI	10% KDD-DoS attack	99.97 ± 0.01	0.10 ± 0.03	99.99 ± 0.00			
	CAI	NSL-KDD	98.92 ± 0.10	0.92 ± 0.11	98.63 ± 0.29	98.33 ± 0.20	1.37 ± 0.29	
	CAI	2% KDD binary	99.88 ± 0.03	0.27 ± 0.14	99.92 ± 0.03	99.93 ± 0.04	99.73 ± 0.14	0.08 ± 0.03
	CAI	UNSW-NB15 binary	82.74	36.46	98.41	76.78	63.54	1.59
Chiba et al. [10]	A12_MinMax_Hrule1_ActS	KDD Cup 99	98.66	1.13	98.59	99.62	98.87	1.41
	A34_Zscore_Hrule1_ActS		99.10	1.60	99.33	99.47	98.41	0.67
Zhao et al. [11]	SVM-RPFMI	DoS (KDD)	99.772	0	99.99	99.55		
		USR (KDD)	96.19	0.37	65.233	95.139		
		R2L (KDD)	91.077	9.907	10.835	99.403		
		Kyoto 2006+	97.749	1.788	97.285	98.196		
Thaseen et al. [12]	SVM	NSL-KDD	98	0.13				

(continued)

Table 3 (continued)

Author	Approaches	Dataset used	Performance metrics					
			Accuracy	False alarm/positive rate	True positive/recall/detection rate	Precision	True –ve rate	False –ve rate
	multiclass SVM	NSL-KDD	96.01 (normal) 95.8 (probe) 99.87 (DoS) 96.37 (R2L) 76.92(U2R)					
Safaldin et al. [13]	SVM	NSL-KDD	96	0.03	96			
Eesa et al. [14]	CFA (no. of features = 10)	KDD Cup 99	92.837	3.9	92.051			
Zhong et al. [15]	ICLN	KDD 99	99.58		99.59	98.59		
	SICLN		99.66		99.60	98.92		
Al-Yaseen et al. [16]	hybrid SVM + ELM	10% KDD	95.75	1.87	95.17			
Gu et al. [17]	SVM + Naïve Bayes	UNSW-NB15	93.75	7.33	94.73			
		CICIDS2017	98.92	3.00	99.46			
		NSL-KDD	99.36	0.54	99.25			
		Kyoto 2006+	98.58	2.62	99.73			
Hajisalem et al. [18]	ABC-AFS method	NSL-KDD	99	0.01	99			
	ABC-AFS method	UNSW-NB15	98.9	0.13	98.6			

5.1 Single Classifier

Wang et al. [9] applied equality constrained optimization-based ELM (C-ELM), an approach proposed by Huang et al. that detects intrusion. They also proposed an adaptively incremental learning strategy, namely construction with adaptive increments (CAI) which derives the finest count of hidden neurons. Their approach eliminates the computation of weights from the scratch when the numbers of neurons are increased as suggested in C-ELM approach. Their approach overcomes the drawback of C-ELM which caused wastage of time during computation. They conducted their experimental work on KDD-DoS dataset 10% KDD dataset, NSL-KDD dataset, and UNSW-NB15. They used the traditional method of converting categorical values into numeric for preprocessing. They have done comparison of their algorithm with few approaches such as Tan's, Lin's, Hu's, Singh's, SVM, Xu's, and MLP and shown improvement in accuracy, recall, false alarm rate, precision, time, false negative rate, and specificity. The authors suggested that concept drift can be used as a future work.

Chiba et al. [10] introduced optimal anomaly network intrusion detection system (ANIDS) approach based on BPNN. They adapted a learning algorithm, namely back propagation to develop a new architecture. They utilized modified Kolmogorov–Smirnov correlation-based filter (CBF) algorithm and information gain algorithm for dimensionality reduction. The authors build 48 IDSs by combining the classifiers. Their proposed ANIDS have four modules, namely feature selection, data preprocessing, normalization, and detection. They considered performance metrics such as false alarm rate, detection rate, F-measure, ability to avoid false classification (AUC) to choose the best two IDSs. They employed the dataset KDD CUP 99 for their experimental study. The comparative analysis of their proposed IDS was done with several techniques. Their approach showed performance improvement with regard to detection rate, F-score, accuracy, score, and lower false alarm rate. As their future work, they may improve the performance of IDS using an optimization algorithm that uses momentum term and learning rate as parameters.

Zhao et al. [11] developed a novel algorithm that utilizes FB feature selection based on MI called the RPFMI algorithm. In their proposed algorithm, they considered three factors, namely redundancy among features, the relationship among candidate features and classes, and the impact among selected features and classes in order to increase relevancy, and reduce redundancy among features. They used Kyoto 2006+ and KDD Cup 99 datasets in their experiment. The accuracy rate on the DoS data is 99.772%, USR data is 96.19%, and R2L data is 91.077% which is better than all other compared algorithms. The Kyoto 2006+ dataset achieves the highest accuracy of 97.749% when compared to other algorithms. As a future work, the authors suggested to use the proposed RPFMI algorithm with Byzantine fault tolerance to detect anomaly.

Thaseen et al. [12] proposed a feature selection (chi-square) and SVM (multiclass) model for intrusion detection by adapting over fitting constant (C) and gamma (γ) as parameters to optimize the RBF kernel. They used the dataset NSL-KDD for their experimental works. Their algorithm showed high true positive rate and low false positive rates when compared with other traditional approaches.

Safaldin et al. [13] developed an enhanced intrusion detection system (IDS) which used modified binary gray wolf optimizer for feature selection and SVM classifier for classification. They varied the number of wolves to find the exact number of wolves. They used NSL-KDD dataset as benchmark to compute accuracy, detection rate, and processing time. The seven wolves GWOSVM-IDS outperformed existing algorithms.

Eesa et al. [14] developed a feature selection model based on the cuttlefish optimization algorithm (CFA) and decision tree classifier. CFS is used to produce the best feature subsets, and decision tree is used to improve the quality of the created feature subsets. The proposed model is evaluated using KDD Cup 99 dataset. This algorithm yields better true positive rate, and accuracy, and lower false positive rate when a maximum of 20 features are chosen. They suggested using CFA as a rule generator for IDS.

Zhong et al. [15] proposed two new clustering algorithms for network intrusion detection. One of the algorithms is unsupervised algorithm, namely the improved competitive learning network (ICLN), and the other is supervised improved competitive learning network (SICLN) to detect network intrusion. The authors have done comparative analysis of performance of the proposed algorithms with both SOM and K-means. The datasets used for their experimental work are the KDD 99, vesta transaction data, and iris data. Their experimental results showed that ICLN achieved similar accuracy when compared with other unsupervised clustering algorithms. But, SICLN performs better than other algorithms in solving classification problems using clustering approaches.

5.2 *Hybrid Classifier*

Al-Yaseen et al. [16] introduced a multilevel hybrid model that uses both SVM and ELM. They also introduced a modified K-means algorithm that builds high-quality training datasets. Their approach showed improved performance than multilevel SVM and multilevel ELM. They used 10% KDD dataset for their work. In their proposed work, they used the equivalent numerical attributes for the symbolic ones, then they normalized data to [0, 1], and the instances of 10% KDD training dataset are separated into five categories such as normal, probe, DoS, R2L, and U2R.Then, they applied modified K-means on each separated category and trained both SVM and ELM with these those training datasets. Finally, testing is done with these datasets. They achieved an overall accuracy of 95.75%, true positive rate of 95.17%, and false positive rate of 1.87. As a future extension, the authors recommended to construct an efficient model to classify new attacks with better performance.

Gu et al. [17] proposed a hybrid classifier based on SVM and naive Bayes feature embedding. They utilized naive Bayes technique for feature enhancement and SVM for classification. The experiments are done using the datasets, namely UNSW-NB15, NSL-KDD, Kyoto, and CICIDS2017. Their experiments have shown an accuracy of 93.75% on UNSW-NB15 dataset, 98.92% accuracy on CICIDS2017 dataset, 99.35% accuracy on NSL-KDD dataset, and 98.58% accuracy on Kyoto 2006+ dataset.

Hajisalem et al. [18] proposed a novel hybrid classification approach by combining both artificial bee colony (ABC) and artificial fish swarm (AFS) algorithms. They split the training dataset and eliminated the irrelevant features by applying fuzzy *C*-means clustering (FCM) and CFS techniques. They used CART technique to generate If–Then rules which distinguished the normal and anomaly records for the selected features. The authors trained the proposed hybrid method through the generated rules. They used the datasets UNSW-NB15 and NSL-KDD for their experimental work. They achieved false alarm rate of 0.01% and detection rate of 99%. In their proposed method, they have computed the computational complexity and time.

6 Literatures in Deep Learning for Intrusion Detection

Based on learning techniques, ML algorithms can be classified as shallow learning and deep learning. Algorithms with few layers are known as shallow learning which is better suited for less complex datasets. The emerging technique which uses more layers of neural network is referred as deep learning which is used for complex target function and larger datasets. Table 4 shows the comparison of different deep learning algorithms with classifier type, datasets used, performance metrics, and techniques used for comparison. The performance metric comparison for different deep learning approaches discussed in Table 4 is tabulated in Table 5.

6.1 Deep Auto-encoders

Farahnakian et al. [19] developed a deep auto-encoder method to improve the performance of IDS. Their DAE model extracts important features from training data by utilizing a nonlinear activation function, namely sigmoid function. To avoid over fitting and local optima, they pre-trained their model using a greedy layer-wise unsupervised learning algorithm. A softmax classifier is used to denote the preferred outputs (normal or attack type). They used the dataset 10% of KDD Cup 99 their experimental work. The results are done in two scenarios, namely binary classification and multiclassification. In binary classification scenario, the detection rate is 95.65%, false alarm is 0.35, and accuracy is evaluated as 96.53%. In multiclassification scenario, the detection rate is 94.53%, false alarm is 0.42, and accuracy is evaluated as 94.71%. They suggested sparse deep auto-encoders as an approach to enhance the detection efficiency.

Shone et al. [20] introduced a technique for unsupervised feature learning, namely non-symmetric deep auto-encoder (NDAE). They also proposed stacked NDAEs as a classification model that does not use a decoder. The benchmark dataset used for their

Table 4 Comparison of different deep learning algorithms for intrusion detection

Baseline	Author	Technique	Datasets used	Techniques compared	Merits	Demerits
DAE	Farahnakian et al. [19]	Deep auto-encoder (DAE)	KDD CUP'99	DBN, auto-encoder + DBN	Avoid overfitting and local optima	Sparsity constraints not imposed
	Shone et al. [20]	Non-symmetric deep auto-encoder (NDAE)	KDD CUP'99 and NSL-KDD	DBN	Training time less	Do not handle zero-day attack
BM	Zhang et al. [21]	RBM + SVM, RBM + DBN	KDD Cup 99	–	Unsupervised learning is used for feature extraction training time less	
RNN	Yin et al. [22]	RNN-IDS	NSL-KDD	NB tree, random tree, J48, naïve Bayes, random forest, MLP, RNN, SVM	Applicable for binary and multiclass classification	Training time is not reduced
	Kim et al. [23]	LSTM-RNN	KDD Cup 99	GRNN, PNN, RBNN, K-NN, SVM, and Bayesian	Detection rate high	Do not detect U2R, and FAR is to be improved
	Su et al. [24]	BLSTM	NSL-KDD	RNN, DNN, DBN, LSTM, CNN, BLSTM, and BAT	Automatically learn the key features and efficient anomaly detection	–
CNN	Ho et al. [25]	CNN	CICIDS2017	Hierarchical, WISARD, forest PA, J48, LIBSVM, FURIA, random forest, MLP, and naive Bayes	Storage and computation overhead are less and detect innovative attacks	Automated methods can be used to solve class imbalance issues

(continued)

Table 4 (continued)

Baseline	Author	Technique	Datasets used	Techniques compared	Merits	Demerits
DNN	Roy et al. [26]	DNN	KDD Cup 99	SVM	High R2 value so more accurate	–
	Kasongo et al. [27]	FFDNN+ Wrapper method	UNSW-NB15 and AWID	Random forest, SVM, naïve Bayes, decision tree, and K-NN	Applicable for wired and wireless networks	Detection rates of individual classes is not done
	Devan et al. [28]	DNN	NSL-KDD	Logistic regression, SVM, and naive Bayes	Prevent overfitting and faster detection	Not applicable for multiclass classification

Table 5 Performance metrics comparison for different deep learning techniques for intrusion detection

Author	Approaches	Dataset used	Accuracy	Detection rate	False alarm/positive rate	True positive/recall	Precision	False negative
Farahnakian et al. [19]	DAE-IDS	KDD 99	94.71	95.65 (binary) 94.53 (multi-)	0.35 (binary) 0.42 (multi-)	94.42	–	
Shone et al. [20]	S-NDAE	KDD Cup	97.85	97.85	2.15	–	99.99	
	S-NDAE	NSL-KDD	85.42	85.42	14.58	–	100	
Zhang et al. [21]	RBM + DBN	KDD 99	97.160		0.480			3.610
	RBM + SVM		96.310		0.400			4.520
Yin et al. [22]	RNN(KDDTest+)	NSL-KDD	81.29					
	RNN(KDDTest-21)		64.67					
Kim et al. [23]	LSTM-RNN	KDD 99	96.93	98.88	10.04			
Su et al. [24]	BLSTM (KDD test+)	NSL-KDD	84.25	97.50 (normal) 87.55 (DoS) 44.25 (R2L) 20.95 (U2R) 85.76 (probe)	25.70 (normal) 1.52 (DoS) 0.91 (R2L) 0.09 (U2R) 1.15 (probe)			
	BLSTM (KDD test-21)	NSL-KDD	69.42					
Ho et al. [25]	CNN	CICIDS2017	99.78%					
Roy et al. [26]	DNN	KDD CUP 99	99.99					
Kasongo et al. [27]	DNN	UNSW-NB15	94.03 (Full) 92.38 (reduced)					
		AWID binary	98.69 (Full) 99.66 (reduced)					
Devan et al. [28]	DNN	NSL-KDD	97					

evaluation are KDD 99 and NSL-KDD. Their evaluation results show that for KDD 99 dataset, an accuracy of 97.85%, precision of 99.99%, recall 97.85, F-score of 98.15, and false alarm of 2.1 are achieved. With regard to 5-class NSL-KDD classification, an accuracy of 85.42%, precision of 100%, recall 85.42, F-score of 87.37, and false alarm of 14.58 are achieved. With regard to 13-class NSL-KDD classification, an accuracy of 89.22%, precision of 92.97%, recall 89.22, F-score of 90.76, and false alarm of 10.78 are achieved. Their result achieved better accuracy, detection rate, and precision, and reduced training time. As their future work, they have suggested to handle zero-day attack and apply their suggested model to real-world backbone network traffic.

6.2 Boltzmann Machine (BM)

Zhang et al. [21] analyzed the performance and characteristics of deep learning in two hybrid algorithms, namely RBM with SVM and RBM with DBN. They have done their experimental study using KDD cup 99 dataset. The performance metrics used for their evaluation are accuracy, testing time, false negative rate, and false alarm rate. They compared their hybrid algorithms with other traditional algorithms and found that DBN performs better in metrics accuracy and speed. RBM-SVM achieved an accuracy of 96.31%, and RBM-DBN achieved an accuracy of 97.16% when compared with the traditional PCA-BP that attained an accuracy of 92.26%.

6.3 Recurrent Neural Networks (RNN)

Yin et al. [22] introduced a deep learning method, namely RNN-IDS for intrusion detection. The performance study is done using binary classification and multiclass classification. They used NSL-KDD dataset for evaluation. The proposed approach is compared with those of J48, random forest, ANN, and SVM. They reported that RNN-IDS gives better accuracy and that its performance is better in both multiclass and binary classification.

Kim et al. [23] applied long short-term memory (LSTM) architecture for training IDS model in RNN. They normalized all instances from 0 to 1 before using the training dataset. They used input vector with 41 dimensions and output vector with 4 dimensions. LSTM architecture is applied to the hidden layer. For their experiment, they used batch size of 50, time step size of 100, and epoch of 500. They used an optimizer, namely stochastic gradient decent (SGD) and softmax at output layer. Mean squared error (MSE) is used as the loss function. In their first experiment, they analyzed hyper-parameter values and found that hidden layer size and learning rate will produce the best performance in their second experiment. The optimal hidden layer size and learning rate are 80 and 0.01, respectively. KDD Cup 1999 dataset is used for their validation. Their approach achieves an accuracy of 96.93% and detection rate of 98.88% which is better than other compared approaches.

Su et al. [24] proposed variation of BAT model with multiple convolution layers, namely BAT-MC. They utilized BLSTM for traffic classification and attention mechanism to retrieve the key feature data. Attention mechanism conducts feature learning on sequential data composed of data package vectors. The obtained feature information is reasonable and accurate. The benchmark dataset used is NSL-KDD dataset. The experimental results show that the performance of BAT-MC is better than the traditional methods.

6.4 Convolutional Neural Networks (CNN)

Ho et al. [25] developed an IDS with CNN classifier. The dataset used is CICIDS2017 which includes innovative attacks. Their model has shown better detection rate for 10 classes of attacks among 14 and has been used to train and validate the proposed model. The issues found in the dataset are missing value, imbalanced class, and scattered presence. They solved these issues and created a customized database, namely α-Dataset after preprocessing. Their model performed well in terms of metrics accuracy, detection rate, false alarm rate, and training overhead.

6.5 Deep Neural Network (DNN)

Roy et al. [26] accessed the functionality of the classifier (DNN) for validating several attacks that cause intrusion. KDD Cup 99 dataset is used for their validation. They compared their work with support vector machine (SVM). They used rectifier and softmax activation functions. Their experimental results of DNN showed a better performance in accuracy when compared with SVM.

Kasongo et al. [27] used extra trees algorithm for wrapper-based feature extraction and feedforward deep neural network (FFDNN) to develop wireless IDS. UNSW-NB15 and the AWID intrusion detection datasets are used for their experimental study which includes both binary and multiclass types of attacks. A feature vector of 22 attributes is used in UNSW-NB15. For binary and multiclass classification, an accuracy of 87.10 and 77.16% is achieved. A feature vector of 26 attributes is used in AWID. For binary and multiclass classification, an accuracy of 99.66 and 99.77% is achieved.

Devan et al. [28] proposed a method for network intrusion detection which used XGBoost feature selection and deep neural network (DNN) for classification. To optimize the learning rate, they used Adam optimizer and softmax classifier. The dataset used for their experiment is NSL-KDD dataset. Their method has shown improved performance in metrics accuracy, precision, recall, and F1-score.

7 Observations and Future Directions

Regarding datasets, most researchers have done their research on KDD Cup and NSL-KDD datasets. But, Brugger [29] claims that there is problem with this dataset. Therefore, researchers can use UNSW-NB 15, AWID, Kyoto, and CICIDS 2017 datasets for their research. Also, they can do performance analysis using some real traffic datasets. There is huge demand for real-time data set for intrusion detection.

The comparative chart illustrating accuracy of different machine learning algorithms based on different datasets is shown in Fig. 3.

Among all compared machine learning algorithms, hybrid classifier detects attack more accurately when compared with single classifiers. It is observed that classifier with feature selection algorithm shows better detection of attacks. Moreover, the performance of hybrid classifier is better when feature selection algorithm is used. Most of classifier's performance is not better when all features are used. Therefore, feature selection plays a major role in attack detection. Hence, researchers can think of developing the best algorithm for feature selection. Also, the performance of classifiers varies among datasets. So, a better IDS can be developed to sort out this issue.

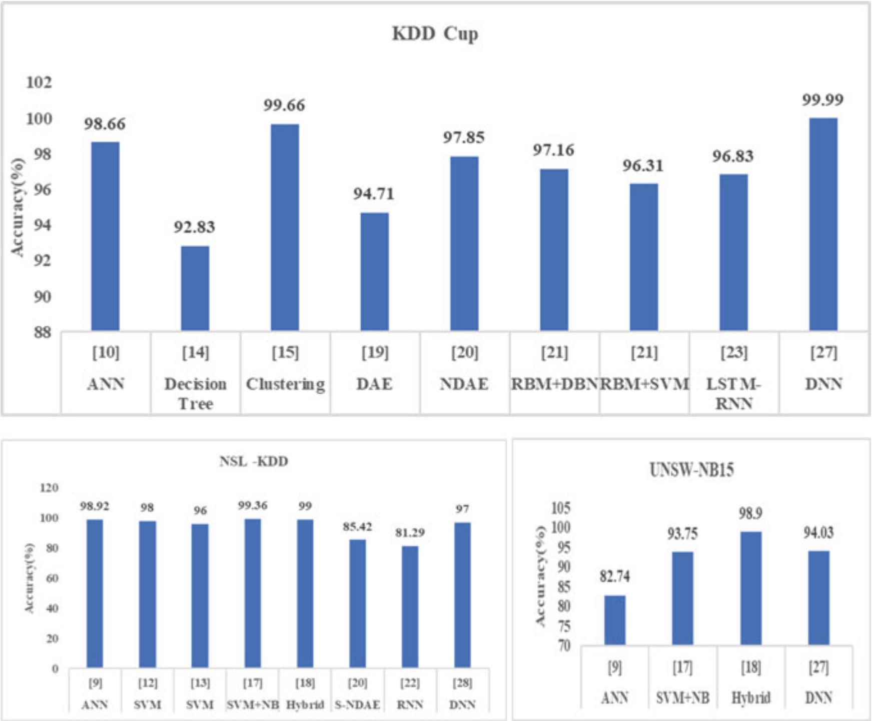


Fig. 3 Comparison of different machine learning algorithms based on accuracy

Among all deep learning algorithms discussed, deep neural networks achieved better accuracy and detection rate. But, some other deep learning approaches such as convolutional neural networks and reinforcement learning can be applied to detect attacks.

8 Conclusion

A taxonomy of different machine learning algorithms used for intrusion detection is discussed. The IDS developed based on machine learning and deep learning algorithms is analyzed. Machine learning algorithms are analyzed based on the classifier type either single or hybrid. Feature selection methods incorporated with machine learning algorithms are also discussed. Machine learning algorithms that used feature selection techniques have shown better accuracy. Deep learning models deal with huge input data. Deep learning IDS has shown better performance in terms of accuracy and running time. GPU-enabled deep learning algorithms can perform execution faster. Future directions to detect intrusion using machine learning algorithm are also discussed.

References

1. Cyber attacks, ALERT: DNS hijacking activity (2019). Online https://www.ncsc.gov.uk/alerts/alert-dns-hijacking-activity
2. C.F. Tsai, Y.F. Hsu, C.Y. Lin, W.Y. Lin, Intrusion detection by machine learning: a review. Exp. Syst. Appl. **36**(10), 11994–1200 (2009)
3. A. Buczak, E. Guven, A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun. Surv. Tutorials 99 (2015)
4. P. Mishra, V. Varadharajan, U. Tupakula, E.S. Pilli, A detailed investigation and analysis of using machine learning techniques for intrusion detection. IEEE Commun. Surv. Tutorials (2018)
5. M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani, A detailed analysis of the KDD CUP 99 data set, in *IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009)* (2009), pp. 1–6
6. N. Moustafa, J. Slay, The evaluation of network anomaly detection systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Inf. Secur. J. **25**(1–3), 18–31 (2016)
7. C. Kolias, G. Kambourakis, A. Stavrou, S. Gritzalis, Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset. IEEE Commun. Surv. Tutor **18**(1), 184–208 (2015)
8. I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion traffic characterization, in *ICISSP 2018—Proceedinggs of 4th International Conference on Information Systems Security and Privacy* (2018), pp. 108–116
9. C.R. Wang, R.F. Xu, S.J. Lee, C.H. Lee, Network intrusion detection using equality constrained-optimization-based extreme learning machines. Knowl. Based Syst. (2018)
10. Z. Chiba, N. Abghour, K. Moussaid, A. El, M. Rida, A novel architecture combined with optimal parameters for back propagation neural networks applied to anomaly network intrusion detection. Comput. Secur. (2018)

11. F. Zhao, Applied sciences a filter feature selection algorithm based on mutual information for intrusion detection (2018)
12. S. Thaseen, A. Kumar, Intrusion detection model using fusion of chi-square feature selection and multi class SVM. J. King Saud Univ. Comput. Inf. Sci. **29**(4), 462–472 (2017)
13. M. Safaldin, M. Otair, L. Abualigah, Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks. J. Ambient Intell. Humaniz. Comput. (2020)
14. A.S. Eesa, Z. Orman, A. Mohsin, A. Brifcani, Expert systems with applications a novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. Exp. Syst. Appl. 1–10 (2014)
15. J. Zhong, A. A. Ghorbani, Neurocomputing Improved competitive learning neural networks for network intrusion and fraud detection. Neurocomputing **75**(1), 135–145 (2012)
16. W.L. Al-Yaseen, Z.A. Othman, M.Z.A. Nazri, Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Exp. Syst. Appl. **67**, 296–303 (2017)
17. J. Gu, S. Lu, An effective intrusion detection approach using SVM with naïve Bayes feature embedding. Comput. Secur. **103** (2021)
18. V. Hajisalem, S. Babaie, A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection. Comput. Netw. **136**, 37–50 (2018)
19. F. Farahnakian, J. Heikkonen, A deep auto-encoder based approach for intrusion detection system, in *International Conference on Advanced Communications Technology* (2018), pp. 178–183
20. N. Shone, T.N. Ngoc, V.D. Phai, Q. Shi, A deep learning approach to network intrusion detection. IEEE Trans. Emerg. Top. Comput. Intell. **2**(1), 41–50 (2018)
21. X. Zhang, J. Chen, Deep learning based intelligent intrusion detection (2017)
22. C. Yin, Y. Zhu, J. Fei, X. He, A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access 21954–2196 (2017)
23. J. Kim, J. Kim, H. Le, T. Thu, H. Kim, Long short term memory recurrent neural network classifier for intrusion detection (2016)
24. T. Su, H. Sun, J. Zhu, S. Wang, Y. Li, BAT: deep learning methods on network intrusion detection using NSL-KDD dataset. IEEE Access 29575–29585 (2020)
25. S. Ho, S. Jufout, S. Al, K. Dajani, M. Mozumdar, A novel intrusion detection model for detecting known and innovative cyberattacks using convolutional neural network. IEEE Open J. Comput. Soc. **2**, 14–25 (2021)
26. S.S. Roy, A. Mallik, R. Gulati, M.S. Obaidat, P.V. Krishna, A deep learning based artificial neural network approach for intrusion detection, in *Mathematics and Computing. ICMC 2017. Communications in Computer and Information Science* ed. by D. Giri, R. Mohapatra, H. Begehr, M. Obaidat, vol 655. (Springer, Singapore, 2017)
27. S.M. Kasongo, Y. Sun, A deep learning method with wrapper based feature extraction for wireless intrusion detection system. Comput. Secur. **92** (2020)
28. P. Devan, N. Khare, An efficient XGBoost–DNN-based classification model for network intrusion detection system. Neural Comput. Appl. 12499–12514 (2020)
29. T. Brugger, KDD Cup ‘99 dataset (Network Intrusion) considered harmful (2007)

Improving Text Classifiers Through Controlled Text Generation Using Transformer Wasserstein Autoencoder

C. Harikrishnan and N. M. Dhanya

Abstract Training good classifiers on imbalanced dataset has always been a challenge, especially if the classifier has to work with textual data. Natural language is one such area where there are abundant imbalanced datasets such as spam filtering, fake news detection, and toxic comment classification. Techniques for generating synthetic data like synthetic minority over-sampling technique fail to train effective classifiers. This paper proposes a technique for generating controlled text using the transformer-based Wasserstein autoencoder which helps in improving the classifiers. The paper compares the results with classifiers trained on data generated by other synthetic data generators. Furthermore, the potential issues of the proposed model for training classifiers are discussed.

Keywords Imbalanced data · Natural language generation · Recurrent neural network · Text classification · Transformers · Wasserstein autoencoder

1 Introduction

A dataset is said to imbalanced when there is a skew in class proportions. This skew is reflected on to the classifiers as when they are trained on skewed data, and their results are also skewed toward the class which has a higher proportion. There are different approaches to balance the data such as oversampling, undersampling, using techniques like Synthetic Minority Over-sampling TEchnique (SMOTE) [13] to generate synthetic data and assigning weights to classes [18]. But each of those approaches has its downsides. There have been impressive works on spam detection and fake news classification in the last few years [1, 14, 17] which showed how well deep learning works better than traditional machine learning algorithms.

C. Harikrishnan (✉) · N. M. Dhanya
Amrita Vishwa Vidyapeetham, Amritanagar, Ettimadai, Tamil Nadu 641112, India
e-mail: cb.en.p2aid19012@cb.students.amrita.edu

N. M. Dhanya
e-mail: nm_dhanya@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_8

Generative models have shown great improvements in the past few years. These models can learn the distribution of the original data and generate samples from that distribution. In the domain of natural language, variational autoencoder (VAE) was proven to be effective [3] than generative adversarial networks [6] (GAN). This is because the architecture of GAN is not suitable for generating discrete samples, as the text is generated by non-differentiable operation like sampling. The VAE architecture used recurrent neural networks for encoder and decoder. The encoder encodes the text into a continuous space, and the decoder takes this latent vector and generates back the input text. The very same architecture with few modifications was used to perform controlled text generation [8], and this model was able to obtain meaningful sentences by restricting the sentence length and better accuracy with sentiment attributes.

After the introduction of the transformer architecture which was a simpler architecture solely based on attention mechanism [16] .Transformers proved to be much better at producing results on natural language tasks such as summarization, question answering, and machine translation. This resulted in most of the RNN-based systems moving from RNN-based architecture to transformers. Naturally, this included generic text generation [12] and controlled text [5, 9] using transformers.

The novelty of this paper is a transformer-based Wasserstein autoencoder which is used for controlled text generation which in turn is used to train a classifier.

2 Related Work

2.1 Strategies for Balancing Datasets

- **Oversampling**: It is the process where the data belonging to the minority class is replicated randomly to match the number of instances in the majority class. The disadvantage of this approach is that it can lead the model to overfit on the training data
- **Undersampling**: It is the process where the instances belonging to the majority class are removed to match the count of instances in the minority class. There is a chance of losing important information which will lead to poor generalization of the model.
- **SMOTE**: SMOTE is used for generating synthetic data for the minority class. These instances are generated by interpolating the points between nearest neighbors. While SMOTE has shown some promise in numerical datasets, but it does not work well with text data
- **Navo Minority Oversampling Technique (NMOTe)**: NMote [4] modifies the original SMOTE algorithm where the data points are synthesized by percentage, where this percentage is a multiple of 100. This was shown to be superior to SMOTE in binary classification problems.

2.2 Wasserstein Autoencoder

Wasserstein autoencoder (WAE) [15] uses the same architecture as variational autoencoder [10], while VAE uses Kullback–Leibler divergence for minimizing the distance between the prior and the posterior distribution WAE replaces this by using a discriminator network that assigns a score how much does the posterior distribution resembles the prior. This is achieved by a min-max game played by the encoder and the discriminator. The discriminator is trained on the following objective function

$$\frac{\lambda}{m}\Sigma_i^m \log(D_\gamma(z_i)) + \log(1 - D_\gamma(\hat{z_i})) \tag{1}$$

The above objective is maximized by performing a gradient ascent. The encoder is trained on the following objective function which is to be minimized.

$$\frac{1}{m}\Sigma_i^m c(x_i, G_\theta(\hat{z_i})) - \lambda \cdot \log(D_\gamma(\hat{z_i}) \tag{2}$$

2.3 Transformers

The transformer is a type of architecture that works with sequence to sequence tasks. This architecture owes its performance to the self-attention mechanism to understand the weightage of each word in the sentence. The self-attention mechanism is further enhanced using multi-head attention, where there are h number of heads and each head performs the self-attention operation. This helps in interpreting the different meanings of a single sentence. The self-attention operation can be expressed in the following mathematical expression.

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{(QK^T)}{\sqrt{d_K}})V \tag{3}$$

2.4 Decoding Strategies

When generating text from a model, the diversity of the generated text depends on the decoding strategies used. Some of the decoding strategies are as follows:

- **Greedy Decoding** This is one of the simplest decoding strategies. While generating the text, the next word is chosen by picking the word with the highest probability. This process goes on until the maximum number of words or the end of sentence tag is encountered.

- **Topk Sampling** This decoding strategy takes top k probabilities and samples a word from it. This strategy helps to introduce words that do not come up often in sentences.
- **Softmax with Temperature**: Here, a parameter T for temperature is used to manipulate the output probabilities of the model [7]. The value of T is used to divide the probabilities before the exponential operation in softmax.

$$\text{softmax}(x)_i = \frac{e^{\frac{y_i}{T}}}{\Sigma_j^N e^{\frac{y_j}{T}}} \tag{4}$$

3 Proposed Work

To ascertain the results, two imbalanced natural language datasets were chosen:

1. COVID Fake News Dataset [2]
2. Spam Identification [11]

The COVID dataset has a total of 10201 headlines of which 9727 headlines are real news and 474 headlines are fake. In the spam or ham dataset, there are a total of 5572 mail subjects of which 4825 are not spam and 747 are spam mails. The proportion of data can be seen in Fig. 1

For preprocessing text data, after tokenization, the words that do not repeat more than once were replaced by the <unk> token. The numbers were replaced by <num> token. To train the transformer WAE start of sentence <sos> and end of sentence <eos> tokens were appended at the beginning and end of the text.

The transformer WAE was trained using teacher forcing with negative log-likelihood as the reconstruction loss and divergence loss determined by the discriminator. The training is performed on the complete dataset. The architecture of the transformer WAE is shown in Fig. 2. After training, the encoder is used to train a

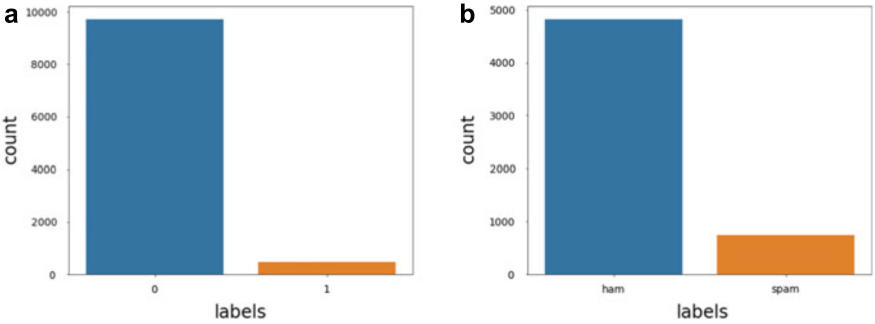


Fig. 1 Proportion of imbalance in **a** COVID dataset **b** Spam dataset

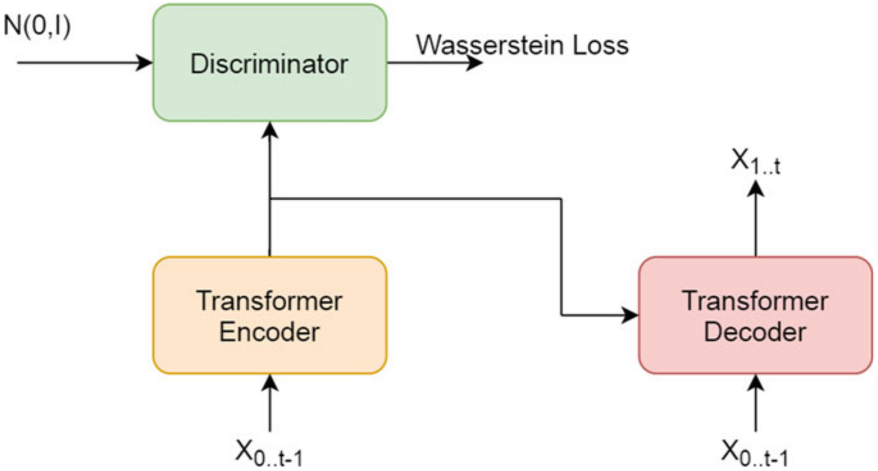


Fig. 2 Transformer Wasserstein autoencoder

controlling network. The dataset is downsampled, and the balanced dataset is encoded into a latent representation by the encoder. The controlling network is trained to distinguish between the latent representation by class. The setup for training controller network is shown in Fig. 3

For controlled text generation, a random noise z is sampled from a unit Gaussian distribution. This noise is then passed to the controller network C_z. The label output by the controller network is set as the expected output, and a cross-entropy loss is calculated with respect to noise z. The noise z is then updated by gradient descent after scaling it by a factor of η

$$y = \text{argmax}(C_z(z)) \tag{5}$$

$$L = \frac{1}{m}\Sigma_1^m y \cdot \log(C_z(z)) \tag{6}$$

$$z = z + \eta \cdot \frac{\mathrm{d}L}{\mathrm{d}z} \tag{7}$$

The noise is updated iteratively to convert it to the value that the controller is confident about. The combination of decoding strategy is used while generating the text, and the softmax with temperature was applied over the probabilities, and greedy decoding was implemented. Topk sampling was used to find a replacement word when a $< unk >$ token was encountered.

The classifier is first trained on a balanced downsampled dataset for 100 epochs. Later the same classifier is fine-tuned for five epochs on the dataset that is a combination of the downsampled dataset with the generated dataset. The model trained on

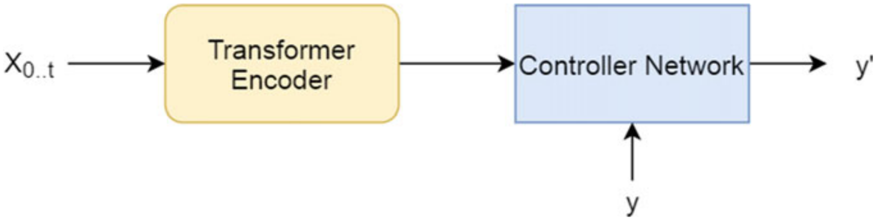


Fig. 3 Training controller network

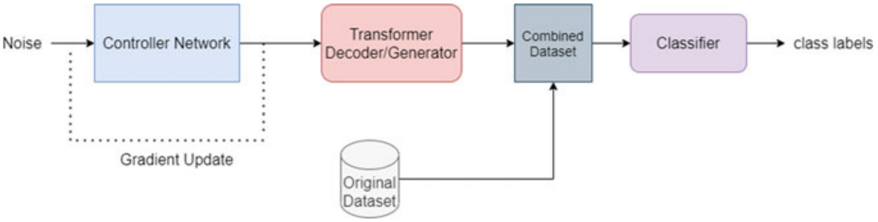


Fig. 4 Training the classifier

downsampled on the same downsampled dataset to prevent the model from forgetting the original data.

Two different types of classifiers were chosen to validate this approach—a normal RNN classifier and an RNN with an attention decoder classifier. The previous works on spam and fake emails [1, 14] were able to achieve best performance on RNN classifier. This is the reason for conducting the experiment with same architecture. The embedding dimension was chosen as 64, the hidden size was set as 256. As for the architecture, two layers of LSTM were stacked together. For the second classifier, one LSTM encoder and two LSTM decoders were part of the architecture. A scaled dot product was performed on the output of the encoder and output of the first decoder. The results of the dot product were input to the second decoder.

The setup shown in Fig. 4 was modified by swapping out the transformer model with RNN variational autoencoder's (RNN VAE) decoder. This was done to compare how transformer-based text generation affects the classifier.

Lastly, another set of classifiers were trained on a combination of the real data and synthetic data generated by SMOTE. This was done to understand how much does the SMOTE helps in text classification.

4 Results

For comparing the different models explained in the previous section, accuracy and F1 score were chosen as the metrics. The models were tested on validation set which was not the part of the training set. The validation set was taken such that there is no significant skewness among the proportion of class instances.

For the COVID fake news dataset, after downsampling and balancing the data, there were 948 data points, which was split in a 80:20 ratio for training and testing. The testing set had 84 real news and 106 fake headlines. For the spam identification dataset, downsampling reduced the number of data points to 1493. After the 80/20 split, the testing set had 157 real mails and 142 spam mails.

From the results, it can be inferred that fine-tuning the classifier on the text generated by the transformer-based model produces better results. From Tables 1 and 2, it is evident that SMOTE does not function well with text classification and prevents the classifier from generalizing.

The proposed architecture suffers from two main limitations :-

- The text generated by the generator is dependent upon the random noise input. This way the user has no control over what kind of text will be generated.
- Another issue would be memory consumption. While training the classifier, at least three models are loaded into the memory, i.e., the generator model, the controller network, and the classifier network. If the size of the models is large, it can result in causing an out-of-memory error.

Due to the above-mentioned memory issues, it was difficult to perform the experiment with larger pre-trained models like BERT and RoBERTa.

Table 1 COVID fake news detection results

Model	Accuracy	F1 score
RNN classifier	0.8105	0.8392
RNN with attention	0.8315	0.8446
RNN classifier on SMOTE	0.4315	0.1147
RNN with attention on SMOTE	0.4578	0.1889
RNN classifier trained on RNN VAE	0.7736	0.7902
RNN with attention trained on RNN VAE	0.8157	0.8292
RNN classifier trained on transformer WAE	0.8421	0.8514
RNN with attention trained on transformer WAE	**0.8421**	**0.8543**

Table 2 Spam detection results

Model	Accuracy	F1 score
RNN classifier	0.8695	0.8849
RNN with attention	0.9364	0.9396
RNN classifier on SMOTE	0.4682	0.1928
RNN with attention on SMOTE	0.4715	0.2882
RNN classifier trained on RNN VAE	0.8695	0.8876
RNN with attention trained on RNN VAE	0.9096	0.9184
RNN classifier trained on transformer WAE	0.8862	0.8950
RNN with attention trained on transformer WAE	**0.9397**	**0.9407**

5 Conclusions

This paper proposes a new approach to train better models on an imbalanced dataset. The experiments showed that the existing synthetic data generation techniques such as SMOTE proved to be ineffective in the natural language domain and the proposed approach was able perform quite well compared to other text generation techniques. The effectiveness of the classifier depends upon the random noise given as the input to the model. The proposed architecture is further more memory intensive while training as several models are loaded in the memory together.

References

1. B. Anjali, R. Reshma, V. Geetha Lekshmy, Detection of counterfeit news using machine learning, in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2019* (2019), pp. 1382–1386. https://doi.org/10.1109/ICICICT46008.2019.8993330
2. S. Banik, *Covid Fake News Dataset* (Nov 2020)
3. S.R. Bowman, L. Vilnis, O. Vinyals, A.M. Dai, R. Jozefowicz, S. Bengio, Generating sentences from a continuous space, in *CoNLL 2016 - 20th SIGNLL Conference on Computational Natural Language Learning*, Proceedings (2016), pp. 10–21. https://doi.org/10.18653/v1/k16-1002
4. N. Chakrabarty, S. Biswas, Navo minority over-sampling technique (NMOTe): a consistent performance booster on imbalanced datasets. J. Electron. Inf. **2**(2), 96–136 (2020). https://doi.org/10.36548/jei.2020.2.004
5. S. Dathathri, A. Madotto, J. Lan, J. Hung, E. Frank, P. Molino, J. Yosinski, R. Liu, Plug and play language models: a simple approach to controlled text generation, pp. 1–34 (2019)
6. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020). https://doi.org/10.1145/3422622
7. G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network 1–9 (2015). http://arxiv.org/abs/1503.02531
8. Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, E.P. Xing, Toward controlled generation of text, in *34th International Conference on Machine Learning, ICML 2017*, vol. 4 (2017), pp. 2503–2513
9. N.S. Keskar, B. McCann, L.R. Varshney, C. Xiong, R. Socher, CTRL: a conditional transformer language model for controllable generation, pp. 1–18 (2019)
10. D.P. Kingma, M. Welling, Auto-encoding variational bayes, in *2nd International Conference on Learning Representations, ICLR 2014—Conference Track Proceedings (Ml)* (2014), pp. 1–14
11. B. Klimt, Y. Yang, The enron corpus: a new dataset for email classification research, 217–226 (2004)
12. D. Liu, G. Liu, A Transformer-based variational autoencoder for sentence generation, in *Proceedings of the International Joint Conference on Neural Networks* **2019-July**(July), 1–7 (2019). https://doi.org/10.1109/IJCNN.2019.8852155
13. Y. Mansourifar, W. Shi, Deep synthetic minority over-sampling technique **16**, 321–357 (2020)
14. S. Srinivasan, V. Ravi, M. Alazab, S. Ketha, A.M. Al-Zoubi, S. Kotti Padannayil, Spam emails detection based on distributed word embedding with deep learning. Stud. Comput. Intell. **919**(December), 161–189 (2021)
15. I. Tolstikhin, O. Bousquet, S. Gelly, B. Schölkopf, Wasserstein auto-encoders, 1–20 (2017)
16. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need. Adv. Neural Inf.Proc. Syst. **2017-Decem**(Nips), 5999–6009 (2017)

17. R. Vinayakumar, K.P. Soman, P. Poornachandran, S. Akarsh, *Application of Deep Learning Architectures for Cyber Security* (Springer International Publishing, No. June, 2019)
18. V. Vishagini, A.K. Rajan, An improved spam detection method with weighted support vector machine, in *2018 International Conference on Data Science and Engineering, ICDSE 2018* (2018). https://doi.org/10.1109/ICDSE.2018.8527737

Topological Analysis of Cube Based and Linearly Extensible Networks Using MATLAB

Faizan Nasir, Abdus Samad, and Jamshed Siddiqui

Abstract Speed of Internet is getting faster year after year and the need of high-processing infrastructure becomes essential. In order to satisfy the desired computing power, it is obligatory to improve computing resources particularly. A number of multiprocessor architectures have been proposed to speed up the parallel executing tasks. These topologies classified into cube, star, linear and tree based architectures. This paper presents the properties of cube and linear based networks, and a comparative study is carried out to identify the best choice of topology. Analysis is carried out on the grounds of notable characteristics like degree, diameter, average distance, cost, bisection width, message density and number of links. Results are shown in the form of tables and graphs to explore the conclusion. The article concludes the overall performance variations in both the network families.

Keywords Computing · Interconnection network · Performance analysis · Data centers · Cube architectures · Linearly extensible architectures · Topological properties

1 Introduction

Performance analysis of parallel architecture research provides a better close idea in solving any complex computing problem. A parallel architecture topology is the leading factor in designing of any multiprocessor system, which is described as interconnection network. An interconnection network allows the transportation of data within processors or between processors and memory modules. Many networks have already been proposed which are placed in different classes on their structural binding's scheme [1, 1]. Various architectures are developed in such a way that they deliver better performance in terms of particular parameters. Some topologies

F. Nasir (✉) · J. Siddiqui
Department of Computer Science, Aligarh Muslim University, Aligarh, India

A. Samad
Department of Computer Engineering, Aligarh Muslim University, Aligarh, India
e-mail: abdussamad@zhcet.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_9

focuses on reduction of its average distance, diameter, communication delay and some focuses on making it more fault tolerant and scalable.

Analysis of topological characteristics can lead to get the behavioral performance of the network. Outcome of such study may lead in identification of various better interconnection network performances in terms of various topological parameters. This analysis may further help in the process of designing a parallel architecture system and also in proposing a new interconnection network for high-performance computing. In our study cube based designs such as hypercube, folded hypercube, cross cube, folded cross cube, dual cube, folded, de burjin, necklace hypercube and linearly extensible based architectures like LEC, LCQ, LET, LEΔ are been compared on the bases of variation in their topological properties. Graphs have been drawn using MATLAB programs. For comparison equal nodes are considered for each network.

Paper is divided into 5 sections. Section 1 encapsulates the introduction, while Sect. 2 describes the various parameters that affects the overall performance of the network. Section 3 defines briefly about the literature review and related work about the networks considered in this article. Section 4 clearly explains the comparative analysis on the bases of topological properties for networks discussed in Sect. 3 with the help of tables and graphs. Section 5 concludes the overall results of the analysis.

2 Upshots of Parameter

(a) Footprints of Degree (d):
Degree is the total number of links connected to a node in a graph. As the degree (d) increases, the complexity increases and performance degrades.

(b) Range of Diameter (D):
More would be the diameter more would be the communication latency between two nodes. Furthermore, complexity increases with the rise in diameter.

(c) Expedition of average Distance (a):
The average shortest path of a network, respective sum of all shortest paths between the nodes divided for the total number of nodes. Overall delay increases with the increase in average distance and hence performance degrades with the increase in average distance.

(d) Cost dependency (C):
Cost depends on the product of the two factors degree and diameter. Hence, any of these two factors increases cost gets high values and performance of the network degrades.

(e) Thickness of Message Density (m):
By reducing message density, the wait time or delay in communication can be reduced. Networks having lesser message density lead a better network performance.

(f) Impact of bisection width (B):

Bisection width is the minimum number of links that must be removed to divide the network into two equal halves. The more bisection width indicates that more fault tolerant would be the network.

(g) Nexus of the Nnumber of Links (E):
Overall number of links decides the complexity of the network and delay in communication delay. Complexity increases with the increase in number of links.

3 Related Work

In recent years, many interconnection network architectures have been proposed, few of them have been reviewed and discussed in this section. One of the most famous design is cube based architecture in which nodes increase exponentially. A number of hypercube variants like hypercube, folded hypercube, cross cube, folded cross cube, dual cube, folded dual cube, meta cube, folded meta cube, de burjin and necklace hypercube have been introduced. Hypercube has logarithmic diameter and it is also called as n-cube. Hypercube has high bisection width as 2^{n-1} and have more number of link [3]. In case of folded hypercube (FHC), it is a special case of traditional hypercube, which have some extra links between its nodes. Its bisection width is 2^{n} and diameter is $\ulcorner n/2 \urcorner$ [4]. Cross cube have almost half diameter as compared to folded hypercube. Cross cube has higher number of links. Folded cross cube have extra links called as complimentary links are connected to design folded cross cube by using cross cube. It also have higher number of link [5]. Another network called as dual cube is a special case of meta cube network [6]. It is designed for large multiprocessor systems. Folded dual cube is an extended network of dual cube which is constructed by adding few more link connectivity to the network. It is designed to inherit the properties of dual cube as well as folded hypercube. While folded dual is an extended network of dual cube which is constructed by adding few more link connectivity to the network [7]. It is designed to inherit the properties of dual cube as well as folded hypercube by connecting each node farthest from it. With the similar design meta cube is built for very large parallel systems. It has less number of links in comparison with nodes in the network [8]. It has many desired features like small diameter and in case of folded meta cube it consists desired properties of folded hypercube and meta cube together. It is also designed for very large multiprocessor systems [9]. With a slight different design, a versatile than shuffle exchange (SE) and cube connected cycles (CCC) and it is proven to be more fault tolerant than hypercube different versions is called de burjin network. A network called as necklace hypercube has an array of processors attached with a traditional hypercube network. It is designed to make hypercube architecture scalable and its increase in number of links is less in comparison with increment in number of nodes [10]. It is designed for large multiprocessor systems.

A similar architecture which holds some topological features of cube based architecture is called as star based architecture. Star cube is one of the network which is a combination of star graph and hypercube. It becomes more fault tolerant, vertex symmetric and regular than cube based architecture. It holds shortest path routing and has large number of nodes. A large scale multiprocessor system called as varietal cube (n, m), which is a combination of n-star and m-varietal hypercube holds strong connectivity, recursive structure and partionability [11]. A union of cross cube and star called as star cross cube is another hybrid network which have better degree, diameter and cost.

A different architecture from star and cube called as linear architecture in which nodes are incremented linearly and provide better scalability, less complexity than star and cube based architectures. Some of the networks named as linearly extensible cube, linear crossed cube, linearly extensible triangle and linearly extensible tree. LEC and LCQ are very similarly designed as hypercube architecture but they have improvised drawbacks such as complex extensibility, VLSI layout and high-cost [12]. LET is designed over the idea of binary tree, and it has number of nodes ($n + 1$). It holds linear extensibility property like small number of nodes per extension. Whereas LEΔ is based on triangle based multiprocessor network. It has very simple geometry and uses the concept of isosceles triangle [13]. Nodes effortlessly be incremented by adding links and forming isosceles triangle again by joining the vertices. LEΔ is scalable and less complex.

Various topological parameters have been formulated by various researchers and few parameters are depicted together in Table 1 [14, 14]. On the basis of these formulations, we have calculated the various parameters using MATLAB tool by designing 8 node structure for each network and written optimum program code for the evaluation of parameters for each network.

Average distance and message density are related with each other such that if average distance increases the delay in communication also increases. We have evaluated average distance and message density through following expressions.

$$\text{Average Distance (a)} = \frac{1}{n.(n-1)}.\sum_{i \neq j} d\left(\mathrm{v}_i, \mathrm{v}_j\right)$$

where '$d\left(v_i, v_j\right)$' represents the length of shortest path exists between two vertices, 'n' is the number of nodes in the network.

$$\text{Message Density } (m) = \frac{n \times a}{E}$$

where 'n' is the number of nodes in a network, 'a' is the average distance of a network, 'E' is the number of links in a network.

Table 1 Various parameters of cube based and linearly extensible architectures

Networks	Parameters				
	Degree	Diameter	Cost	Links	Bisection width
Hypercube	n	n	n^2	$n*2^{n-1}$	2^{n-1}
Folded hypercube	$n+1$	$\lceil n/2 \rceil$	$\lceil n/2 \rceil *(n+1)$	$(n+1)*2^{n-1}$	2^n
Cross cube	n	$\lceil (n+1)/2 \rceil$	$n*(\lceil (n+1)/2 \rceil)$	$2^{2\hat{}n}$	2^{n-1}
Folded cross cube	$n+1$	$\lceil n/2 \rceil$	$(n+1)*\lceil n/2 \rceil$	$(n+1)*2^{2n-2}$	2^n
Dual Cube	$(n+1)/2$	$n+1$	$(\mathrm{n}+1)^2/2$	$(\mathrm{n}+1)*2^{n-1}$	2^{n-2}
Folded dual cube	$(\mathrm{n}+3)/2$	$\mathrm{n}-1$	$(\mathrm{n}\text{-}1)*(\mathrm{n}+3)/2$	$(\mathrm{n}+3)*2^{n-2}$	$2^n/2$
Meta cube	$n+1$	2^{1+n}	$(n+1)*2^{1+n}$	$(n+1)*2^{2\hat{}n+n-1}$	$2^{2n}/2$
Folded meta cube	$n+1$	$2n-1$	$(n+1)*(2n-1)$	$2^{2\hat{}n+n}$	$2^{2n}/2*(2^{2n}+n-2)$
De aurjin	4	$\mathrm{Log}_2\mathrm{N}$	$4*\mathrm{Log}_2\mathrm{N}$	–	–
Necklace hypercube	$2n$	$n+k$	$2n(n+k)$	$n2^{n-1}+n2^{n-1}(\mathrm{k}+1)$	2^n
LEC	4	$\lfloor n \rfloor$	$4*\lfloor n \rfloor$	$4n$	n
LCQ	4	$\lfloor \sqrt{N} \rfloor$	$4*\lfloor \sqrt{N} \rfloor$	–	–
LEΔ	$N-1$	$\sqrt{N}$	$\sqrt{N}*(N-1)$	–	–
LET	4	$\sqrt{N}$	$4*\sqrt{N}$	–	$2\log(n-2)$

4 Comparative Analysis

Comparative study with different networks is explained in this section to make overall conclusion. Characteristics defined in Sect. 2 are being calculated and represented by dedicated tables. This section is further divided into 3 sub sections. Sections 4.1 and 4.2 describe the performance of cube and linearly extensible based networks on the bases of their properties defined earlier section. Whereas Sect. 4.3 makes comparative study of those networks which have performed better in Sects. 4.1 and 4.2. With the help of separate graphs drawn for performed better networks, it is easy to understand the analysis more deeply.

Table 2 Evaluation of various topological parameters at 8-node cube based networks

S. No	Network	Average distance	Diameter	Degree	Cost	Bisection width	Message density
1	Hypercube	1.7143	3	3	9	12	1.1429
2	Folded Hypercube	1.4286	2	4	8	16	0.7143
3	Cross Cube	1.5714	2	3	6	12	1.0476
4	Folded Cross Cube	1.4286	2	4	8	16	0.7143
5	Dual Cube	2.2857	4	2	8	8	2.2857
6	Folded Dual Cube	1.7143	3	3	9	12	1.1429
7	Meta Cube	2.2857	4	2	8	8	2.2857
8	Folded Meta Cube	1.7143	3	3	9	12	1.1429
9	De Burjin	1.6429	3	4	12	16	0.8214
10	Necklace Hypercube	1.7778	3	4	9	14	1.10159

4.1 Performance of Cube Based Networks

4.1.1 Average Distance (a)

According to the data shown in Table 2, folded hypercube and folded cross cube have value 1.43 as lowest for average distance which means performance of these networks are better in comparison with others. Whereas dual cube and meta cube have value 2.28 which is largest value among those networks. Hence, dual cube and meta cube performed worst by having largest average distance values.

4.1.2 Diameter (D)

The trend of variations in diameter is depicted in Table 2. It clearly depicts that folded hypercube, cross cube, folded cross cube performed better by having value 2 and dual cube and meta cube have diameter value 4 which means performed worst among the all networks.

4.1.3 Degree (d)

In Table 2, it is clearly represented that dual cube and meta cube have less degree of value 2 than other networks and hence performed better. While worst networks are folded hypercube, folded cross cube, de burjin, necklace hypercube with degree 4.

4.1.4 Cost (C)

In terms of cost cross cube have the less value as 6 and performed best among the networks, whereas de burjin performed worst by having cost as 12 in Table 2.

4.1.5 Bisection Width (B)

In accordance with table, folded hypercube and folded cross cube shown in Table 2 has the highest bisection width as 8 and performed best. While cross cube and meta cube has the lowest value as 2 which means worst performance between all networks.

4.1.6 Number of Edges (E)

Dual cube and meta cube have lowest number of edges/links according to the results shown in Table 2 the value is 8, hence better performance of the networks. Whereas folded hypercube, folded cross cube and de burjin have highest value as 16 that means worst the performance in this case.

4.1.7 Message Density (M)

The results shown in Table 2 shows that folded hypercube and folded cross cube have the lowest message density as 0.71 which means they have lesser message delay. Whereas dual cube and meta cube have the largest value as 2.28 for message density, hence worst performed.

4.2 Performance of Linearly Extensible Networks

4.2.1 Average Distance (a)

According to the results shown in Table 3, LEΔ has value 1.36 as lowest for average distance which means performance of these networks are better in comparison with others. Whereas LET has value 1.78 which is largest value among those networks. Hence, LET performed worst by having largest average distance values.

Table 3 Evaluation of various topological parameters at 8-node linearly extensible based networks

S. No	Network	Average distance	Diameter	Degree	Cost	Bisection-width	Number of edges	Message density
1	LEC	1.4286	2	4	8	8	16	0.7413
2	LCQ	1.4286	2	4	8	8	12	0.7413
3	LET	1.7857	3	3	9	6	11	1.2987
4	LEΔ	1.3571	2	7	14	9	18	0.6032

4.2.2 Diameter (D)

In terms of diameter data in Table 3 clearly depicts that LEC, LCQ, LEΔ performed better by having value 2 and LET have diameter value 3 which means performed worst among the all networks.

4.2.3 Degree (d)

In Table 3, it is clearly represented that LET has less degree of value 3 than other networks and hence performed better. While worst network is LEΔ with degree 7.

4.2.4 Cost (C)

Cost is calculated and shown in Table 3. It is observed that as compared to other networks LEC and LCQ has a cost equal to 8 which is significantly less and proved to be better when calculating complexity and cost. Whereas LEΔ performed worst by having cost as 14 in table.

4.2.5 Number of Edges (E)

LET has lowest number of edges/links according to data in Table 3 have the value as 11, hence better performance of the networks. Whereas LEΔ has highest value as 18 that means worst the performance in this case.

4.2.6 Message Density (M)

In Table 3, LEΔ has the lowest message density as 0.60 which means they have lesser message delay. Whereas LET has the largest value as 1.29 for message density, hence worst performed.

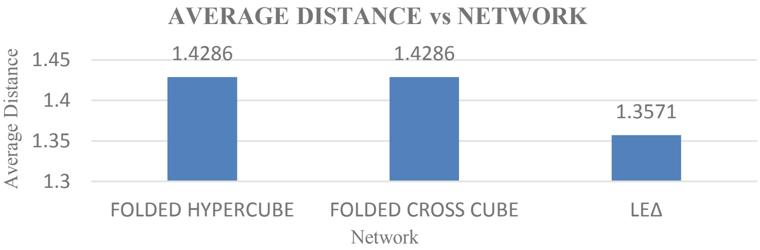


Fig. 1 Performance in terms of average distance

4.2.7 Bisection Width (B)

In accordance with data shown in Table 3, LEΔ has the highest bisection width as 9 and performed best. While LET has the lowest value as 9 which means worst performance between all networks.

4.3 Comparative Analysis of Better Performing Networks

After analyzing the performance of two different class of network, it is revealed that LEC and LCQ networks are out performing among cube based networks. Similarly, when linearly extensible networks are considered it is observed that LEΔ is producing better results in terms of average distance and message density. In this section, a comparative study is carried out of different class of networks for the parameters for which they are performing better. For each calculated property, a graph is made to show the variation between various networks. Separate graphs are produced to clearly distinguish between performed better networks.

4.3.1 Average Distance (a)

Average distance is the average shortest path of a network. The graph shown in Fig. 1 clearly depicts that linearly extensible architecture LEΔ has performed much better over cube based networks in terms of average distance.

4.3.2 Diameter (D)

The diameter of a network is the maximum distance between the pair of nodes. The less is the diameter the less would be the network. According to the graph shown in Fig. 2, every network performed the same better in terms of diameter property of the network.

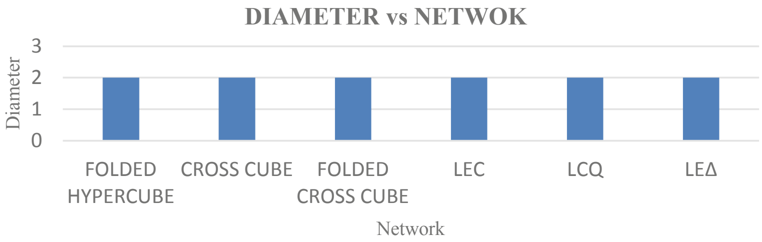


Fig. 2 Performance in terms of diameter

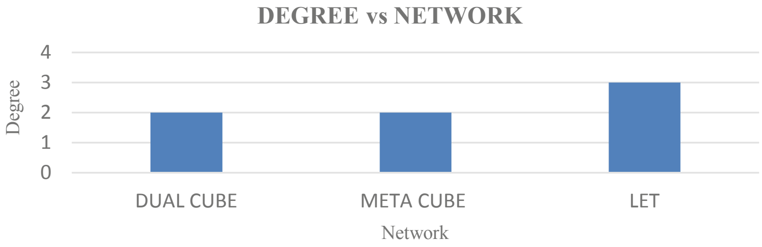


Fig. 3 Performance in terms of degree

4.3.3 Degree (d)

Degree is the number of links/edges associated with a node. More will be the degree more would be the links and hence more complexity of the network. While considering degree shown in Fig. 3 cube based networks dual cube and meta cube have performed much better than linearly extensible network LET.

4.3.4 Cost (C)

Cost factor depends on the product of degree and diameter of the network. As shown in Fig. 4, cross cube has less diameter as 2 and it has degree value 3, hence product

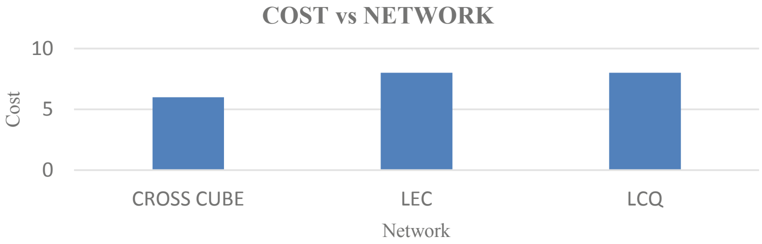


Fig. 4 Performance in terms of cost

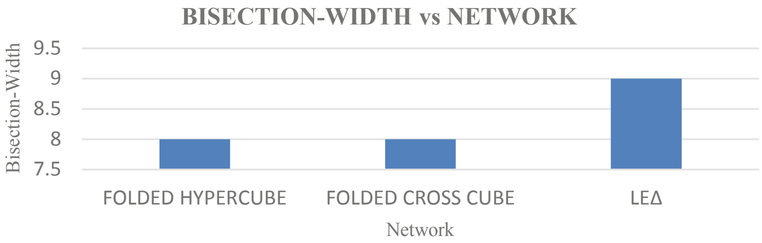


Fig. 5 Performance in terms of bisection width

of these two parameters gives cost as 6. So cross cube have performed better than linear based networks LEC and LCQ in terms of cost.

4.3.5 Bisection Width (B)

Bisection width is very important factor which directly related with fault tolerance of the network. In the graph shown in Fig. 5, LEΔ is more fault tolerant than cube based network folded hypercube and folded cross cube. Hence, linearly extensible network LEΔ comes out to be better network in our study.

4.3.6 Number of Edges (E)

In terms of edges which is just a number of links in any network, dual cube and meta cube have less number of edges than linearly extensible LET network as shown in Fig. 6. Thus, dual cube and meta cube come out to be less complex, while considering nexus of edges.

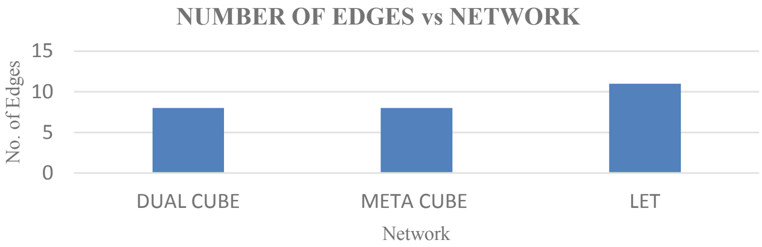


Fig. 6 Performance in terms of number of edges

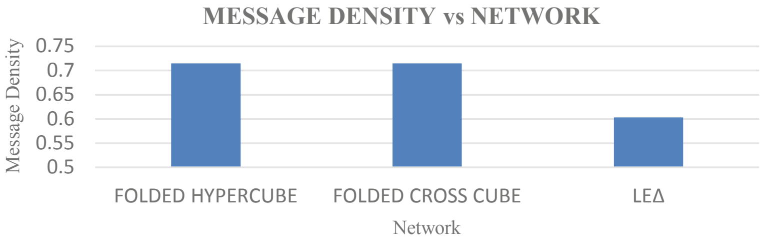


Fig. 7 Performance in terms of message density

4.3.7 Message Density (M)

Message density simply guides about how much communication delay can face a network due to message congestion. In Fig. 7, LEΔ has very less message density in comparison with cube based networks folded hypercube and folded cross cube. Hence, LEΔ is best performed networks dealing with message congestion.

5 Conclusion and Future Work

In this article, various topological properties are discussed in the context of their effect on any network architecture. The topological properties are evaluated for various cube based and linearly extensible networks using MATLAB programs. After complete comparative study of cube based and linearly extensible networks, it is concluded that LEΔ is giving better performance in terms of average distance, bisection width and message density. Meta cube and dual cube performed well in degree, number of edges. While considering cost dual cube comes out to be best network. When diameter is taken into consideration folded hypercube, crossed cube, folded cross cube, LEC, LCQ and LEΔ have performed best.

The comparative analysis concludes that cube based networks are complex in nature and networks having linearly extensible in nature could be used to design efficient interconnection networks that could be used for massively parallel systems. In context with issues, the design of an interconnection networks could be more cost efficient and more fault tolerant. While choosing an interconnection network may influence many topological properties like cost, scalability, complexity and fault tolerance. The present study can be extended by considering large node structures. The scope of study may give a better opinion for designing a high-performance multiprocessor data center/server.

References

1. Z.A. Khan et al., Topological evaluation of variants hypercube network. Asian J. Comput. Sci. Inf. Technol. **9**, 125–128 (2013)
2. F. Nasir, J. Siddiqui, Comparative analysis of cube and star based networks. Int. J. Comput. Sci. Eng. **6**(11), 51–59 (2018). https://doi.org/10.26438/ijcse/v6i11.5159
3. J.P. Hayes, T. Mudge, Hypercube supercomputers. Proc. IEEE. **77**(12), 1829–1841 (1989). https://doi.org/10.1109/5.48826
4. A. El-amawy, S. Latifi, Properties and performance of folded hypercubes, **2**, January, 31–42 (1991)
5. N. Adhikari, The folded crossed cube: a new interconnection network for parallel systems **4**(3), 43–50 (2010)
6. Y. Li, S. Peng, Dual-cubes: a new interconnection network for high-performance computer clusters
7. N. Adhikari, C.R. Tripathy, Folded dualcube: a new interconnection topology for parallel systems. https://doi.org/10.1109/ICIT.2008.49
8. Y. Li, S. Peng, Metacube—a new interconnection network for large scale parallel systems, **24**(4), 29–36 (2001)
9. N. Adhikari, C.R. Tripathy, Folded metacube: an efficient large scale parallel interconnection network. 2009 IEEE Int. Adv. Comput. Conf. IACC 2009. April 2009, 1281–1285 (2009). https://doi.org/10.1109/IADCC.2009.4809200
10. M. Monemizadeh, H. Sarbazi-Azad, The necklace-hypercube: a well scalable hypercube-based interconnection network for multiprocessors. Proc. ACM Symp. Appl. Comput. **1**, 729–733 (2005). https://doi.org/10.1145/1066677.1066842
11. B. Nag et al., Star varietal cube: a new large scale parallel interconnection Network. Ii, 37–44 (2011)
12. Z.A. Khan, Linear crossed cube (LCQ): a new interconnection network topology for massively parallel system. February, 18–25 (2015). https://doi.org/10.5815/ijcnis.2015.03.03
13. J.M. Islamia, N. Delhi, A Δ -based linearly extensible multiprocessor network. **4**(5), 700–707 (2013)
14. A. Samad, et al., Properties and performance of cube- based multiprocessor architectures **7**(1), 63–78 (2016). https://doi.org/10.4018/IJAEC.2016010105
15. S. Gautam, A. Samad, Properties and Performance of Linearly Extensible Multiprocessor Interconnection Networks. Springer Singapore (2019). https://doi.org/10.1007/978-981-13-2372-0_1

Implementation of Rule Based Testing for Digital Circuits Using Inductive Logic Programming

Shubhangee Kishan Varma, Vanita Agarwal, and Ashok Chandak

Abstract Inductive logic is widely known for implementing a rule based system. The most significant inspiration for utilizing inductive logic programming is that it defeats the llustrative impediments of property estimation learning frameworks. Such frameworks utilize a table based portrayals where the occurrences compare to lines in the table, the ascribes to sections, and for each case, a solitary worth is doled out to every one of the credits. The second inspiration for utilizing ILP is that it utilizes rationale, a declarative langauge. This infers implies that theories are justifiable and interpretable. By utilizing rationale, inductive rationale programming frameworks are likewise ready to utilize foundation information in the acceptance cycle. This system is additionally helpful for AI, specifically for defining furthermore, creating ideas like the covers connection, over-simplification, and refinement operators, see likewise the rationale of consensus. With the help of ILP, rules get induced and induced rules are helpful for testing instead of testbences. This rule mainly represents basic logic for digital circuits. In this paper, authors are implementing optimized inductive logic, i.e., Metarule. This metarule is optimized from inductive logic programming. The main aim of the paper is to simplify the digital testing process with the help of metarule. Using metarule we can implement the circuit logic with some simple code. This code based designs make a quick and simple solution for digital circuit testing. With the help of rule based testing, verification get faster and more reliable. In this paper, authors are implementing rule based testing for digital circuits using ILP and also comparing project file size of verilog and prolog for the same.

Keywords PL—prolog (logic programming language) · IBK—ınformation background knowledge · ILP—ınductive logic programming · MIL—Meta-interpretive learning

S. Kishan Varma (✉) · V. Agarwal · A. Chandak
Department of Electronics and Telecommunication, Cusrow Wadia Institute of Technology, Pune, India

V. Agarwal
e-mail: vsa.extc@coep.ac.in

V. Agarwal
Department of Electronics and Telecommunication, College of Engineering, Pune, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_10

1 Introduction

In this paper, the writers are implementing a digital logic using metarule based rule using ınductive logic programing.

- Inductive logic writing computer programs is the subfield of AI that utilizes first request rationale to address theories and information.
- Since first request rationale is expressive furthermore, revelatory, inductive rationale programming specfically targets issues including organized information and foundation information.
- Inductive logic programming handles a wide assortment of issues in AI, including classification, relapse, bunching, and support learning, frequently utilizing 'updates' of existing propositional AI frameworks, it depends on rationale for information portrayal and thinking purposes.
- Thoughts of inclusion, consensus, and administrators for crossing the space of speculations are grounded in rationale, moreover the logic of over-simplification. Inductive logic programming frameworks have been applied to significant applications in bio-and chemo-informatics, regular language preparing, also, web mining.
- In this paper, authors are implementing rule based testing for digital circuits using ILP. The first step is to implement or design a metarule by which we are able to implement digital circuit logic. The selected metarule is used to implement the digital circuit using inductive logic programming.
- The rule obtained by inductive logic programming used for testing of digital circuits. Rule based testing is a part of formal verification, validation and testing techniques.
- Induction, inference and logical deduction are acts of qualifying conclusions on the basis of background knowledge.
- An argument is true if the phases used to progress from the background knowledge to the conclusion imitate to established rules of inference.
- Metarule learning is an ILP method which uses higher-order metarules to support build invention and learning of recursive definitions. The metarules conclude the structure of allowable rules which in turn defines the assumption space.
- In this paper, we demonstrate that rules can be induced by using ILP and can be used for verification of digital circuits. When this approach is applied for testing of digital circuits, it will be result in effective verification, and hence, we compare. In our experiments, we compare the project size of Verilog and SWI Prolog software.
- The paper is organized as follows: Sect. 2, we introduce inductive logic programming (ILP) with illutration; Sect. 3, we formally describe the implementation platform to simulate ILP codes; Sect. 4, we provide the Metagol library details with meta-interpretive learning (MIL) and procedure to implement the SWI prolog on windows and raspberry pi; Sect. 5, the results of simulations of the test benches on verilog and ILP codes on SWI prolog as well as on raspberri pi; Sect. 6, the comparative result of project file size of prolog code using ILP and verilog test benches is given; Sect. 7, conclusion of work done is written; acknowledgment

2 Inductive Logic Programming

ILP is an assessment zone formed at the intersection purpose of ML and logic programming. ILP structures make predicate depictions from models and establishment data. The models, establishment data and last depictions, are totally portrayed as method of reasoning projects. Before long viable applications regions for ILP systems consolidate the learning of structure-development rules for cure plan, restricted part work assessment design rules, basic assistant desire for protein structure and weakness end rules for satellites [1].

From computational rationale, inductive rationale programming acquires its authentic formalism, its semantical direction, and different grounded strategies. Rather than most different ways to deal with inductive learning, inductive logic writing computer programs is intrigued in properties of derivation rules, in union of calculations, and in the computational intricacy of strategies. Numerous inductive rationale programming frameworks profit by utilizing the consequences of computational rationale. Extra advantage might actually be gotten from utilizing work on end, types and modes, information base refreshing, algorithmic troubleshooting, kidnapping, requirement rationale programming, program combination, and program investigation [2].

ILP expands the hypothesis and practice of computational rationale by exploring acceptance instead of derivation as the essential method of deduction. While present computational rationale hypothesis depicts deductive induction from rationale recipes favorable to ided by the client, inductive rationale programming hypothesis depicts the inductive deduction of rationale programs from cases and foundation information. As such, ILP may add to the act of rationale programming, by giving instruments that help rationale favorable to grammers to create and check programs [2].

ILP is an important method to deal with the issue of finding a great deal of speculations covering positive models and at the same time notwithstanding negative models. It uses first-demand method of reasoning as a uniform depiction for models and hypotheses. The ILP strategy gives us a phase to make a couple of measures that show the association between the pancreatic malady and the related factors. ILP gives a computation to learn hypotheses, imparted in method of reasoning, from a database by tolerating the followings:

1. Foundation data background as a prolog program;
2. Some language specification L portraying the hypotheses;
3. An optional game plan of goals I on commendable theories;
4. A finite set of models (Positive and Negative)

Here, E is the relationship of non-void game plan of ‘positive’ models $E+$ and a great deal of ’negative’ models $E-$. The purpose of an ILP is to find a great deal of rules (H), as a method of reasoning system that spread all the positive models without negative models [3]. ILP has specific focal points than other data mining methodology since it can empower the relationship among individuals and PCs by using establishment data to restrict the interest space and return human-reasonable results, along these lines misusing both the PC's speed and the human's data and capacities [4].

ILP Setup:

First-order clauses learn the theory that is rule using IBK, positive and negative examples [5, 6].

Example:

Learn grandparent (P, Q) relation.

- Background Knowledge:
  - father (dashrath, rama) father(rama,love) father(rama, kush)
  - mother(sita, gita) mother(sita, siya)
  - male(love) male(kush)
  - female(gita) female(siya)
  - parent(P, Q):-father(P, Q).
  - parent(A, B):-mother(A, B).
- Examples
  - Positive examples:
  - grandparent(dashrath, love), grandparent(dashrath, kush)
  - Negative examples:
    grandparent(dashrath, rama), grandparent(rama, love)
- Learning a Rule:
  - Top Down Approaches

    start with most general rule (cover everything)
    specialize the rule until it covers only (mostly) positives
  - Bottom up Approaches

    start with most specific rule (covers a single example)
    generalize the rule as long as it doesn't cover any (to many) negatives [2]
  - Rules are

    grandparent(A, B):-mother(A,B).
    grandparent(A,B):-father(A,B).

3 Implementation Platform

Prolog is a reliable and an informative logic programming language. The prolog is short for programming in logic. Prelog's inheritance recalls the assessment for speculation provers and other modernized determination structures made during the 1960s and 1970s. The surmising system of prolog depends on Robinson's goals rule (1965) along with instruments for separating answers proposed by Green (1968).

These thoughts met up powerfully with the approach of direct goals methods. Express objective coordinated straight goals methods, for example, those of Kowalski and Kuehner (1971) and Kowalski (1974), offered driving force to the improvement of a universally useful rationale programming framework. The 'main' prolog was 'Marseille Prolog' in view of work by Colmerauer (1970). The main point by point depiction of the prolog language was the manual for the Marseille prolog translator (Roussel, 1975). The other significant effect on the idea of this first prolog was that it was intended to encourage normal language preparing [7].

Prolog is the significant case of a fourth era programming language supporting the revelatory programming worldview. The Japanese fifth-generation Computer Project, reported in 1981, embraced prolog as an improvement language, and along these lines concentrated on the language and its capacities. The projects in this instructional exercise are written in 'standard' (University of) Edinburgh Prolog, as indicated in the great prolog course book by writers Clocksin and Mellish (1981,1992). The other significant sort of prolog is the prolog II group of prologs which are the relatives of Marseille prolog. The reference to Giannesini, et al. (1986) utilizes a form of prolog II. There are contrasts between these two assortments of prolog; some portion of the thing that matters is linguistic structure, and part is semantics. Nonetheless, understudies who adapt either sort of prolog can without much of a stretch adjust to the next kind [7].

4 Metagol Library Working

Metagol is an inductive rationale programming (ILP) framework dependent on meta-interpretive learning. Prolog language is used to write metagol and runs with SWI prolog software. Metarules are used as an input to metagol. Metarules characterize the type of provisos allowed in a speculation and along these lines the pursuit space [8]

Meta-interpretive learning (MIL) uses metarules. A MIL student is given as information of two arrangements of sets that address positive and negative instances of an objective idea, IBK depicted as a rationale program, and a bunch of higher-request Horn provisos called metarules (Table 1). A MIL student employments the metarules to develop a proof of the positive models also, none of the negative models, and structures a speculation utilizing the replacements for the factors in the metarules [9]

Table 1 Metarules examples [9]

Name of metarule	Metarule
Ident rule	P(A; B) ⟵ Q(A; B)
Precon rule	P(A; B) ⟵ Q(A); R(A; B)
Postcon rule	P(A; B) ⟵ Q(A; B); R(B)
Chain rule	P(A; B) ⟵ Q(A; C); R(C; B)

For example, given positive and negative instances of the grandparent connection, foundation information with the parent connection, and the metarules in Table 1, a MIL student could utilize the chain metarule with the replacements fP/grandparent, Q/parent, R/parent to actuate the hypothesis [10–12]:

grandparent(A,B) ⟵ parent(A,C), parent(C,B) [9].

Metagol bolsters deciphered foundation information (IBK). IBK is typically used to learn higher-request programs.

For example, one can characterize the ‘map/3’ develop as follows:

‘prolog’.

ibk([map,[],[],_],[]).

ibk([map,[A|As],[B|Bs],F],[[F,A,B],[map,As,Bs,F]]). ‘”.

Given this IBK, metagol will attempt to demonstrate it through meta-translation, and will likewise attempt to get familiar with a sub-program for the molecule ‘[F,A,B]’.

The higher-request and ibk models specifies the same.

Metagol settings [13]

Metagol looks for a theory utilizing iterative developing on the quantity of provisions in the arrangement. You can determine a most extreme number of provisos:

Metagol looks for a speculation utilizing iterative extending on the quantity of conditions in the arrangement.

You can determine a most extreme number of statements:

‘”prolog metagol:max_clauses(Integer). % default 10‘”.

Code body anticipated, rules choice, learning rules are the style to write ILP program. The coding style is quite certain. As demonstrated as follows.

:- use_module(‘metagol’).

%% metagol settings.

%% foundation information.

%% metarules.

%% learning task:-

%% positive models.

Pos = [].

%% negative models.

Neg = [].

learn(Pos,Neg) [13].

As its help just hardly any standard we ought to be explicit to our advanced circuit segment. Author should choose the circuit which should ready to execute utilizing this standard.

Demonstrating advanced circuit in metarule based library.

Attempting to actualize reality table and working in metarule utilizing library based standard [10–12].

Introducing SWI Prolog:

Windows and Mac:

1. Download SWI-Prolog.
2. Windows: We suggest downloading the 64-bit version.

3. Install SWI prolog by adhering to the installer directions.

Beginning SWI prolog on Windows:

1. Select SWI prolog from start Button.
2. Click SWI prolog, this will begin SWI prolog support
3. Click on document and Create new record with.pl augmentation.
4. Save the document and Compile.
5. Consult the accumulated from record counsel.
6. After stacking a program, one can get some information about the program.

Establishment and Simulation of prolog on Raspberry Pi:

1. Installing swi prolog: sudo well-suited get introduce swi prolog
2. Launch the mediator: swipl
3. Start the supervisor: emacs.

5 Results

In this paper, the author simulates the results on prolog language as well as verilog and demonstrate the prolog programming on raspberry pi.

5.1 Simulation Result on Verilog

Figure 1 shows output waveform of EX-OR gate, where ‘a’ and ‘b’ are inputs and ‘out’ is output. It is the output of EX-OR test bench in verilog. The output will pull high when inputs are unequal and vice versa.

Output waveform of EX-NOR gate is shown in Fig. 2, where ‘a’ and ‘b’ are inputs and ‘out’ is output. It is the output of EX-NOR test bench in verilog. The output will

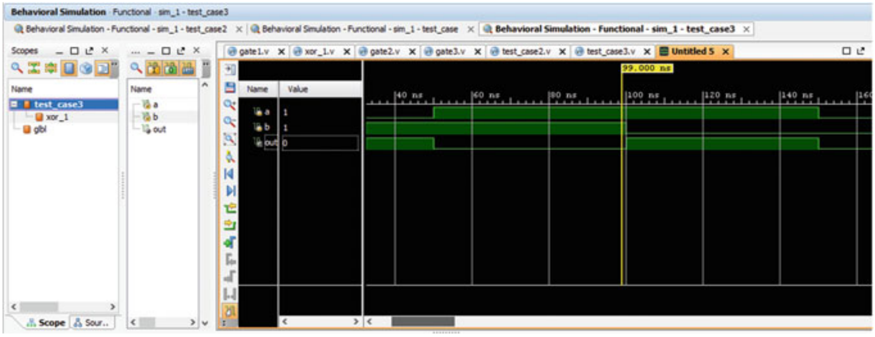

Behavioral Simulation - Functional - sim_1 - test_case3
Behavioral Simulation - Functional - sim_1 - test_case2
Behavioral Simulation - Functional - sim_1 - test_case
Behavioral Simulation - Functional - sim_1 - test_case3
Scopes
Name
test_case3
xor_1
glbl
Name
a
b
out
gate1.v
xor_1.v
gate2.v
gate3.v
test_case2.v
test_case3.v
Untitled 5
Name
Value
a 1
b 1
out 0
99.000 ns
40 ns
60 ns
80 ns
100 ns
120 ns
140 ns
Scope
Sour...

Fig. 1 Output waveform of Ex-OR gate result in Verilog

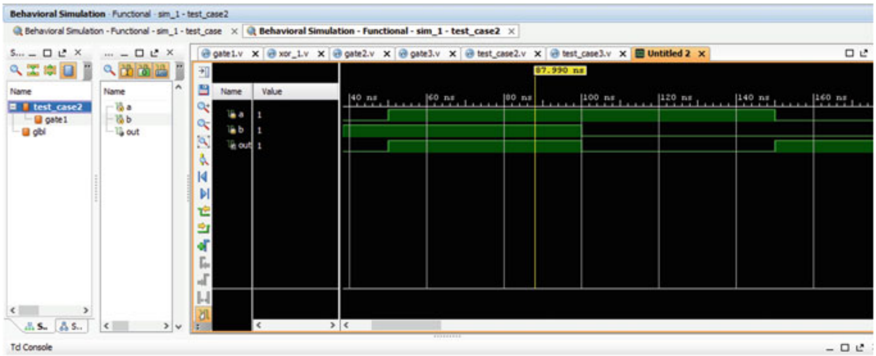


Fig. 2 Output waveform of Ex-NOR gate result in Verilog

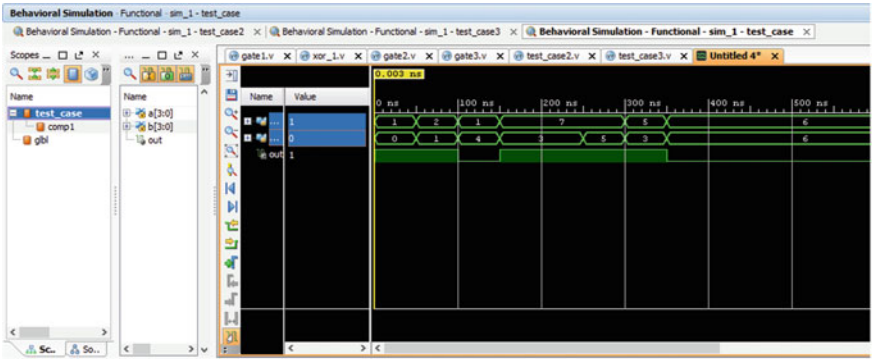

Behavioral Simulation - Functional - sim_1 - test_case
Behavioral Simulation - Functional - sim_1 - test_case2
Behavioral Simulation - Functional - sim_1 - test_case3
Behavioral Simulation - Functional - sim_1 - test_case
Scopes
gate1.v
xor_1.v
gate2.v
gate3.v
test_case2.v
test_case3.v
Untitled 4*
Name
test_case
comp1
glbl
Name
a[3:0]
b[3:0]
out
Name
Value
1
0
out 1
0.003 ns
0 ns
100 ns
200 ns
300 ns
400 ns
500 ns
1
2
1
7
5
6
0
1
4
3
5
3
6

Fig. 3 Output waveform of Comparator gate result in Verilog

pull high when both inputs to the EX-NOR gate is high, and when both inputs are different the output of EX-NOR gate will pull down.

In Fig. 3, we can see that input ‘a’ and ‘b’ are getting compared and ‘out’ is output of comparison. It is the output of comparator test bench in verilog. The output will get pull high if ‘a’ is greater than ‘b’ else it will pull down to low.

5.2 Simulation Result on Raspberry PI

Figure 4 shows prolog editor on raspberry pi. To open prolog editor on raspberry pi, ‘emacs’ command is used.

Figure 5 shows simulation of full adder code in prolog on Raspberry Pi. ‘sum’ is the predicate written to get sum of the inputs and ‘carry’ is the predicate written to get carry of summation.

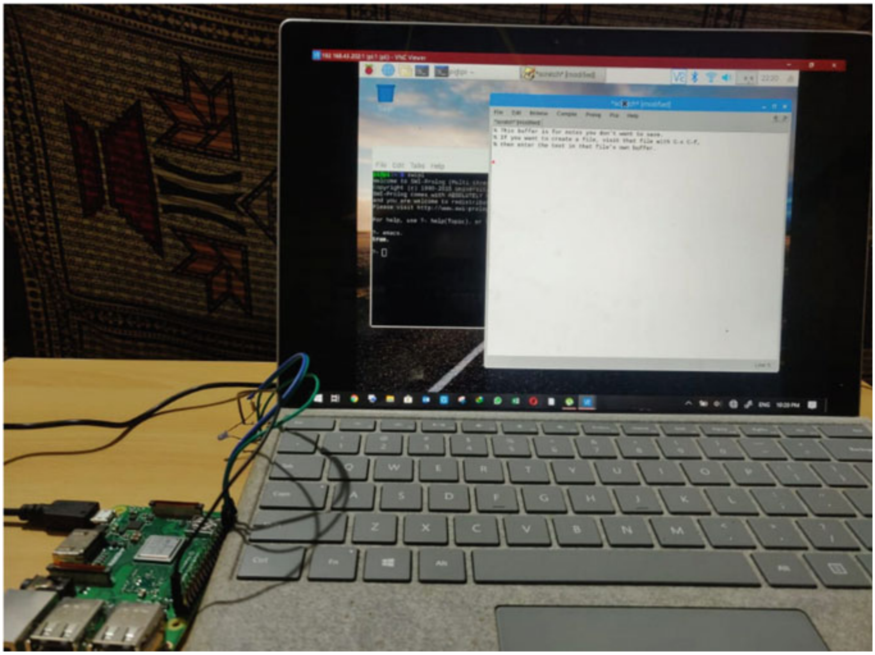


Fig. 4 Prolog Editor on Raspberry Pi

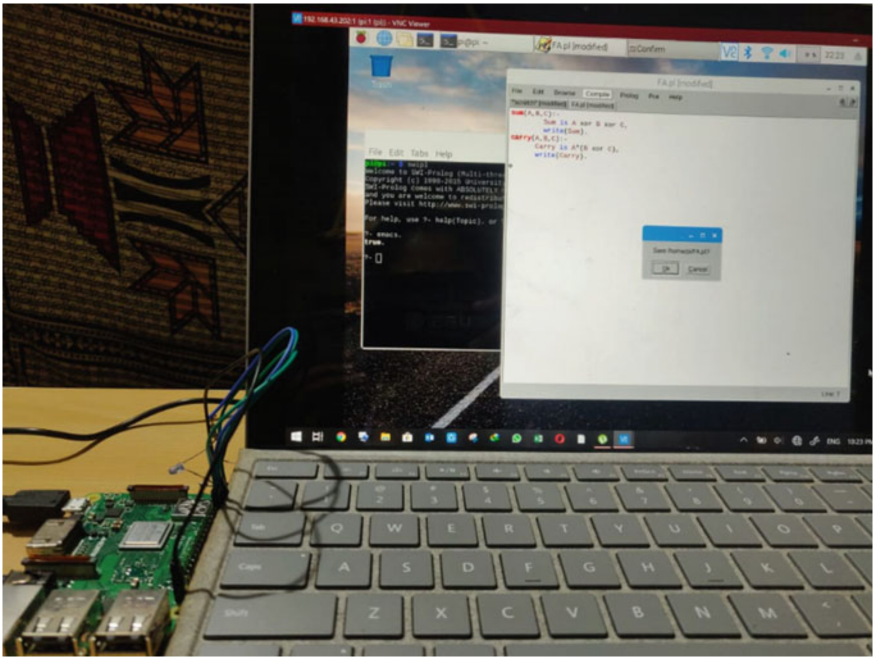
Figure 6 shows result of full adder code in prolog on raspberry pi. ‘sum’ is the predicate with inputs (1, 0, 0) and hence output of sum is 1. ‘carry’ is the predicate with inputs (1, 0, 1) and hence output of carry is 1.

5.3 Simulation Result of ILP Codes on Prolog

Figure 7 shows output of Ex-OR gate. This is the output of Ex-OR gate code written in SWI prolog using ILP. Hence, Fig. 7 shows the EX-OR rule obtained after compiling the code.

Figure 8 shows output of Ex-NOR gate. This is the output of Ex-NOR gate code written in SWI prolog using ILP. Hence, Fig. 8 shows the EX-NOR rule obtained after compiling the code.

Figure 9 shows output of comparator. This is the output of comparator code written in SWI prolog using ILP. Hence, Fig. 9 shows the comparator rule obtained after compiling the code.



```
sum(A,B,C):-
        Sum is A xor B xor C,
        write(Sum).
carry(A,B,C):-
      Carry is A*(B xor C),
      write(Carry).
```

```
pi@pi:~ $ swipl
Welcome to SWI-Prolog (Multi-threa
Copyright (c) 1990-2015 Universit
SWI-Prolog comes with ABSOLUTELY
and you are welcome to redistribu
Please visit http://www.swi-prolo

For help, use ?- help(Topic). or

?- emacs.
true.

?- 
```

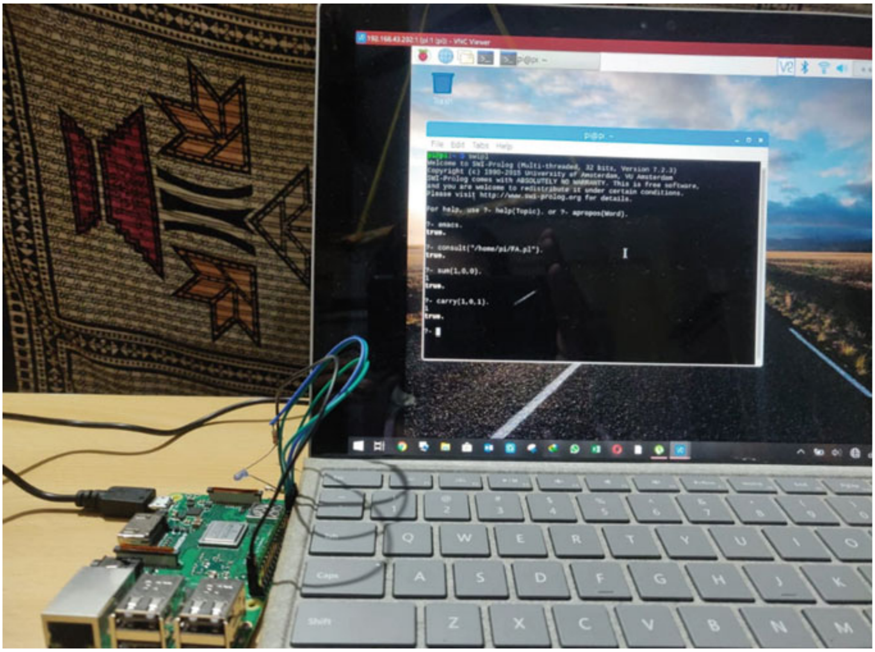
Fig. 5 Full adder Code in Prolog on Raspberry Pi

6 Comparative Result

Table 2 shows Comparative result of project file size of prolog code using ILP and verilog code. Its gives information about file size written in both the software platform. The table shows that the project file size is less in SWI prolog as compared to verilog. Hence, SWI prolog is very good software platform to simulate ILP codes.

7 Conclusion

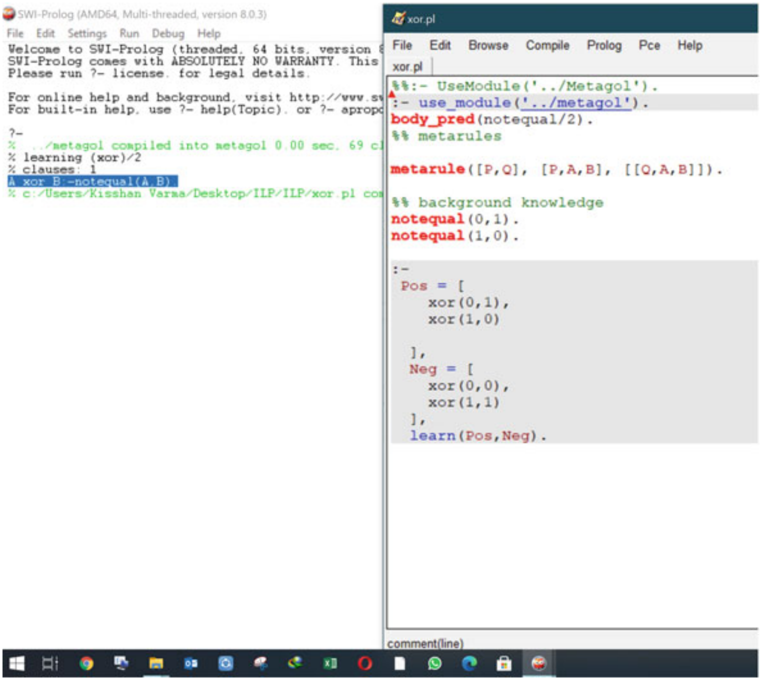
In this paper, writers have structured a framework, which will be utilized for testing of computerized circuits. In customary procedures, they need to execute a ton of scientific models for testing of computerized circuits. This model structure is a tedious and complex procedure to decrease coding overhead. Authors are executing an improved principle based basic arrangement. This incorporates composing a standard which totally speaks to the advanced circuit rationale and impersonates like an inactive computerized circuit. Metarule based execution makes your framework easy to rapidly actualize an answer. ILP programming gives rules so that testing of digital circuits become very effective and easy as compared to other. From table 2, one can



```
pi@pi:~ $ swipl
Welcome to SWI-Prolog (Multi-threaded, 32 bits, Version 7.2.3)
Copyright (c) 1990-2015 University of Amsterdam, VU Amsterdam
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free software,
and you are welcome to redistribute it under certain conditions.
Please visit http://www.swi-prolog.org for details.

For help, use ?- help(Topic). or ?- apropos(Word).

?- emacs.
true.

?- consult("/home/pi/FA.pl").
true.

?- sum(1,0,0).
1
true.

?- carry(1,0,1).
1
true.

?- 
```

Fig. 6 Full adder Code in Prolog on Raspberry Pi Result

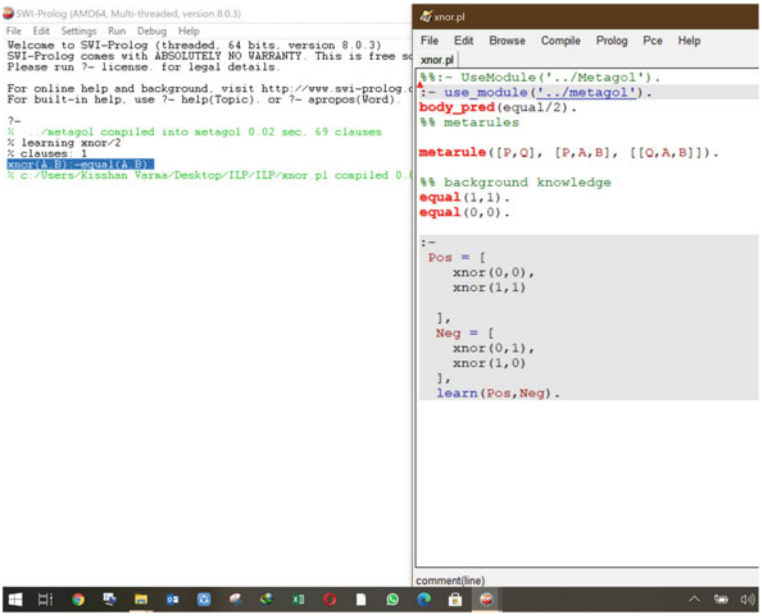
conclude that the project size is less in prolog, so it is faster to execute the file as compare to other.


SWI-Prolog (AMD64, Multi-threaded, version 8.0.3)
File Edit Settings Run Debug Help
xor.pl
File Edit Browse Compile Prolog Pce Help
xor.pl
comment(line)

```
Welcome to SWI-Prolog (threaded, 64 bits, version 8
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This
Please run ?- license. for legal details.

For online help and background, visit http://www.sw
For built-in help, use ?- help(Topic). or ?- aproppo

?-
%  ../metagol compiled into metagol 0.00 sec, 69 cl
% learning (xor)/2
% clauses: 1
A xor B:-notequal(A,B).
% c:/Users/Kisshan Varma/Desktop/ILP/ILP/xor.pl com
```

```
%%:- UseModule('../Metagol').
:- use_module('../metagol').
body_pred(notequal/2).
%% metarules

metarule([P,Q], [P,A,B], [[Q,A,B]]).

%% background knowledge
notequal(0,1).
notequal(1,0).

:-
 Pos = [
    xor(0,1),
    xor(1,0)

  ],
  Neg = [
    xor(0,0),
    xor(1,1)
  ],
  learn(Pos,Neg).
```

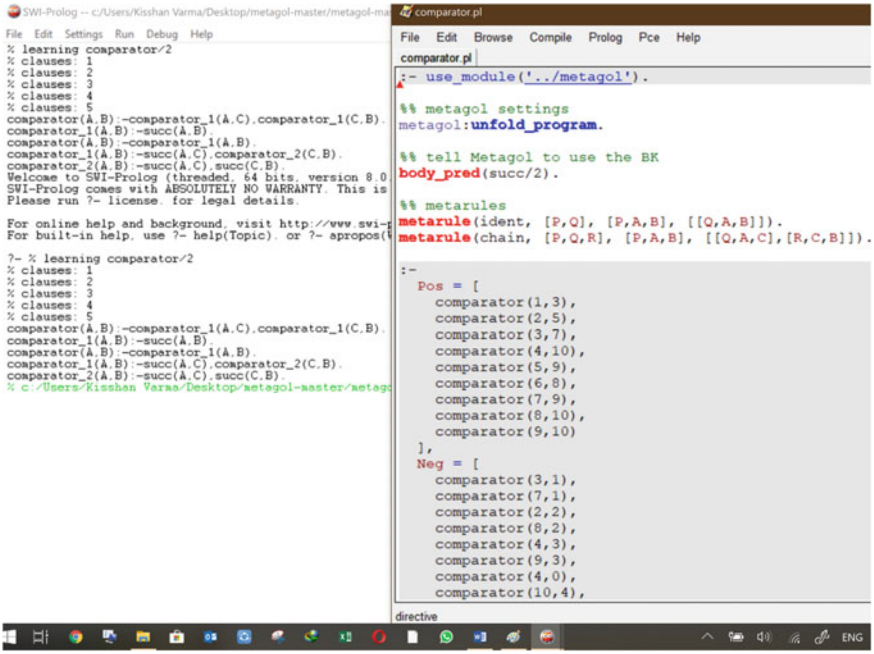
Fig. 7 Result of Ex-OR (Inductive Logic Programming)


SWI-Prolog (AMD64, Multi-threaded, version 8.0.3)
File Edit Settings Run Debug Help
xnor.pl
File Edit Browse Compile Prolog Pce Help
xnor.pl
comment(line)

```
Welcome to SWI-Prolog (threaded, 64 bits, version 8.0.3)
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free so
Please run ?- license. for legal details.

For online help and background, visit http://www.swi-prolog.c
For built-in help, use ?- help(Topic). or ?- apropos(Word).

?-
%  ../metagol compiled into metagol 0.02 sec, 69 clauses
% learning xnor/2
% clauses: 1
xnor(A,B):-equal(A,B).
% c:/Users/Kisshan Varma/Desktop/ILP/ILP/xnor.pl compiled 0.0
```

```
%%:- UseModule('../Metagol').
:- use_module('../metagol').
body_pred(equal/2).
%% metarules

metarule([P,Q], [P,A,B], [[Q,A,B]]).

%% background knowledge
equal(1,1).
equal(0,0).

:-
 Pos = [
    xnor(0,0),
    xnor(1,1)

  ],
  Neg = [
    xnor(0,1),
    xnor(1,0)
  ],
  learn(Pos,Neg).
```

Fig. 8 Result of Ex-NOR (Inductive Logic Programming)


SWI-Prolog -- c:/Users/Kisshan Varma/Desktop/metagol-master/metagol-ma
File Edit Settings Run Debug Help
comparator.pl
File Edit Browse Compile Prolog Pce Help
comparator.pl
directive
ENG

```
% learning comparator/2
% clauses: 1
% clauses: 2
% clauses: 3
% clauses: 4
% clauses: 5
comparator(A,B):-comparator_1(A,C),comparator_1(C,B).
comparator_1(A,B):-succ(A,B).
comparator(A,B):-comparator_1(A,B).
comparator_1(A,B):-succ(A,C),comparator_2(C,B).
comparator_2(A,B):-succ(A,C),succ(C,B).
Welcome to SWI-Prolog (threaded, 64 bits, version 8.0
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is
Please run ?- license. for legal details.

For online help and background, visit http://www.swi-p
For built-in help, use ?- help(Topic). or ?- apropos(W

?- % learning comparator/2
% clauses: 1
% clauses: 2
% clauses: 3
% clauses: 4
% clauses: 5
comparator(A,B):-comparator_1(A,C),comparator_1(C,B).
comparator_1(A,B):-succ(A,B).
comparator(A,B):-comparator_1(A,B).
comparator_1(A,B):-succ(A,C),comparator_2(C,B).
comparator_2(A,B):-succ(A,C),succ(C,B).
% c:/Users/Kisshan Varma/Desktop/metagol-master/metago
```

```
:- use_module('../metagol').

%% metagol settings
metagol:unfold_program.

%% tell Metagol to use the BK
body_pred(succ/2).

%% metarules
metarule(ident, [P,Q], [P,A,B], [[Q,A,B]]).
metarule(chain, [P,Q,R], [P,A,B], [[Q,A,C],[R,C,B]]).

:-
  Pos = [
    comparator(1,3),
    comparator(2,5),
    comparator(3,7),
    comparator(4,10),
    comparator(5,9),
    comparator(6,8),
    comparator(7,9),
    comparator(8,10),
    comparator(9,10)
  ],
  Neg = [
    comparator(3,1),
    comparator(7,1),
    comparator(2,2),
    comparator(8,2),
    comparator(4,3),
    comparator(9,3),
    comparator(4,0),
    comparator(10,4),
```

Fig. 9 Comparator (Inductive Logic Programming)

Table 2 Comparative result

Software platform	Prolog inductive logic programming	Verilog
Comparator	Project file Size-11.2 KB	Project file Size-1.1 MB
Ex-OR Gate	Project file Size-10.8 KB	Project file Size-1.1 MB
Ex-NOR Gate	Project file Size-10.8 KB	Project file Size-1.1 MB

Acknowledgements Authors express genuine thanks to late. Mr. A. B. Patki for guiding through all the phases of research. His ability to explore new research areas and hard work always inspired us. Authors express sincere thanks to Dr. S. P. Mahajan, Head of the E & TC Department, College of Engineering, Pune for his valuable assistance, support, inspiration and self-assurance toward us.

References

1. R. Lima, F. Freitas, B. Espinasse, Relation extraction from texts with symbolic rules induced by inductive logic programming, in *2015 IEEE 27th International Conference on Tools with Artificial Intelligence*
2. L.D. Raet dInductive Logic Programming, Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan, 200A, BE - 3001 Heverlee, Belgium
3. K. Inoue, H. Ohwada, A. Yamamoto Inductive logic programming: challenges, in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*
4. R. Lima, B. Espinasse, H. Oliveira, R. Ferreira, L. Cabral, D. Filho, F. Freitas, R. Gadelha An inductive logic programming-based approach for ontology population from the web, in DEXA 2013, Part I, LNCS 8055 ed by H. Decker et al. (Springer-Verlag, Berlin Heidelberg, 2013), pp. 319–326
5. A. Cropper, S.H. Muggleton, Learning higher-order logic programs through abstraction and invention. in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (IJCAI), ed by S. Kambhampati, (New York, NY, USA, 9–15 July 2016, IJCAI/AAAI Press, 2016), pp. 1418–1424
6. A. Cropper, S.H. Muggleton, Learning efficient logic programs. Mach. Learn. **108**(7), 1063–1083 (2019)
7. K. Benkrid, D. Crookes, A. Benkrid, S. Belkaeemi, A prolog—based hardware development environment, in FPL 2002, LNCS 2438, ed by M. Glesner P. Zipf, M. Renovell (Springer-Verlag, Berlin Heidelberg 2002), pp. 370–380
8. A. Cropper, S. H. Muggleton. Metagol system. https://github.com/metagol/metagol (2016)
9. A. Cropper, Forgetting to learn logic programs. AAAI (2020)
10. Cropper, S. Tourret, Derivation reduction of metarules in meta-interpretive learning. ILP (2018)
11. Cropper, S.H. Muggleton, Logical minimisation of meta-rules within meta-interpretive learning. ILP (2014)
12. A. Cropper, S. Tourret Logical reduction of metarule., arXiv:1907.10952 [cs.LG], 2015
13. Cropper, S. Tourret, Logical minimisation of metarules. Machine Learning (2019)

Design of a High-Speed Binary Counter Using a Stacking Circuit

C. Devika and J. P. Anita

Abstract A novel design of a binary counter is introduced in this paper. A 7:3 binary counter is designed using 5-bit and 2-bit stacking circuits, which is further merged and then converted into binary counts that bestow a binary stacking counter. This new design of binary stacking counter circuit does not have XOR gates or multiplexers in its critical path, which would turn out this circuit into a faster and efficient one. This innovative circuit is faster and has a better power efficiency which outperforms the conventional binary counter circuits. Hence, they find applications in deep learning and big data analysis.

Keywords Binary counters · Critical path · Multipliers · XOR gates · FPGA (Field Programmable Gate Array) · Stacking circuit

1 Introduction

Faster and low-power circuits are a pressing priority in the current environment of VLSI domain. There is a high demand in developing low power and high-performance systems in this era, which is used in all digital applications. Power efficiency and latency also affect the overall operational performance of any circuit. Binary counters can be used in multiplier circuits. This multiplier circuits are indispensable part in the field of digital signal processing which is used in various applications for performing various algorithms, filtering, and convolution. It is also used in ALUs, in a variety of arithmetic techniques, and it plays an essential role in signal and image processing [1, 2].

The binary multiplication of numbers is done using partial products. These partial products in the multipliers are the main component, which drain power and increase the latency in most of the high-speed processors. In multipliers [3], column compression is predominantly used to integrate the partial products. A prominent row compression technique is used in Wallace tree [4], or in Dadda tree [5] and

C. Devika · J. P. Anita (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: jp_anita@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_11

enhanced technique in [6]. These methods use full adders, which would function as a counter. By using suitable parallel counter, Dadda brings the scheme of using full adders in multipliers with initiating the notion of parallel counters. An efficient method to implement a multipliers using decomposition logic is explained in [7]. Wallace uses pseudo adders in finding partial products and array multiplier scheme, which is used to implement in 64 bit booth multiplier and Vedic multiplier in [8, 9].

2 Related Works

Binary counters play an integral role in most of the circuit operations. Methods to optimize the performance of binary counter using adder circuits are discussed in [10]. Here, the counters are using full adders and half adders to get the functionality of binary counters. A 7:3 binary counter using CMOS full adders is depicted in Fig. 1. Drawback of this circuit is that more no. of XOR gates contribute ample delay of the circuit. So many faster architectures came into place for better performance. A parallel 7:3 counter is discussed in [11], which is made using (3,2) counters. A counter design to minimize the no. of multiplexers is used in [12], which is an improved design from a conventional XOR counter to minimize the delay which can be done by replacing second XOR with a multiplexer in order to reduce the timing for switching of the transistors. But all these circuits have a greater number of XOR gates or multiplexers in their critical path.

[10] Introduced a binary counter circuit using symmetrical stacking circuit. Stacking circuit is a kind of a circuit, which reduces the number of XOR gates in the critical path of the circuit, which escalate the performance of the counter design.

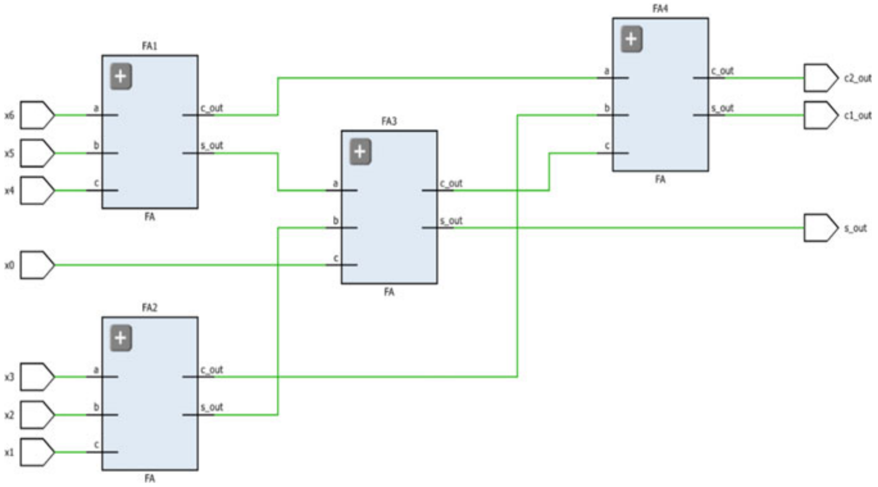


Fig. 1 7:3 binary counter using CMOS full adder circuit

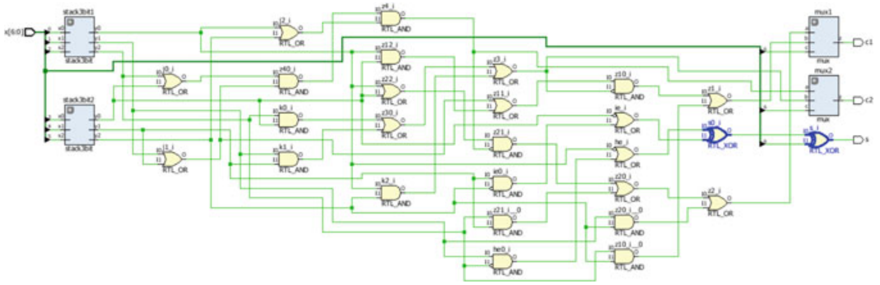


Fig. 2 7:3 symmetrical stacking counter circuit

Major challenges in designing binary counter circuits are in its performance in terms of latency and power of the circuit. Low-power design is a key consideration in all modern design and low-computational speed is a major constraint in the performance of the circuit.

Stacking circuit reduces the numbers of XOR gates. A 7:3 stacking counter circuit is made from 6:3 stacking counter circuits by adding one more additional input and adjusting the output with select lines of two multiplexers as shown in Fig. 2. This design is slightly faster than conventional circuits, but its power consumption is increased due to the addition of multiplexers in the circuit. It does not have adequate performance as that of 6: 3 stacking circuit in [10].

İn order to sweep over the drawback of 7:3 binary symmetrical stacking circuit in [10], a new design of 7: 3 binary stacking counter is proposed in this paper. 2-bit and 5-bitstacking circuits are used for this purpose, in order to build up integrated stacking circuit. 5-bit stacking circuit can take up to five inputs into circuit and will stacks all the input "1" bits together and gives a output where all "1" bits are stacked into leftmost positions. While 2-bit stacking circuit can take 2 input "1" bits and stacks the "1" bits. These 5-bits and2-bit stacking circuits are merged and the pairs are combined in order to make a 7-bit stackingcircuit. Here, seven inputs to the seven-bit stacker circuit, which is formed by merging 5-bit stacker and a 2-bits stacker, gives seven outputs. To make a 7:3 binary stacking counter, this 7-bits stacker circuit output has to be converted into binary values, which will be given as output.

3 Proposed Architecture

A 7: 3 stacking circuits are made of by merging two asymmetrical stacks of 5-bit stacking circuit and a 2-bit stacking circuit.

5-bit stacking circuit. Here, the total numbers of "1" bit in the input as well as in output will be same in number. In output all the "1" bit will be stacked together to leftmost positions and all the"0" bits will follow it. The input is represented as x_0, $x_{1,}$ x_2, $x_{3,}$ $x_{4,}$ and output is represented as y_0, $y_{1,}$ y_2, $y_{3,}$ y_4. Here, y_0 can be obtained by doing OR operations to all the inputs shown in Eq. (1). It will check whether there is any “1” bits in the input and set the y_0. y_1 shown in Eq. (2) will check whether there are more than one “1” bits. y_2 shown in Eq. (3) checks for more than 3 “1” bits. y_3 shown in Eq. (4) will be set, if there are at least 3 “1 bits.” y_4 shown in Eq. (5) will be set if all the inputs are “1″ bits.

$$y_0 = x_0 + x_1 + x_2 + x_3 + x_4 \tag{1}$$

$$\begin{aligned} y_1 = {} & x_0(x_1 + x_2 + x_3 + x_4) + x_1(x_2 + x_3 + x_4) \\ & + x_2(x_3 + x_4) + (x_3x_4) \end{aligned} \tag{2}$$

$$\begin{aligned} y_2 = {} & x_2x_3(x_0 + x_1) + x_0x_1(x_2 + x_3) + x_4x_0(x_1 + x_3) \\ & + x_4x_1(x_2 + x_3) + x_4x_2(x_0 + x_3) \end{aligned} \tag{3}$$

$$y_3 = x_4x_1x_2(x_0 + x_3) + x_4x_0x_3(x_1 + x_2) + x_0x_1x_2x_3 \tag{4}$$

$$y_4 = x_0x_1x_2x_3x_4 \tag{5}$$

2-bit stacking circuit. For a 2-bit stacking circuit $y_0 = 1$, if one of the inputs is “1” bit. y_0 can be obtained by doing OR operation to the input, shown in Eq. (6) and $y_1 = 1$, if both the input are one. It can be obtained by doing AND operation to the input, shown in Eq. (7).

$$y_0 = x_0 + x_1 \tag{6}$$

$$y_1 = x_0x_1 \tag{7}$$

7-bit stacking circuit counter is designed by splitting into two parts which is made by merging these 5-bit and 2-bit stacking circuits. The outputs from this stacking circuit are converted into binary numbers. The output of the counter circuit is denoted by s, c_1, c_2. The outputs from the stacker circuit are represented by h and i. h_o and i_o are the odd parity bits, which can be obtained from Eqs. (8) and (9). The ‘s’ can be found out by doing XOR operation to odd parity of bits from the stacker circuits, shown in Eq. (10).

$$h_\text{o} = h_0\overline{h_1} \tag{8}$$

$$i_\text{o} = i_0\overline{i_1} + i_2\overline{i_3} + i_4 \tag{9}$$

$$s = h_0 \oplus i_0 \tag{10}$$

c_1 is set if the inputs are 2, 3, 6, and 7. To check not more than 3, i_3 and k must be 0, k is a vector which represent more than three bits shown in Eqs. (11), (12), and (13) and $(h_0 + i_0) + h_1\overline{i_0} + (\overline{h_0} + i_1))$ will check whether second bit is set. More than 6 "1" bits can be found by $(h_1h_3 + h_0i_4)$. So all together c$_1$ can be found by Eq. (14).

$$k_2 = h_1 + i_1 \tag{11}$$

$$k_1 = h_0 + i_2 \tag{12}$$

$$k = k_1 + k_2 \tag{13}$$

$$c_1 = (\overline{i_3 k}(h_0 + i_0) + h_1\overline{i_0} + (\overline{h_0} + i_1)) + (h_1h_3 + h_0i_4) \tag{14}$$

c_2 is set if the input is more than 4 "1" bits, which can be found out by Eq. (15)

$$c_2 = h_1i_1 + h_0i_2 + i_3 \tag{15}$$

The proposed 7: 3 counter circuit is shown in Fig. 3. The conventional binary counters have many numbers of XOR gates and multiplexers in the critical path, which degrade the performance of the circuit. This proposed circuit does not have XOR gates and multiplexers, in the critical path. So this proposed architecture gives better performance, in terms of latency and power. One of the drawbacks of this circuit is its wiring complexity.

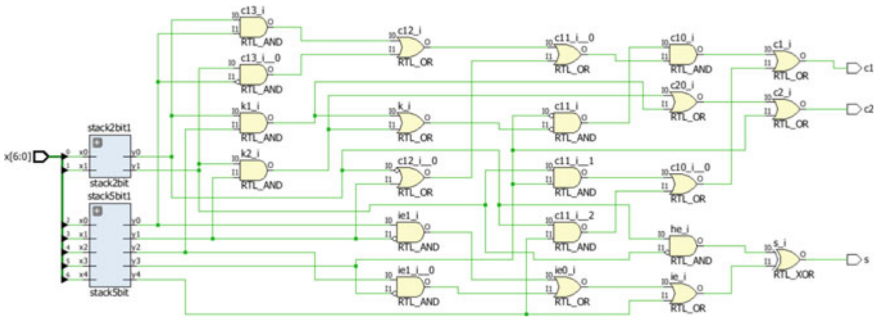


Fig. 3 Proposed architecture of 7:3 binary stacking counter circuit

4 Simulation and Results

The simulation result of proposed 7: 3 stacking counter is shown in Fig. 4. Here, 7 inputs can be given to the circuit and s, c_1, c_2 are taken as output. Circuit counts the no of “1” bits from the input and gives out the binary value of it in the output. The power report of proposed architecture is depicted in Fig. 5, which have a total on chip power of 1.978 W. The area utilization of this circuit is shown in Fig. 6. It depicts the estimation of the capacity of the overall circuit. The proposed architecture functionality is validated using Basys 3 Artix 7 FPGA board shown in Fig. 7, which depicts the outputs of the 7:3 proposed architecture. The LED lights shows the binary output 1–7 when various inputs are given.

For the performance evaluation of the proposed architecture, the latency and power of the circuit are evaluated and compared with the conventional available counter circuit discussed in the literature, shown in Table 1. The proposed architecture does not have any power degrading gates such as XOR or MUX in the critical path. So this circuit outperforms any conventional circuits.

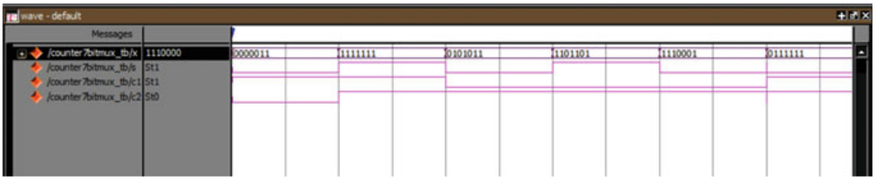
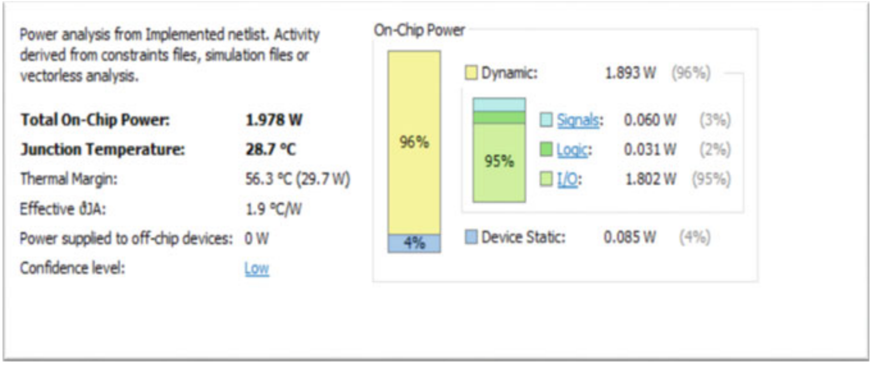


Fig. 4 Simulation result of proposed architecture

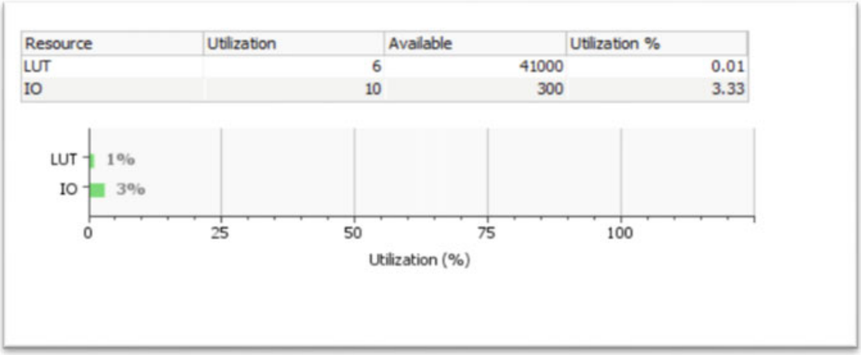

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**1.978 W**
Junction Temperature:	**28.7 °C**
Thermal Margin:	56.3 °C (29.7 W)
Effective ϑJA:	1.9 °C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Low

On-Chip Power

Dynamic:	1.893 W	(96%)
Signals:	0.060 W	(3%)
Logic:	0.031 W	(2%)
I/O:	1.802 W	(95%)
Device Static:	0.085 W	(4%)

Fig. 5 Power report of proposed architecture



Resource	Utilization	Available	Utilization %
LUT	6	41000	0.01
IO	10	300	3.33

Fig. 6 Area report of proposed architecture

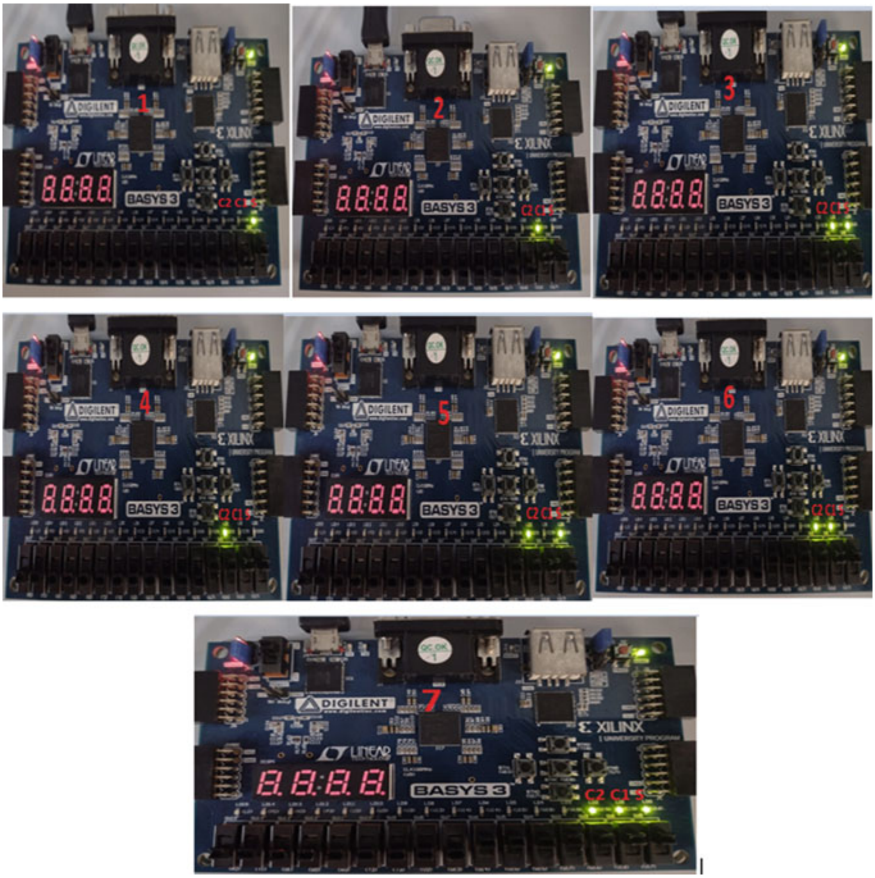


Fig. 7 7:3 proposed architecture in Basys 3 Artix 7 FPGA board

Table 1 Comparison of counter circuits

Circuit design	No. of XOR and mux gates in the critical path	Latency (ns)	Power (W)
CMOS full adder counter [10]	6 XOR	2.435	2.12
Parallel counter [11]	4 XOR	2.315	2.21
Mux based counter [12]	1 XOR + 3 MUX	2.215	2.32
Symmetrical bit stack counter [10]	1 MUX	1.859	2.41
Proposed architecture (Basys 3 Artix 7 FPGA board utilized)	0 XOR or MUX	1.796	1.978

5 Conclusion

In this paper, a novel design of 7:3 binary stacking counter using asymmetrical stack combination of 5-bit and 2-bit stack circuit is proposed. Area, power, and latency of the proposed architecture are analyzed, and the functionality of the proposed architecture is validated using Basys 3 Artix 7 FPGA board. The performance of the proposed architecture in terms of power and latency is compared with the conventional 7:3 counter circuits. The power consumption of this proposed architecture is lower than any other conventional circuits. The CMOS full adder-based counter has 6 XOR gates in the critical path. Parallel counter has 4 XOR gates. Mux based counter has 1 XOR gates and 3 multiplexers. And the symmetrical bit stack counter has 1 multiplexer in the critical path. While the proposed architecture doesn't have any XOR or MUX in the critical path. So, this architecture outperforms conventional counter circuits. This proposed binary stacking counter can be extended to higher order counter as a future work, which finds application in big data analysis and deep learning, etc.

References

1. H.H. Shravani, J.P. Anita, Structured DFT based analysis of standard benchmark circuits. Lecture Notes Electr. Eng. **569**, 705–715 (2020)
2. N. Mohan., J.P. Anita, Compact test and diagnosis pattern generation for multiple fault pairs in single run. J. Eng. Sci. Technol. **15**(6), 3820–3835 (2020)
3. P. Gupta, D. Samnani, A. Gupta, A. Asati, Design and ASIC implementation of column compression wallace/dadda multiplier in sub-threshold regime, in *International Conference on Computing for Sustainable Global Development* (2015)
4. C.S. Wallace, A suggestion for a fast multiplier. IEEE Trans. Electron. Comput. **EC-13**(1), 14–17, 1964
5. L. Dadda, Some schemes for parallel multipliers. Alta Freq. **34**, 349–356 (1965)
6. Z. Wang, G.A. Jullien, W.C. Miller, A new design technique for column compression multipliers. IEEE Trans. Comput. **44**(8), 962–970 (1995)
7. W.J. Stenzel, W.J. Kubitz, G.H. Garcia, A compact high speed multiplication scheme. IEEE Trans. Comput. **C-26**, 948–957 (1977)

8. P.K. Somayajulu, Ramesh, S.R, Area and power efficient 64-bit booth multiplier, in *International Conference on Advanced Computing and Communication Systems* (2020), pp. 721–724
9. D. Lachireddy, S.R. Ramesh, Power and delay efficient ALU using vedic multiplier. Lecture Notes Electr. Eng. **672**, 703–711 (2020)
10. C. Fritz, A.T. Fam, Fast binary counters based on symmetric stacking. IEEE Trans. Very Large Scale Integ. (VLSI) Syst. **25**(10) (2017)
11. B. Sajeeva Rao, A. Ramkumar, FPGA implementation of algorithmic counter based wallace tree. Int. J. Electr. Commun. Technol. **8**(3) (2017)
12. G. Divya, S.M.E. Vengatesh Kumar, Mux based novel counter design for low power application. SSRG Int. J. Electr. Electron. Eng. (2018)

Feature-Level Fusion of Multimodal Biometric for Individual Identification by Training a Deep Neural Network

B. Nithya and P. Sripriya

Abstract Digital era needs obligatory requirement of multimodal biometric to spot a person to access a specific environment. To provide such a high secure platform, the proposed system used fingerprint and face modalities to check an individual's identity. The objective of this research work is to give complete recognition accuracy without performing any pre-processing on the acquired images of face and fingerprint. To obtain this, the features are extracted using histogram of oriented gradients (HoG) and Speeded up Robust Features (SURF) algorithms. These features are fused to provide as input to train a deep network. The images are acquired from public databases, AT & T are used for face images, and FVC 2004 is used for fingerprint images. The pre-defined convolutional neural network (CNN) models, AlexNet, GoogLeNet, VGG16 and ResNet50 are also tested with the acquired images. But the proposed system well-behaved and has given highest recognition accuracy than other CNN methods.

Keywords Fingerprint · Face · Multimodal · Classification · Deep network · Feature fusion · Speeded up robust features · Histogram of oriented gradients · Sequence inputs

1 Introduction

Biometric is a unique inherent physical or behavioral characteristics. To find individuality of a person, these peculiar traits are used; this is known as biometric recognition system. If only one modal is used for identifying an individual is known as uni-modal biometric authentication structure. If more than one modality is used for recognizing a person is called multimodal authentication system. To improve the security, the later one can be applied for knowing an individual and the multiple biometric sources can provide more uniqueness to a person than uni-modal. But multimodal biometric is

B. Nithya (✉)
Department of Computer Science, VISTAS, Chennai, Tamil Nadu, India

P. Sripriya
Department of Computer Application, VISTAS, Chennai, Tamil Nadu, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_12

also have some disadvantages such as extra hardware is needed to acquire additional biometric modal, extra cost and storage resources. Intra-class variability is another challenge when biometric image acquisition. Because, at the time of acquisition, user's biometric trait is stored in the database will be different at the time of verification. Even though this identification system has these limitations, this is a most promising identification system to get more accuracy (e.g., in national border identifying a soldier is a most important one) mentioned in the book by Jain et al. [1]. Multi-biometric system is divided into multi-modal, multi-sample, multi-sensor, multiple algorithms and multiple instances. The multiple sensors, multi-algorithm, multi-sample and multiple instances techniques are use one modality. But multimodal uses two or more biometric traits, and this has to be fused. The fusion can be done in different ways, feature-level fusion, rank level, decision and score level fusion. Sensor and feature fusion can be done before matching process. But score level, rank and decision level fusion are done after matching process. Feature fusion is features of the input images are extracted and concatenated and then matched. Sensor level fusion is fusing images acquired from multiple-sensors. Score level fusion is combining the matching scores of input modalities. Rank fusion is matching ranks of processed inputs with the database, where highest rank is the best match that will be fused. The proposed system uses hybrid multi-biometric technique because it combines multimodal biometric and multiple samples of modalities.

There are five reasons to use multimodal, they are to obtain accuracy, increased identification, security enhancement and vulnerability and user acceptance. When compared with the feature-level fusion, the score-level fusion is applied most due to its easy achievable. This is said by Yadav et al. [2]. But feature-level fusion has some challenges in fusing stage because file formats will differ when acquiring different modalities in different sensors and also the dimensionality mismatch will occur after the features are acquired this is tried by Mane et al. [3]. Almabdy et al. [4] proved that instead of using conventional matching techniques for biometric recognition, biometric patterns are given to deep network for training and classification to increase the overall performance. The recent research works on multimodal biometric recognition are increased with deep neural network (DNN) training. Deep neural network has three types of layers, input, output and hidden layers. This layered architecture is used to train the features extracted from biometric traits. This extracts the features of input images automatically and learns the features of images, and it gives the classified result. The features can also be extracted externally, and the feature sets can be given for training. This method of learning the features from feature sets is an improved version of traditional features matching algorithms. For training, a deep network with images and its feature classification, pre-trained convolutional neural network (CNN) methods are available like Alexnet, ResNet50, VGG16, etc. But instead of using pre-trained CNN for training the fused biometric, the proposed one uses a novel deep neural network classification strategy with sequence input learning method. The biometric recognition accuracy of proposed model is compared with the pre-trained models' identification accuracy. The contribution of this research work is:

(i) Image acquisition from public databases
(ii) Different feature extraction methods for face and fingerprint without any pre-processing techniques
(iii) Normalizing feature sets
(iv) Concatenation (feature-fusion)
(v) Dimensionality Reduction
(vi) Training a deep network with the extracted features with reduced dimension
(vii) Finally, comparing the classification accuracy results with various pre-trained CNN (convolutional neural network) models and with the proposed one.

The proposed model handled the challenge of intra-user variability by acquiring multiple samples from same biometric trait and given these samples for the deep neural network training.

2 Related

Various research works are carried out on multimodal biometrics with deep learning techniques. This section describes the recent studies. Alay et al. [5] have proposed a multimodal recognition with three modalities of finger vein, iris and face. The pre-trained CNN method VGG 16 was applied to perform training the inputs. Hybrid fusion was tried and got 100% accuracy at score-level fusion and 99.39% accuracy at feature-level fusion. The authors Arora et al. [6] have described various combinations of CNN techniques with two modalities (face, left and right iris). Here, they combined the features and classification is accomplished with the softmax classifier. Before acquiring the features using CNN, pre-processing was carried out to enhance input images. VGG gives face features and using convnet iris features were extracted. The images were taken from the public databases such as CASIA-face V5, IITD iris and got 99.80% accuracy. Al-Waisy et al. [7] have presented a novel on real-time and effective hybrid fusion of face, left and right iris. Here, also some pre-processing steps were followed to get enhanced images. For extracting features, deep belief network was applied with 3 layers with softmax classifier. Score-level and rank-level fusion methods gave 100% accuracy at recognition stage. Sengar et al. [8] have proposed a deep learning based multimodal authentication using fingerprint and palmprint. Histogram equalization and segmentation are carried out before the images were given for deep network training. The researchers Infantraj et al. [9] show training results before feature extraction and after doing pre-process and feature extraction. Iris and face were used for this research work and concluded that pre-process and feature acquirement gives better results than without doing pre-process. Silva et al. [10] have suggested modified VGG for feature extraction from two modalities of iris and periocular. Those are taken from the UBIRIS-V2 dataset and NICE dataset. Feature-level fusion was tried by selecting features by particle swarm optimization (PSO) technique. PSO searches all the possible combinations and gives satisfied result. The writers Soleymani et al. [11] pre-processed the face, fingerprint and iris images and fused different algorithms and various fusion settings. And various database images were taken for comparing the proposed method's performance. Results show that multi-abstract fusion stretch 99.91% accuracy at BIOMDATA multimodal database and 99.34% at BioCop dataset.

The both uni-modal and multimodal techniques were implemented to find which one gives better results Gowda [12]. When resizing inputs and given these to novel neural network with two layers, 100% accuracy was obtained in recognition at multimodal. The fusion can be done not only with the modalities, but also layers can be fused if deep networks were applied by Xu [13]. The face, iris and palmprints are given to Alexnet network, and the fusion is carried out at first convolution then features are fused at next layer. This innovative method for multimodal biometric provides better results. A feature—level fusion is proposed by Tiong et al. [14] but pre-processing was done to eliminate illumination changes. Deep network was trained with the fused features acquired from handcrafted texture descriptors of face and periocular traits. Finally, scores are also fused. Different score fusion methods have proposed to fuse scores such as, maximum, minimum and concatenation of scores. Prabu et al. [15] suggested a hybrid adaptive fusion of two methods (Effective Linear Binary Patterns—ELBP and SIFT). Iris and hand geometry images are taken from public datasets and pre-processed using median filter and image enhancement techniques. They proposed bi-modal fusion methods, the method one extracted features using SIFT from iris and ELPB from hand geometry. The method two extracted features using ELPB from hand geometry and SIFT from iris traits. Then, the response time and accuracy were calculated. With 700 test images, 99.42% accuracy was attained as highest recognition rate. The authors Preetha et al. [16] proposed a decision fusion with KNN classifier and compared the results with some machine learning techniques. Features are extracted with the help of hybrid wavelet with pre-processed images of face and iris. When observing this survey, we come to know that the preference is given to score-level fusion than the other three methods of fusions. And every research work in this survey had pre-process step for input enhancement and one of the researchers proved that after doing the pre-process the accuracy had been improved. The following sections describe the multimodal recognition with the use of face and fingerprint by training a deep network.

3 Feature Abstractions

The proposed method uses two modalities, face and fingerprint biometric traits. The pre-trained CNN models extract features internally from the input images and trains the corresponding network. But the proposed work uses two different algorithms, the histogram of oriented gradients (HoG) is applied on face images to mine discriminant features. And SURF (Speeded up Robust Features) algorithm is applied on fingerprint for gaining unique features.

3.1 HoG Descriptor for Face

The histogram of oriented gradients (HoG) is one of the most used feature descriptor techniques. It mainly concentrates on the nature of an object. This algorithm slides a window of block which is nothing but a pixel grid. The HoG algorithm finds a pixel inside the grid whether it is an edge or not. The grid size can be any number in the form of n × n. The different grid size is calculated on the face image which is depicted in the figure (Fig. 1). Here, three different sizes are used, 2 × 2, 4 × 4 and 8 × 8 and the length has also been showed. The length is higher in number if the cell size is lower, and the length is lower in number if the cell size is higher. So the number of features is reduced if the cell size is increased.

In figure (Fig. 2), the 64 bit of block is looked closure, and the grid values of the pixel is showed. If the pixel value is known then the gradient can be calculated with the help of following formula. For every pixel, the horizontal and vertical gradient are calculated. The magnitude is calculated as:

$$\sqrt{G_x^2} + G_y^2 \tag{1}$$

Here, G_x is base magnitude, and G_y is perpendicular magnitude. Then, the direction of the pixel is calculated by:

$$\tan(\emptyset) = G_x/G_y \tag{2}$$

The tan gives the direction of the pixel. Using the values received from the previous two equations, the HoG feature descriptor finds the discriminant features.

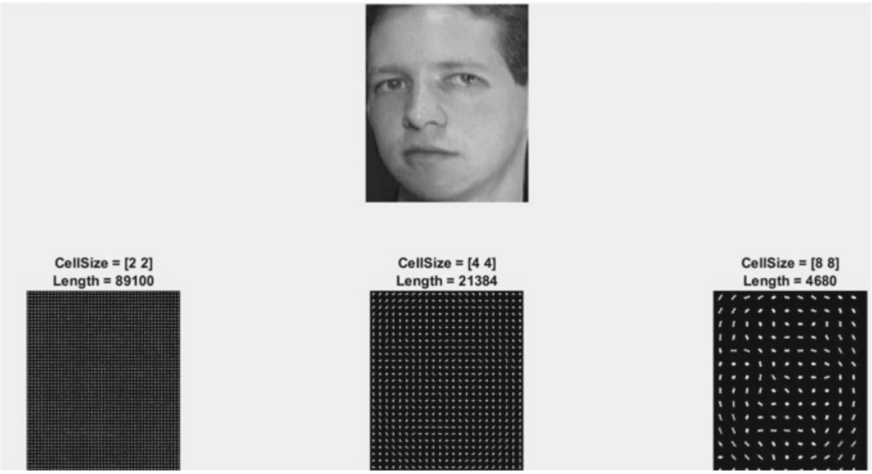


Fig. 1 Different grid size of HoG calculation on a face image

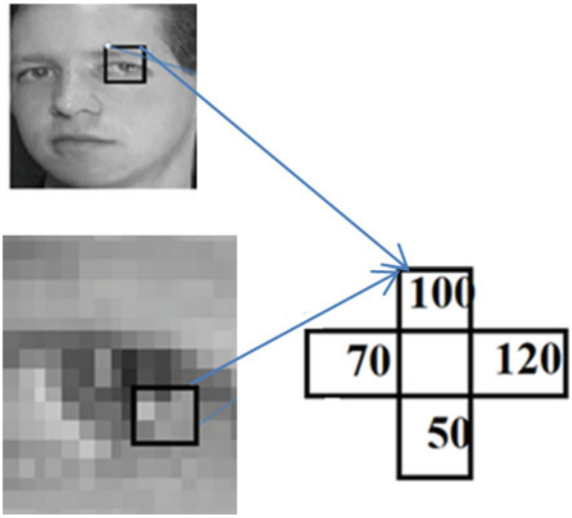


Fig. 2 Closure look on grid of block

3.2 SURF Features

The SURF (Speeded up Robust Features) is a fast and forceful algorithm in detecting features from an image. Here for fingerprint, biometric trait features have been extracted with the aid of SURF algorithm. It uses $n \times n$ matrix box as a filter to get local interest points, and these are discovered by Hassian Matrix Approximation. It has the advantage of lower computation time and highest accuracy. In a given fingerprint image, it does convolution operation to find average intensity. To calculate this:

$$I_{\sum}(x) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i, j) \tag{3}$$

The $I_{\sum}(x)$ represents the input image as integral picture at the location x which can be calculated as $x = (x, y)T$, this denotes the summation of all pixels in I. And the hessian pixel calculation is done by:

$$H(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial_x \partial_y} \\ \frac{\partial^2 f}{\partial_x \partial_y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \tag{4}$$

(x, y) is every pixel in the given image and the $H(f(x, y))$ is hessian pixel calculation. The given is then filtered using Gaussian kernel at a pixel point $X = (x, y)$, then the scale σ is computed in the Hessian matrix $H(x, \sigma)$ which is defined as:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \tag{5}$$

The $L_{xx}(x, \sigma)$, $L_{yy}(x, \sigma)$ and $L_{yy}(x, \sigma)$ describe the box filters of the Gaussian second order derivative of input image I at the point x. The Gaussian derivatives used due to its computing cost, and SURF algorithm is considered to be a fast one.

4 Processing Sequence and Feature-Level Fusion

4.1 Acquisition and Processing Sequence

Acquisition Sequence

Serial, parallel and hierarchical are three types of biometric acquisition methods. The serial method is not needed any special caring on sensors organization. Each source of biometrics is gained sequentially that is fingerprint is gained first, and the face image of the person is gained next. Parallel method of acquisition needs special kind of sensor arrangements to acquire biometric traits. Multiple sensors are required simultaneously to acquire biometric traits in parallel way. That is, both modalities are acquired simultaneously and processed at the same time. In hierarchical mode of acquisition, both serial and parallel modes are combined.

Processing Sequence

The acquired biometrics can be processed sequentially or in parallel fashion. The received modalities are processed one by one and fusing the processed images is known as sequential processing. If the modalities processed simultaneously and fused, then this is known as parallel processing. Generally, multimodal architecture uses the parallel way of processing to reduce the error rate. The proposed system sequentially acquired the samples of a person and fused the fingerprint and face features and then the processing were carried out.

4.2 Feature-Level Fusion

This is a challenging module of choosing the type of fusion method when multiple modalities are used for identification. Feature-level fusion is fusing the discriminants features attained from different biometrics. This process is done before matching and integrating the data at primary stage are considered to be a more operative then the combining at a later stage. This scheme is divided into two categories, namely, heterogeneous feature fusion and homogeneous feature fusion. Many samples of same biometric modal's features have been extracted using same algorithm is known as homogenous. If different algorithm or different samples of various biometric modalities is used for extracting discriminant features, is known as heterogeneous feature fusion. In proposed system, the later technique is applied that different samples and different modalities are used such as fingerprint and face. The extracted features are normalized and then concatenated the features. If diverse modalities are used, the big challenges on concatenation of two feature sets since the modalities' feature sets size are different. So before concatenating, converting the two feature sets into the same size is the important process. After concatenation, the size of the feature sets has been increased. To handle this problem, the principal component analysis algorithm is applied for reducing the concatenated feature sets' dimensionality.

5 Proposed System

The proposed method is having three stages, normalizing feature sets, concatenation and feature selection. This is done after extracting the features. The third stage eliminates the dimensionality problem by applying principal component analysis algorithm.

5.1 *Normalizing Feature Sets*

The objective of normalizing feature sets is to change the location (mean) and variance (scale) of features values. This process helps to identify outliers in the values and this may not be necessary if the values are already compatible. Here, in proposed system, the normalization is accomplished with min–max technique and it calculates:

$$\hat{x} = \frac{x - \min(h_x)}{\max(h_x) - \min(h_x)'} \tag{6}$$

where x is generated by the method h_x and all the probable minimum and maximum x values are perceived by $\min(h_x)$ and $\max(h_x)$. After the normalization, the new and normalized features are $\widehat{x_1} = \widehat{x_1^1}, \widehat{x_2^2}, \ldots, \widehat{x_1^{d1}}$ and $\widehat{x_2} = \widehat{x_2^1}, \widehat{x_2^2}, \ldots, \widehat{x_2^{d2}}$.

5.2 *Concatenation*

After the successful normalization, the new and modified feature sets $\widehat{x_1}$ and $\widehat{x_2}$ are attained. The proposed system of multimodal biometrics system, where the objective is to fuse diverse-length feature sets $\widehat{x_1}$ and $\widehat{x_2}$ to obtain fresh feature set $\widehat{x_f} = \widehat{x_1^1}, \widehat{x_2^2}, \ldots, \widehat{x_1^{d1}}, \widehat{x_2^1}, \widehat{x_2^2}, \ldots, \widehat{x_2^{d2}}$. The resultant fused vector $\widehat{x_f}$ should have the dimensionality d, $d < (d_1 + d_2)$. D_1 and d_2 are the dimensionalities of $\widehat{x_1}$ and $\widehat{x_2}$.

5.3 Feature Selection

There are many methods are in use to get selected feature after the concatenation or fusion of two or more feature sets. PCA (principle component analysis) is one of the popular methods of feature section or transformation. The reason for doing this step is, when compared to the larger feature sets, the smaller sets are easy to handle and training these with deep network will be faster. This method has mainly three steps, standardization, computation of covariance matrix and the computation of Eigen values and Eigen vectors which identifies the principal components. Standardization is to reduce the domination over small range variable by larger range variables in feature sets. This can be done by subtracting the mean of features from the values and divided by the standard deviation of every value in the each variable.

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \tag{7}$$

The importance of computing covariance matrix $p \times p$ is to make understand about the values of input variables' variation against the mean respect to each other. Then, the Eigen values and vectors can be extracted from the covariance matrix which gives the reduced vectors of the input feature sets. After the successful completion of reducing dimensionality using PCA algorithm, the newest and highly important feature set $x_f = [x_1, x_2, \ldots, x_d]$ can represent as a biometric template, and this could be given for the further process in the identification system.

5.4 Proposed Model

This research work has three steps, training stage, validating stage and testing. The training step (Fig. 3) first acquires images from virtual dataset, and it doesn't do any pre-process as the authors did in review of literature.

Preliminary Steps of Proposed

The raw images of fingerprint and face are given to feature extraction stage; here HoG method gets unique features from face images. SURF method gets distinct features from fingerprint images. For each person, four samples have selected for each modality. After extracting features from the fingerprint and face modalities, the feature sets are given for the normalization process. Where the outliers are removed and then given for the fusion stage. Here, the features are concatenated like the first sample of face's feature is fused with first sample of fingerprint; the second sample is fused with second sample of another trait, and so on. So after the fusion, again the four fused vectors are gained. Then, the dimensionally is reduced by principle component analysis method. These vectors are given for the training process to classify the inputs.

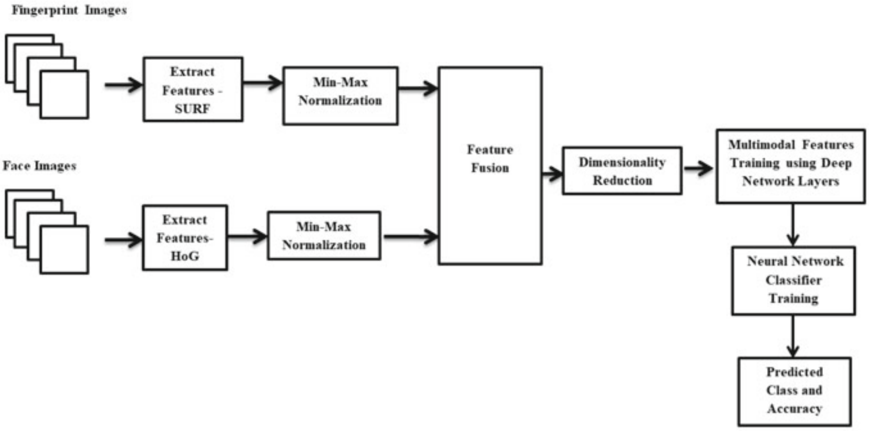


Fig. 3 The proposed multimodal biometric system

Training and Testing

In this stage, the deep neural network is designed with five layers. Sequence input layer, LSTM (Long Short-term memory) layer, fully connected layer, softmax and finally classification layer. In this research work, the inputs are given as sequence of 2-D vectors with the size of 64 dimensions in the first layer. The inputs given in this layer is the concatenated and dimension reduced feature vectors as sequence. And the second layer is LSTM; the long short-term memory layer has hidden state which has 100 units used to back up the previous stages of outputs. Then, the fully connected layer has final discriminant features, followed by softmax and classification layers (Table 1).

At the validation stage, the classifier is trained with already trained network of five layers, and this gives the accuracy results of given images. Finally the testing stage, it checks with one pair (fingerprint and face) of images and does the same process

Table 1 Layers of the proposed deep training network

Name of the layer	Type	Activations	Learning factors
Sequence	Sequence input with 64 dimensions	64	Sequence of features as vectors
LSTM (long short—term memory)	LSTM with 100 hidden units	100	Input weights 400 × 64 Recurrent weights 400 × 100
Fully connected	20 fc layers	20	Weights 20 × 100
Softmax	softmax	20	Feature Vectors
Class output	Classification-Crossentropyex with ‘1’ of 19 other classes		Prediction of the subject

to identify the performance of the trained system, whether it correctly classifies the given input or not. The following sections describe about the results attained by proposed system.

6 Experiment Results and Discussions

6.1 Testing Setup and Dataset

The experiment is carried out on the framework MATLAB version 2020b with the system configuration of Intel Core i5 with the RAM speed of 8 GB and 2 GB capacity of NVIDIA graphics card. For fingerprint images, FVC2004 dataset is selected and for the face images AT and T dataset are chosen. These are uni-modal databases but our proposed system is multimodal. So virtual database is made and believed the biometric traits are from same person. From each person in the face and fingerprint dataset, four samples have taken. The first sample of face is fused with the first sample of face and so on. Like this, 20 subjects, each one has four samples totally 80 vectors received after the concatenation. From that the training vectors are 60 and for testing 20 were given.

6.2 Results and Discussions

The sample images from virtual dataset for proposed multimodal biometric are portrayed in figure (Fig. 4). From each person, four face images from AT and *T* database and four fingerprint images from FVC 2004 database have selected to extract features from each image. The reason for selecting multiple samples from each modality is to reduce the error when it faces intra-user variability problem. One face may pose in different way and in different scenario, so when training the deep network with multiple samples with various scenarios has given better improvement

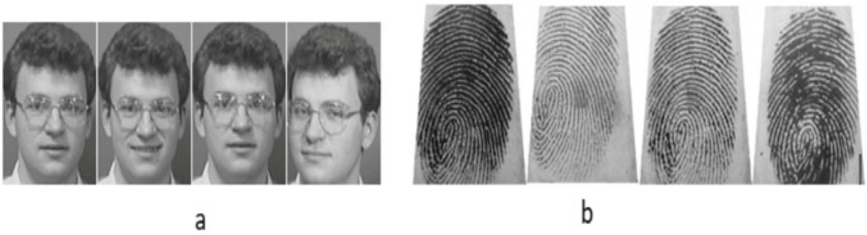


Fig. 4 Virtual database for face and fingerprint (sample images) **a** four samples from AT and *T* dataset, **b** four samples from FVC2004 dataset

in the proposed method. The extracted feature points have been marked in the images by applying HoG algorithm on face and SURF algorithm on fingerprint are depicted in figure (Fig. 5).

Here we have not pre-processed, just the raw images have taken and pointed the features. The virtual database images have given to pre-trained CNN models, Alexnet, Resnet50, VGG16 and GoogLeNet. Here, the models extract features and concatenate internally to train these models, and the accuracies are calculated.

The total number of layers of GoogLeNet is 144, Vgg16 is 41, Resnet50 is 177 and Alexnet is 25 but the proposed model has 5 layers of network (Table 2). This accurateness is compared with other pre-trained CNN models and with the proposed model's accuracy. The projected model gives 100% of accuracy in multimodal recognition on proposed virtual database. The resnet50 gives 95%, and this is the highest among other pre-trained CNN classification models. The reason for getting 100%

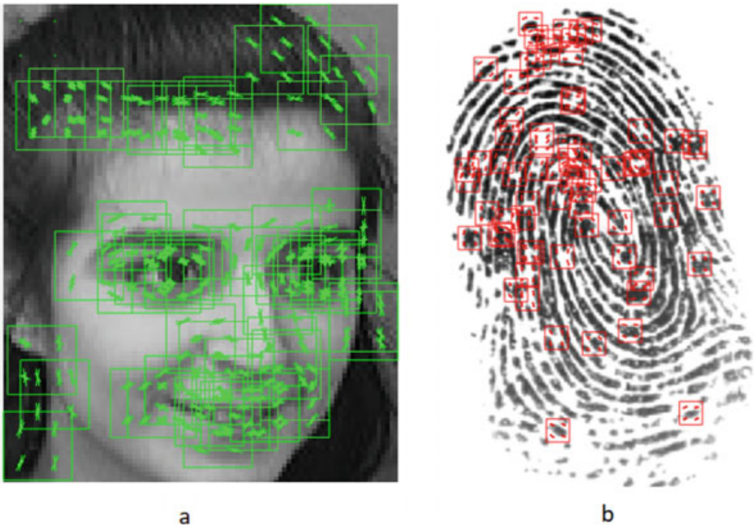


Fig. 5 Detected features **a** HOG features of face **b** SURF features of fingerprint

Table 2 Accuracy results of proposed and pre-trained CNN models

Model	Total No. of layers	Image input	Fully connected layer (weights)	Accuracy (%)
GoogLeNet	144	224 × 224 × 3	1000 × 1024	45
Vgg16	41	224 × 224 × 3	1000 × 4096	60
Resnet50	177	224 × 224 × 3	1000 × 2048	95
Alexnet	25	227 × 227 × 3	1000 × 4096	60
Proposed	5	64 sequence dimensions	20 × 100	100

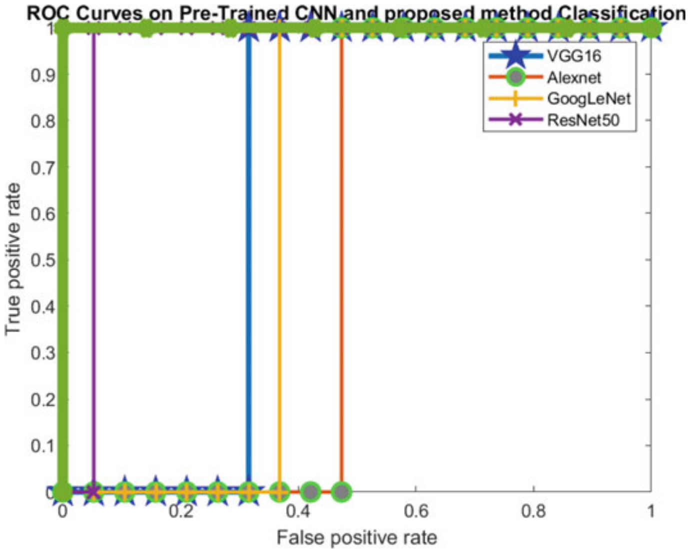


Fig. 6 ROC for the proposed multimodal biometric system

accuracy in the proposed method is the pre-trained CNN methods are taking inputs as images in the size of 227 × 227 × 3. But in the proposed system takes the inputs as sequence of vectors with reduced dimensions.

The receiver operating characteristics (ROC) curve is an effective and a graphical tool to show a method's accuracy. In figure (Fig. 6), ROC curve shows the proposed method and other pre-trained CNN models' accuracy as a graphical representation. This ROC diagram showed the proposed method is green in color, which lies only on the true positive rate. This says that all the given images are correctly or truly identified as positive. The ResNet50 model is marked in violet color with cross mark, VGG16 is in blue color with star mark. The AlexNet is in red color with circle filled and the GoogleNet is showed in orange color. The different colors used here is to make a difference on the CNN methods and the proposed method. We can understand through the graph that the proposed method outperforms than others because it gives 100% accuracy in recognition. Resnet50 also shows good performance and the least performance was by Alexnet model. The novelty of the proposed work is bringing the raw images as feature vectors by doing 3 step processes (normalization, concatenation and dimensionality reduction) and training the feature vectors as sequence inputs. Even without enhancing, the raw images, the highest accuracy has been acquired in this proposed model. Also the intra-user variability problem is handled by giving multiple samples of the biometric traits.

7 Conclusions

Multimodal biometric is an important and inevitable one to escape from known threats. But the challenge here is to provide 100% accuracy in recognition and raw data handling. To face this difficulty, in literature review, many of the researchers had experimented pre-processing techniques to acquire highest accuracy. But the proposed system has a sequence of steps like biometric images acquisition, different feature extraction methods (HoG feature extraction for face images, SURF extraction method for fingerprint images), and these features are normalized to eliminate the outliers in the feature sets. After the normalization, the features are concatenated and dimensionality is reduced. Finally, the feature sets have been given as sequence inputs to the training process. Here, the highest accuracy of 100% is achieved even without doing pre-processing. The same acquired biometric images are fused and are given to the pre-trained CNN models for classification purpose. The pre-trained models are designed to make computation with only on images. But the feature fusion gives numerical feature sets and when it is doing the classification on numerical feature sets the pre-trained models are not giving much satisfied recognition result. But the proposed system's identification accuracy is higher than other pre-trained CNN techniques because it handles the feature sets in well-organized manner.

In this work, the accuracy is the only factor in the comparison stage with the other CNN techniques. But the future work will include other performance factors like time taken, throughput and speed of the proposed work. Also the upcoming work will add extra biometric traits with increased samples and will increase number of subjects. The fusion was done in the way of sequence, but the future work may fuse all the fingerprint images with all the four samples of face images.

Acknowledgements Dedicated to Professor Dr. P. Sripriya, Department of Computer Applications, Vels institute of Science, technology and advanced studies, Chennai, Tamil Nadu, India.

Conflict of Interest The authors confirm that there is no conflict of interest to declare for this publication.

References

1. A.K. Jain, A.A. Ross, K. Nandakumar, *Introduction to Biometrics* (Springer, New York Dordrecht Heidelberg London, 2011)
2. N. Yadav, J.K. Gothwal, R.S. Shyam, Multimodal biometric authentication system: challenges and solutions. Glob. J. Comput. Sci. Technol. **11**(16), 57–60 (2011). ISSN 0975-4172
3. V. Mane, J. Dattatray, Review of multimodal biometrics: applications, challenges and research areas. Int. J. Biom. Bioinforma. **3**(5), 90–95 (2009)
4. S. Almabdy, L. Elrefaei, An overview of deep learning techniques for biometric systems, in *Artificial Intelligence for Sustainable Development: Theory, Practice and Future Applications* (2020), pp. 127–170

5. N. Alay, H.H. Al-Baity, Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. Sensors **20**, 5523 (2020). https://doi.org/10.3390/s20195523(2020)
6. S. Arora, M.P.S. Bhatia, H. Kukreja, A multimodal biometric system for secure user identification based on deep learning, in *Proceedings of 5th International Congress on Information and Communication Technology*. ICICT 2020. Advances in Intelligent Systems and Computing (2021), 1183p
7. A.S. Al-Waisy, R. Qahwaji, S. Ipson, S. Al-Fahdawi, A multimodal biometrie system for personal identification based on deep learning approaches, in *2017 7th International Conference on Emerging Security Technologies (EST)* (2017), pp. 163–168. https://doi.org/10.1109/EST.2017.8090417 (2017)
8. S.S. Sengar, U. Hariharan, K. Rajkumar, Multimodal biometric authentication system using deep learning method, in *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)* (2020), pp. 309–312. https://doi.org/10.1109/ESCI48226.2020.9167512
9. A. Infantraj, M. Augustine, A multimodal biometric recognition system using principal component analysis and feed forward neural network for mobile applications. Int. J. Adv. Res. Edu. Technol, 17–21 (2017)
10. P.H. Silva, E. Luz, L.A. Zanlorensi, D. Menotti, G. Moreira, Multimodal feature level fusion based on particle swarm optimization with deep transfer learning, in *2018 IEEE Congress on Evolutionary Computation (CEC)* (2018), pp. 1–8. https://doi.org/10.1109/CEC.2018.8477817
11. S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, N.M. Nasrabadi, Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification, in *2018 24th International Conference on Pattern Recognition (ICPR)* (Beijing, China, 2018), pp. 3469–3476. https://doi.org/10.1109/ICPR.2018.8545061
12. S.H.D. Gowda, M. Imran, G.H. Kumar, Feature level fusion of face and iris using deep features based on convolutional neural networks, in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (Bangalore, 2018), pp. 116–119. https://doi.org/10.1109/ICACCI.2018.8554683
13. H. Xu, M. Qi, Y. Lu, Multimodal biometrics based on convolutional neural network by two-layer fusion, in *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (2019), pp. 1–6. https://doi.org/10.1109/CISP-BMEI48845.2019.8966036
14. L.C.O. Tiong, S.T. Kim, Y.M. Ro, Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion. Multimed Tools Appl. **78**, 22743–22772. https://doi.org/10.1007/s11042-019-7618-0
15. S. Prabu, M. Lakshmanan, V.N. Mohammed, A multimodal authentication for biometric recognition system using intelligent hybrid fusion techniques. J. Med. Syst. **43**, 249 (2019). https://doi.org/10.1007/s10916-019-1391-5(2019)
16. S. Preetha, S.V. Sheela, New approach for multimodal biometric recognition, in *Machine Learning for Predictive Analysis. Lecture Notes in Networks and Systems*, vol. 141, ed. by A. Joshi, M. Khosravy, N. Gupta (Springer, Singapore, 2021). https://doi.org/10.1007/978-981-15-7106-0_45

Design of Multistage Counters Using Linear Feedback Shift Register

Neethu B. Nair and J. P. Anita

Abstract Applications such as single-photon detection require the use of large array of counters within a small area. Linear feedback shift registers (LFSR) can be considered as the best option for such applications, where the area can be considerably reduced. Compared to a conventional binary counter, these counters enhance area and performance. In existing literature, only many-to-one LFSR structure was used. However, if LFSR counter is used for first stage and binary counters for subsequent stages could increase performance of the counter designs. In this paper, the LFSR counters are combined with binary counters and the performance in terms of area, speed and power are compared with existing multistage counters. Also instead of using a decoding logic for each stage as in existing literature, a single decoder with a multiplexer is used. This gives a reduced area. A two stage LFSR counter is implemented in Xilinx Vivado 2017.2, and the results are validated. Due to its efficacy, they can be used in deep learning and big data analysis.

Keywords Linear feedback shift register (LFSR) · Decoding logic · Sequence extension logic · Ripple carry logic · Binary counters · Single-photon detection

1 Introduction

With the new developments in applications like single-photon detection, there arises a need for implementation of huge number of arrayed counters in small areas. Also for the applications involving photon-counting cameras the main aim is to bring down the area utilized by counters as each camera pixel consists of separate counter. So with increase in numbers of camera pixels, area used by counters will also increase. Linear feedback shift registers are generally used for generating pseudorandom numbers but it can also be used as synchronous counters [1]. LFSR counters are used for applications like CMOS pixel and for detection of single-photon arrays [2]. The speed of clock in LFSR is not dependent on the number of bits in the counter and

N. B. Nair · J. P. Anita (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: jp_anita@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_13

goes through each and every state besides the zero state. Although, LFSR count order is pseudorandom but additional logic is needed to convert the LFSR states into binary order.

The comparison of methods used for decoding LFSR states into binary is discussed in [1]—the iteration method, direct lookup table (LUT) method and time-memory trade-off algorithm. The iteration technique does the iteration of whole LFSR count sequence, and all values are compared to the values of counter. An n-bit LFSR on an average needs 2^{n-1} comparisons. For the direct LUT technique, n x n LUT is used which converts the LFSR states directly. The time-memory trade-off algorithm merges both the techniques by keeping $2^{(N/2)}$ LFSR values in a table and iterates the LFSR values till the matching count sequence is obtained from the table. To get the decoded value, the number of iteration should be deducted from the value stored in table. For applications that involves enormous arrays need all the cells to be converted to binary order for additional processing. Therefore, decoding logic should be fast and prone to error as lot of conversions are taking place.

A new counter design is proposed in this paper in which multiple LFSR stages are used where additional logic is used for decoding LFSR states into binary values. Also here ripple signal is used instead of direct concatenation of LFSR counters so as to avoid any reduction in performance.

2 Related Work

LFSR used for generation of pseudorandom numbers is a shift register where the input bit is a linear function of its previous stage. The sequence produced by the register is dependent on the current state of the register. With correctly chosen, feedback function LFSR produces random sequences of maximum length. LFSR can also be used as an event counter by shifting the values in the register whenever an event occurs. All the achievable non-zero values are generated using XOR along with the shift register. A few sizes of shift register need more than one XOR gate for implementing LFSR of maximum length. Clark and Weng [3] propose the utilization of single-gate LFSRs which generate sequence of maximum length.

Morrison et al. [4] propose multistage LFSR counter design for applications which needs large array of counters. This design also has better improvement in area and performance than the binary counters. Although some additional logic is used for decoding the count order into binary values. The LFSR used in this method has sequence extension logic in addition to the feedback logic. As the basic LFSR covers only maximum of 2^{n-1} states for n bits. To cover all the states, sequence extension logic is used which uses NOR and XOR gate to include the missing states. Ajane et al. [1] give comparison of LFSR counters with binary counter. The comparison is made between both the methods on the basis of speed, area and power consumption. Here, an efficient algorithm is used for decoding the pseudorandom sequence generated by LFSR counter to binary value. Moreover, different sizes of LFSR counters and binary counters are implemented, and performance of both the counter techniques is compared. On comparing, it is observed that LFSR counter is better in terms of area and speed than the binary counters. LFSR based counter has better performance compared to binary counters. However, additional logic is required to convert the count order to binary. Sushmita et al. [5] provide four-stage multistage LFSR counter design that retains the advantages of single-LFSR stage which is fabricated in 90 nm CMOS technology.

Low-power PRBS generator is designed in [6] which is used to test multilane multi-Gb/s transmitter to avoid crosstalk. This generator produces three uncorrelated PRBS sequences. For reducing the power, quarter-clock rate power efficient LFSR is used. Ring generator architecture can also be used as LFSR feedback. Ring generator is constructed using a ring of memory elements. Ring generator also produces maximum length sequence. Ring generator lowers the use of XOR logic by using only single two input XOR gate between the flip-flops when feedback logic is there [7]. The LFSR uses the tap configuration described in [8].

On the basis of different applications, Fibonacci and Galois LFSRs can be used. The requirements of applications can be attained using random length and tag configuration. As the LFSR structure in most of the design are fixed and can't change the type of feedback used. However, the method in [9] finds a solution to this problem by using reconfigurable LFSR architecture where Fibonacci and Galois LFSR can be selected according to various applications. This design prevents attacks due to complexity of design. [10] uses LFSR for generating test vectors as well as for address generation for memory testing in BIST architectures. With the help of this modified LFSR design, single circuit can be used instead of two circuits thus reducing the overall circuit complexity.

For applications like cryptography, PRNGs are required. So [11] implemented multibit LFSR which generates pseudorandom sequences. Compared to traditional LFSR where only one bit is shifted in each clock cyle multibit LFSR shifts multiple bits on each clock cycle. In [12], a pattern generator for BIST structures is introduced which can be used for VLSI circuit testing [13]. BIST will reduce the power consumption and improves fault coverage [14]. This TPG also lowers the switching activity. The patterns generated by counter and gray code generated are XOR-ed with seed produced by low-power LFSR [15, 16].

3 Methodology

3.1 Basic LFSR

LFSR are shift registers which can be used as counters also. Based on the characteristic polynomial, LFSR generates pseudorandom sequence. So LFSR are known as pseudorandom generators. An n-bit LFSR produces only upto maximum 2^{n-1} sequences. Therefore, characteristic polynomial should be chosen in such a way that it produces maximum length sequence. A 4-bit LFSR contains 4 registers and the

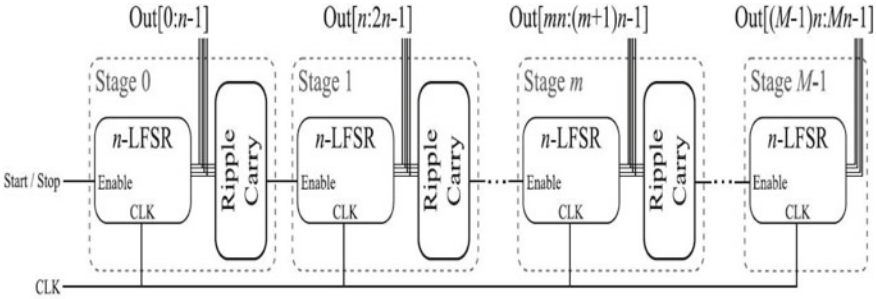


Fig. 1 Block diagram of existing multistage LFSR counter [4]

contents are shifted to right by one position on application of clock. The output of XOR gate is fed to leftmost register as feedback. As LFSR counts only upto 2^{n-1} states a sequence is missing from basic LFSR. Here, the state 0000 is lockup state so the LFSR cannot be reset to this value. Instead the basic LFSR is set to value 1111. The disadvantge of the basic LFSR is that it misses 0000 state. Therefore, some additional logic should be used so that the LFSR counts upto maximum of 2^n sequences.

3.2 Existing Multistage Counter Design

The existing multistage LFSR counter design is shown in Fig. 1. The *n*-LFSR stages are activated using an enable signal such that for a particular state change in one stage of LFSR enable signal activates next stage of LFSR. This allows the bit state space of all stages of LFSR to be covered. The limitations of existing mutistage counter design are that as it uses ripple carry logic and sequence extension logic, area utilization is more compared to normal counter design. If multistage stage counter which uses LFSR counter as first stage and binary counter for the rest of the stages could have better performance in terms of area utilization and consumption of power which is the motivation for the proposed modification in the multistage counters using binary counter.

3.3 Proposed Multistage Counter Designs

The existing multistage LFSR uses only many-to-one structure for all LFSR stages. So in the proposed method includes using one-to-many structure and random counters for the multistage LFSR model.

Multistage many-to-one style LFSR counter grouped with one-to-many LFSR structure. The LFSR used in the existing work is many-to-one style for the all the LFSR stages. In this modified method, many-to-one style is used as first stage and for subsequent stages one-to-many style feedback style is used. In many-to-one style, LFSR external XOR feedback is given, whereas in one-to-many structure internal XOR is used. Also compared to many-to-one structures, one-to-many structures has shortest clock-to-clock delay path. The advantage of using one-to-many structure the bits which are taps are XOR-ed with the output and then stored in the adjacent position. In one-to-many style, LFSR XOR-ing of bits happens within the LFSR. Therefore, the propagation delay is minimized, and thus, value of each tap is calculated in parallel. Thus, the execution speed can be increased. For LFSR with internal XOR logic also count only upto 2^{n-1} states. Therefore, sequence extension logic which consists of NOR and XOR gate is used to extend the sequence to 2^n as shown in Fig. 2.

Multistage LFSR counter grouped with binary counter. Here, the binary counter used is down counter. The first stage of multistage LFSR counter is LFSR counter and for second or for subsequent stages down counter can be used as shown in Fig. 3. By this grouping, the complexity of the circuit is reduced. Apart from the first stage, for the subsequent stages any of the binary counters like up counter

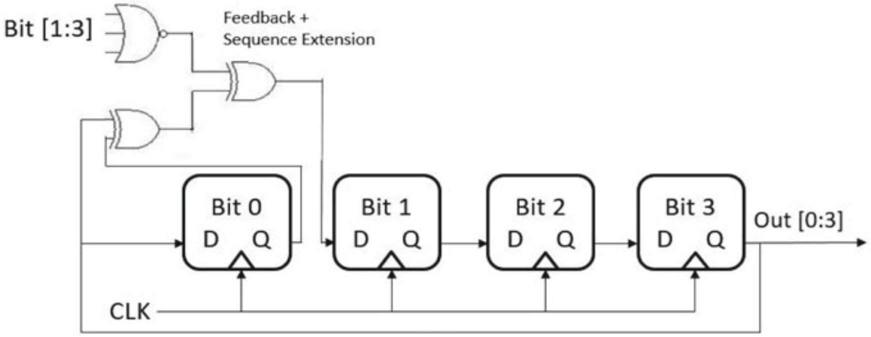


Fig. 2 Proposed one-to-many 4-LFSR with sequence extension logic

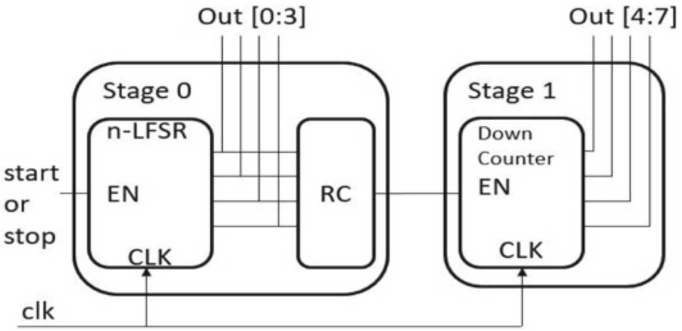


Fig. 3 Proposed multistage LFSR counter grouped with binary counter

or random counter can be used. The only disadvantage of using down counter is that as the sequence is predictable it is prone to attacks. This can be avoided by using random counters which do not generate numbers in binary order. As random counters generates numbers randomly, it can be used in testing applications. However, by grouping, LFSR counters with binary counters decrease circuit complexity as well as power utilization can be minimized.

Multistage LFSR with single-decoding logic using MUX. In this modification, each LFSR stage output is given to multiplexer as shown in Fig. 4. So the MUX will select which LFSR stage to be sent to the decoding logic for converting into binary value on the basis of select pins. This figure shows 4-bit LFSR with 2 stages connected to a 2:1 MUX and the output of MUX is given to the decoding logic When select pin = 0, the MUX will select the first stage and and decoding logic starts counting. Once the counting of first stage is done then the ripple signal is generated. The ripple signal activates the second stage and will begin the decoding.

This modified LFSR model is also comprised of sequence extension logic, ripple carry logic and decoding logic. The additional logic used here is the multiplexer compared to the existing logic. Also no need for separate decoding logic for each stage. Thus, this modification utilizes less area, and it can be used in applications of large array of counters where only one LFSR need to be activated at a time. This design is also synchronized with clock and every transition is happening with respect

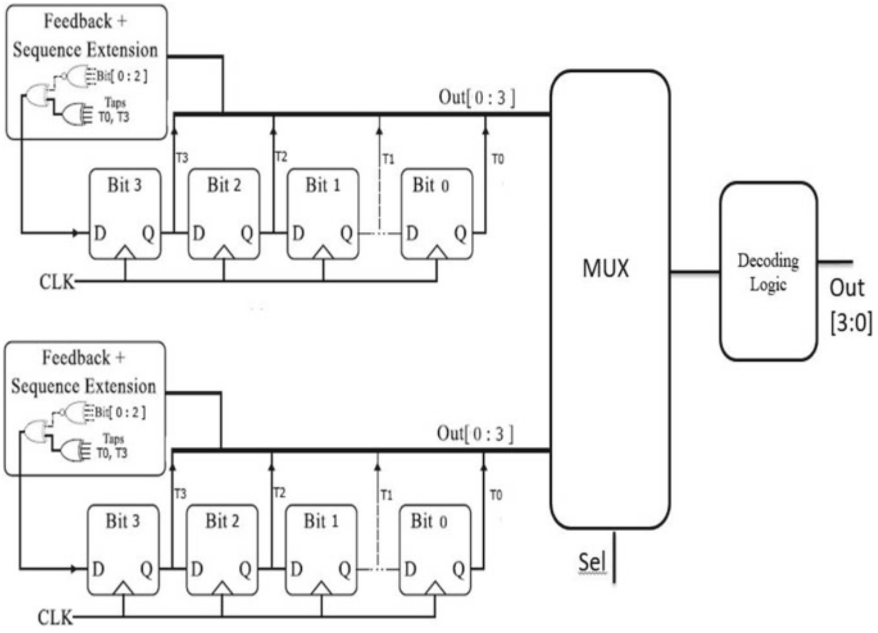


Fig. 4 Proposed multistage LFSR with single-decoding logic using MUX

to clock signals. The counter is designed in a way such that LUT can directly decode each stage correctly in a single-clock cycle.

3.4 Sequence Extension Logic

The n-LFSR stage used is of maximal length. The n-LFSR has maximum length up to 2^{n-1} so extra logic should be used for including the missing state. This is attained using sequence extension logic. This logic consists of NOR gate and XOR gate to deactivate the feedback logic when 0001 state is reached.

The sequence extension logic which can be used to extend the length of the sequence to $2n$ instead of $2n - 1$. So this additional logic is added to the basic LFSR to extend the sequences to maximum length. Thus, this design can be applied in applications like self-starting counters which requires all states to be covered. Many feedback methods can be used like many-to-one, one-to-many and ring generators. Ring generators are considered as the ideal choice for LFSR implementation. But as sequence extension logic needs extra logic along with the feedback logic in the existing counter design many-to-one style is used.

3.5 Ripple Carry Logic

As each LFSR will traverse through all the states, so it should comprises of transition from 1111 to 0111. Every LFSR design comprises of this transition as it is a gray code transition. Thus, this transition can be considered better for the ripple carry logic transition. So 0111 becomes the start of every LFSR stage and can be decoded as 0000.

If LUT is designed in such a way that it can directly decode all stages in one clock cycle, then ripple signal need to go through each stage and check if the transition happens in every stage. This leads to addition of extra logic to the counter which will decrease the counter performance. Therefore, the ripple signal is used only to activate the next stage. The ripple for succeeding stages is transferred to upcoming clock cycle so that no additional stage will be add on to the counter. Figure 5 shows the ripple carry logic.

For every LFSR stage, when transition occurs from 1111 to 0111 first ripple signal RIPPLE0 is produced. This ripple signal is then applied to activate the next LFSR stage to begin from 1111 to 0111 transition. Hence, again ripple signal is enabled which will activate the subsequent LFSR stage. In this manner, ripple carry logic leads to a delay in transition edge for one clock cycle in each stage. This delay in transition may lead to error in decoding logic and can be avoided using additional logic along with decoding logic.

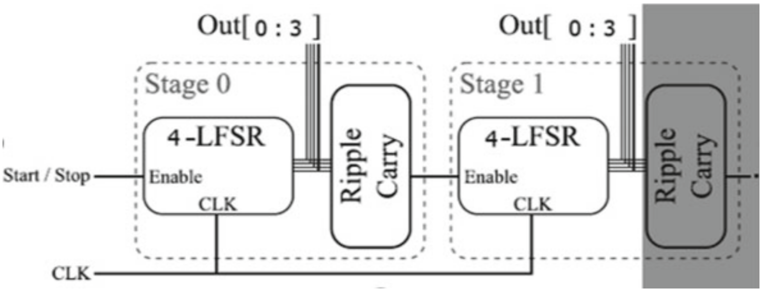


Fig. 5 Ripple carry logic

3.6 Decoding Logic

The pseudorandom sequence from the LFSR stage is given as input to the decoding logic. This logic converts the count order into binary. The decoding logic consists of an LUT where all the states are predefined to a known count value. So that each LFSR stage is decoded separately by the decoding logic. But extra logic is essential to rectify the errors due to the delay in transition. There are mainly two types of errors arise due to transition delay: initial errors and overflow errors [4]. Initial errors arises as the counter is finished on the clock cycle prior to transitions of *m*th stage. Another error called overflow error occurs when previous stage value is decoded as 0x…FF due to some error. Whenever an error is detected on the next stage, invalid next stage register is activated, and the error is rectified in the coming clock cycle. Errors are usually one value lesser compared to the actual value. Therefore, the invalid next stage register either choose the correct LUT output or one is added to the output of LUT. Overflow error can be spotted by doing AND operation of the invalid next stage register with incrementer or carryout.

This decoding logic may find very useful in applications involving large array of counters. The delay from each LFSR stage is compensated in this logic and also it saves area and improves performance compared to binary counters.

4 Results and Discussion

Figure 6 shows the existing multistage counter designed and simulated in Xilinx Vivado 2017.2. The circuit is operating based on the clock and reset value. By enabling the first LFSR, its will start it transitions and decoder will start its counting. When first LFSR output becomes 1111, ripple carry logic signal is activated. This ripple will act enable for the next LFSR stage and next stage transitions will occur. Likewise, the subsequent stages will work. At the same time, all decoders will also start to operate.

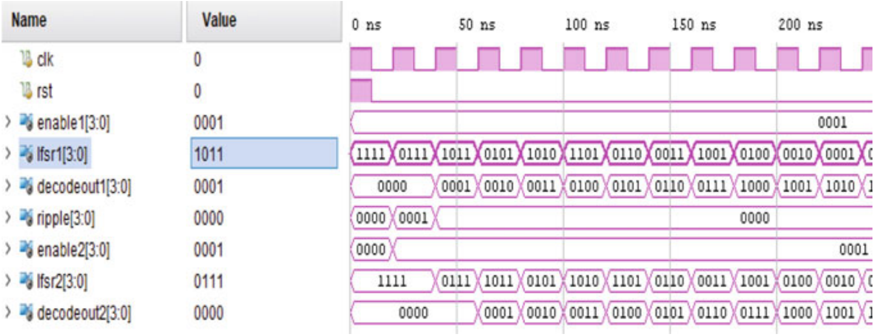


Fig. 6 Simulation result of multistage LFSR counter using methodology proposed in [4]

Figure 7 shows the proposed multistage LFSR counter with many-to-one structure as first stage and one-to-many as the second stage. Here, also with the transition from 1111 to 0111 in the first stage, ripple signal is generated. This ripple signal will activate the second stage. Except for the first stage for all the subsequent stages one-to-many style LFSR is used which will comparatively decrease the delay compared to the first stage as one-to-many structure has shortest clock-to-clock delay path.

The proposed modification 2 multistage LFSR counter is designed with first stage using external XOR feedback and for rest of the stages binary counter can be used. Here down counter is used for the second stage. This improves the area, delay and power compared to the existing counter. Figure 8 shows the proposed multistage LFSR counter with first stage using external XOR feedback and subsequent stages with down counter. This method combines the use of both LFSR counters and binary counters.

Figure 9 shows proposed multistage LFSR counter where decoding logic is selected using MUX. For the existing design each LFSR stage has separate decoding logic. But in this method after counting the first stage only the decoder will start

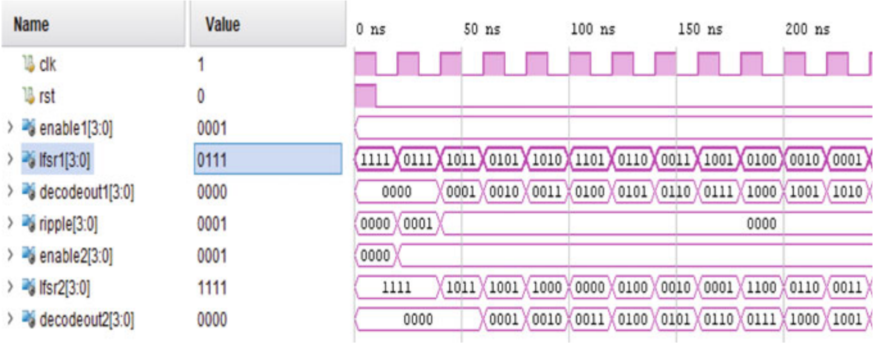
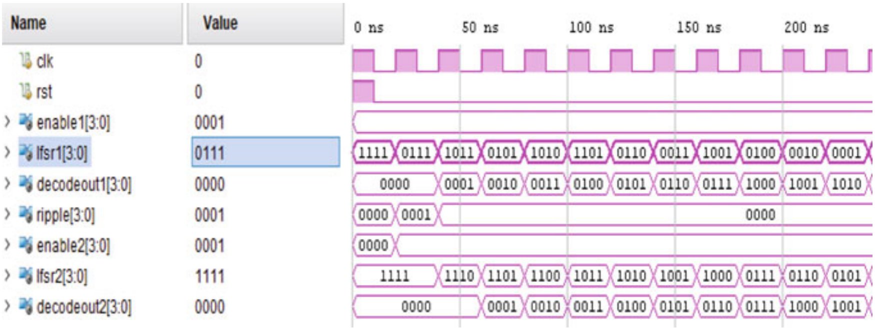
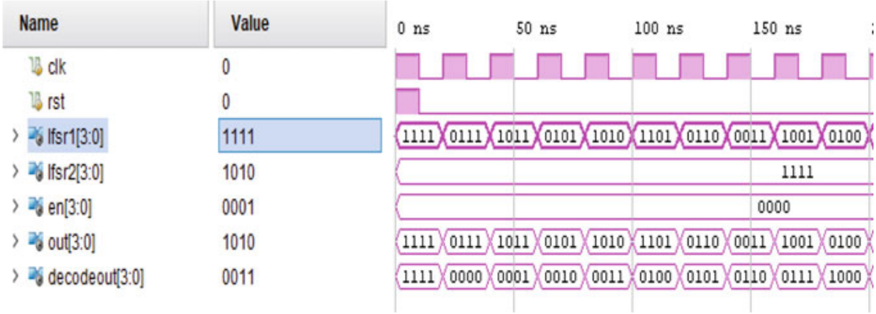


Fig. 7 Simulation result of proposed multistage many-to-one LFSR counter grouped with one-to-many LFSR structure



Name	Value
clk	0
rst	0
enable1[3:0]	0001
lfsr1[3:0]	0111
decodeout1[3:0]	0000
ripple[3:0]	0001
enable2[3:0]	0001
lfsr2[3:0]	1111
decodeout2[3:0]	0000

Fig. 8 Simulation result of proposed multistage LFSR counter grouped with down counter



Name	Value
clk	0
rst	0
lfsr1[3:0]	1111
lfsr2[3:0]	1010
en[3:0]	0001
out[3:0]	1010
decodeout[3:0]	0011

Fig. 9 Simulation result of proposed multistage LFSR counter using single decoding logic using MUX

counting the second stage. With the use of MUX only one stage decoding is enabled at a time. So this method can be used in applications of large array of counters which requires the activation of one LFSR at a time. Using the multiplexer for selective LFSR decoding logic gives improvement in area, power and delay.

All the proposed models are compared with the existing models in terms of area, power and delay as shown in Tables 1 and 2. From Table 1, it is observed that compared to existing multistage LFSR counter, the proposed multistage LFSR counters have increased speed and less power utilization. For the proposed modification 1, one-to-many structure is used as the second stage. For this case, a slight decrease in the delay is there compared to the existing work. As for LFSR with external XOR feedback various bits have to XOR together or cascaded, whereas with internal XOR uses 2 input XOR gate so propagation delay is reduced. The limitation of the proposed modification 1 is that as the LFSR stage increases the one-to-many LFSR structure and sequence extension logic can't be put to single-logic block for minimizing the logic. The proposed modification 2 is consuming less power as LFSR counter is used only for first stage and for rest of the stages down counter or any binary counter is

Table 1 Comparison of power and delay of existing work with the proposed multistage LFSR counter designs

Type of LFSR design		Power (W)	Delay (ns)
Methodology proposed in [4] designed and simulated in Xilinx Vivado 2017.2		14.084	1.833
Proposed modification 1	Stage 1: many-to-one style Stage 2: one-to-many style	14.029	1.783
Proposed modification 2	Stage 1: many-to-one style Stage 2: Down counter	11.876	1.783
Proposed modification 3	Two stage LFSR with single-decoding logic using MUX	9.238	1.657

Table 2 Comparison of area of existing work with the proposed modified multistage LFSR counter designs

Type of LFSR design		Resource	Utilization	Available	Utilization (%)
Methodology proposed in [4] designed and simulated in Xilinx Vivado 2017.2		LUT	12	41,000	0.03
		FF	23	82,000	0.03
		IO	30	300	10
Proposed modification 1	Stage 1: many-to-one style Stage 2: one-to-many style	LUT	10	41,000	0.02
		FF	21	82,000	0.03
		IO	30	300	10
Proposed modification 2	Stage 1: many-to-one style Stage 2: Down counter	LUT	12	41,000	0.03
		FF	19	82,000	0.02
		IO	30	300	10
Proposed modification 3	Two stage LFSR with single decoding logic using MUX	LUT	10	41,000	0.02
		FF	9	82,000	0.01
		IO	22	300	7.33

used which has less circuit complexity. So power consumption is reduced to value 11.876 W.

In the proposed modification 3, multiplexer is used to allot the decoding logic for each stage of multistage LFSR. Here, only single decoding logic is used. Therefore, the number of decoding logic is minimized. So the power consumption is less compared to the existing work as shown in Table 1.

From Table 2, it is observed that for proposed modification 2, as down counter is used as second stage the circuit complexity is reduced which results in less area utilization compared to the existing counter design in [4] simulated using 2017.2. The limitation of the proposed modification 2 is as down counter is used, and it is prone to attacks since the sequence generated can be predicted. This can be avoided with the use of any random counter for the subsequent stages along with LFSR counter. For the proposed modification 3, as single-decoding logic is used for converting sequences into binary, the area utilization is reduced compared to existing LFSR counter design. The limitation of the proposed modification 3 is that as the LFSR stage increases, the delay of the design is expected to increase as single-decoding logic is used. However, the area utilization will be reduced with the increase in number of LFSR stages as only one decoding logic is there for converting sequences to binary by using multiplexer. So with the increased number of LFSR stages, there will be trade-off between area and delay. Here as only two LFSR stages are used, delay is less compared to the existing counter design.

5 Conclusion

The modified counter comprises of LFSR and binary counters which are triggered on the basis of a particular transition in the previous stage. With the use of binary counters like down counters the area utilization of the modified LFSR counter is less compared to the existing multistage LFSR counters. Grouping the LFSR counters with the binary counter also minimizes the power consumption. The grouping of Galois LFSR with the Fibonacci LFSR also saw improvement in speed as propagation delay for internal XOR feedback is less. The multistage LFSR counter with single-decoding logic has less area utilization due to less circuit complexity and minimum power utilization. This can be used for applications of large array counters where only one LFSR is required to be activated at a time. The limitations of the proposed method are that counters used are not of varying length. As a future work, this multistage LFSR design can be generalized where different counters with varying length can be used for each stage with the help of same ripple carry logic.

References

1. A. Ajane, P.M. Furth, E.E. Johnson, R.L. Subramnayam, Comparison of binary and LFSR counters and efficient LFSR decoding algorithm, in *Proceedings of IEEE 54th INt. Midwest Symp. Circuits Syst. (MWSCAS)* (2011), pp. 1–4
2. M. Klosowski, W. Jendernalik, J. Jakusz, G. Blakiewicz, S. Szczepanski, A CMOS pixel with embedded ADC, digital CDS and gain correction capability for massively parallel imaging array. IEEE Trans. Circuits Syst. I Reg. Pap. **64**(1), 38–49 (2013)
3. D.W. Clark, L.J. Weng, Maximal and near maximal shift register sequences: Efficient event counters and easy discrete logarithms. IEEE Trans. Comput. **43**(5), 560–568 (1994)
4. D. Morrison, D. Delic, M.R. Yuce, J.-M. Redoute, Multistage linear feedback shift register counters with reduced decoding logic in 130-nm CMOS for large-scale array applications. IEEE Trans. VLSI. **27**(1), 103–115 (2019)

5. V. Sushmita, T. Selva Muniswari, M. Pune Dharshini, S. Rathina Selvi, T. Athieswari, Design of multistage linear feedback shift register based counters using CMOS logic style. Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng. **8**(4), 63–69 (2020)
6. K.J. Sham, S. Bommalingaiahnapallya, M.R. Ahmadi, R. Harjani, A 3x5-Gb/s multilane low-power 0.18-um CMOS pseudorandom bit sequence generator. IEEE Trans. Circuits Syst. II Exp. Briefs. **55**(5), 432–436 (2008)
7. N. Mukherjee, J. Rajski, G. Mrugalski, A. Pogiel, J. Tyszer, Ring generator: an ultimate linear feedback shift register. IEEE Comput. **44**(6), 64–71 (2011)
8. P. Alfke, Efficient shift registers, LFSR counters and long pseudorandom sequence generators, Xilinx Inc., San Jose, CA, USA, Tech, Rep, XAPP 052, 1996
9. W. Li, X. Yang, A parallel and reconfigurable united architecture for Fibonacci and Galois LFSR, in *7th International Conference on Intelligent Human-Machine System and Cybernetics* (2015), pp. 203–206
10. P.K. John, P. Rony Antony, Complete data and address generation for memory testing using modified LFSR structures, in *2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT)* (2017), pp. 1344–1348
11. D. Datta, B. Datta, H.S. Dutta, Design and implementation of multibit LFSR on FPGA to generate pseudorandom sequence number, in *Devices for Integrated Circuit (DevIC)* (2017), pp. 346–349
12. A. Kavitha, G. Seetharaman, T.N. Prabakar, S. Shrinithi, Design of low power TPG using LP-LFSR, in *3rd International Conference on Intelligent Systems Modelling and Simulation* (2012), pp. 334–338
13. N. Mohan, J.P. Anita, Compact test and diagnosis pattern generation for multiple fault pairs in single run. J. Eng. Sci. Technol. **15**(6), 3820–3835 (2020)
14. P.A. Kumar, J.P. Anita, Implementation of hybrid LBIST mechanism in digital circuits for test pattern generation and test time reduction, in *Proceedings of the International Conference on Communication and Electronics Systems* (2020), pp. 243–248
15. N. Mohan, M. Aravinda Kumar, D. Dhanush, J. Gokul Prasath, C.S. Jagan Sai Kumar, Low transition dual lfsr for low power testing. Lecture Notes Netw. Syst. **145**, 397–406 (2021)
16. R.S. Durga, C.K. Rashmika, O.N.V. Madhumitha, D.G. Suvetha, B. Tanmai, N. Mohankumar, Design and synthesis of LFSR based random number generator, in *Proceedings of the International Conference on Smart Systems and Inventive Technology* (2020), pp. 438–442

Machine Learning Based and Reconfigurable Architecture with a Countermeasure for Side Channel Attacks

Babu Illuri, Deepa Jose, Shiela David, and M. Nagarjuan

Abstract Cryptographic devices, which are embedded in SoC systems and mathematically secured algorithms, are used for operating it. But a side channel leakage may cause the secret data in these systems to be at high risk. Particularly, cryptography circuits like ECC algorithm are prone to power attacks like correlation power analysis (CPA) and differential power analysis (DPA). Hence, it is necessary to safeguard the sensitive information using proposed method having ECC with countermeasure. This paper focuses on protecting sensitive information using chaotic countermeasure as reconfigurable architecture by adopting ARTIX-7 board and offline ELM algorithm to detect the attack. In this, an effective countermeasure is designed and implemented using the FPGA for attack detection. Moreover, the proposed chaotic architecture has been integrated with ECC bit key mounted on ARTIX-7 board and tested nearly with 250 power traces recorded from the architecture. We have compared the proposed chaotic maps with the other current countermeasure technique, such as WDDL, IVR, and inductive ECC, in which the proposed design outperforms the above described existing structures in terms of area usage, power overhead, and frequency overhead. And to verify the strength of the encryption algorithm by using NIST to ensure that the given random number can be used for cryptographic purposes.

Keywords ECC · Chaotic · Side channel attacks · ELM · NIST · Countermeasures

B. Illuri · M. Nagarjuan
Department of ECE, Vardhaman College of Engineering, Hyderabad, India
e-mail: i.babu@vardhaman.org

D. Jose (✉)
KCG College of Technology, Chennai, TamilNadu 600097, India
e-mail: deepa.ece@kcgcollege.com

S. David
SRM Institute of Science and Technology, Ramapuram, Chennai, TamilNadu, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_14

1 Introduction

Nowadays, side channel attacks are gaining its importance in the field of automotive especially in the connected cars and many security devices. The advances made in information technology have led to the development of the Internet of Things. There are smart cards, mobile payment facilities, social networking platforms, and more. However, there are also many security risks involved. This has made data security a significant focal point of this century, both in the non-military personnel and military territories.

In 1990s, Kocher [1] coined the term side channel attacks (SCAs) to refer to the physical attacks that make use of additional sources of information such as power consumption, timing information, sound and electromagnetic emissions (EM), etc., to steal secret information. Since then, the designing and deployment of security measures [2] and the assessment of these [3] are extensively performed in the field of cryptography. These attacks reveal the key of secured and cipher devices by collecting the power traces from the different architectures. The application of statistical processing techniques such as correlation power analysis has been used to analyze the power traces which leads to the cipher key modifications and also reveals the critical important information. Many hardware countermeasure methodologies were studied to enhance the cryptographic measures. Many measures such as wave dynamic differential logics (WDDL) [1], inductive AES, and in-voltage regulators [4] were used as the exiting integration in the hardware's to provide an effective countermeasures.

Our contribution is bi-folded. First we have proposed light weight logistic chaotic maps for the countermeasures with the categories of initial conditions for an effective countermeasure methodology. Secondly, we have proposed the area and power centric hardware architecture for the proposed chaotic methodology which can finds its suitability for an effective prevention of side channel attacks.

The remaining paper organized into a sections: In Sect. 2, it tells about the previous work, Sect. 3 explains the proposed work, and result analysis, and finally, Sects. 4 and 5 given conclusion with future scope.

2 .

2.1 Related Works

Side channel attacks on security devices are on the rise, according to this literature review. Since everybody is using the Internet of Things to exchange confidential information, security risks are rising [5]. Generally, side channel attack is a reverse engineering method. To implement the side channel attack, required components are 1. CRO 2. Hardware (Target Device) 3. To analysis power traces from target device 4. To retrieve the sensitive information from power traces data base.

Previously, side channel attack was only accessible using an offline method [6], such as collecting power traces from a protected hardware and then analyzing the sensitive information with the MATLAB tool [7]. But now the machine learning concept has been added to enhance the analysis of the sensitive data from hardware devices using machine learning algorithms [8]. Even though a few papers have limits, such as interpreting sensitive information but not offering countermeasure strategies for side channel attacks using machine learning algorithm [9].

Srivastava and Ghosh [10] in the side channel, leakage pre-processing is required, and the leakage is a combination of noisy and statistical analysis methods to multiple traces. Misaligned traces or noise in the traces collected reduce the correlation between the traces and the data processed by the device. So, point of interest is to detect the side channel effect traces by using 70% training and 30% testing process to get better analysis [11].

Shan et al. [6] in this paper, the author presented side channel attack using machine learning algorithm by adopting hamming distance method to prevent the SCA attack.

In paper [7], authors presented a design of the logistic map by means of FPGA. The VHDL-93 hardware description language was introduced, and the effects were simulated using the Altera Model-Sim package. The main of the project was to produce a chaotic system with a low energy and time cost.

The author [8] discussed about the FPGA architectures for logistic maps which are targeted for chaotic encryption. In this methodology, author has proposed the modified logistic maps which works on the random feedback mechanism. The various advantages such as low power and reduced area overheads were the main features of this methodology. But in this paper, we focused on the limitation of the literature survey, that is, the ELM machine learning algorithm with a chaotic countermeasure to detect and protect sensitive information using the hybrid ELM algorithm with a chaotic method. But as per literature survey, only concentrate on retrieve the sensitive information and over the pre-process problem using filter method with machine learning algorithm.

Zhao et al. [9] in this paper, the author presented software-based power analysis using Ring Oscillators. And also side channel attack used as a FPGA- FPGA and FPGA-GPU communication experiments at various power traces are captured.

2.2 *Power Attack Method*

We can obtain the sensitive key from the protected board (used as FPGA) using the power attack process, which includes putting a resister between VSS and VDD across the FPGA board's power line. And the attacks are divided into two category (1) CPA(Correlation Power Analysis) (2) DPA (Differential Power Analysis).

In CPA: CPA is a method of attack that helps us to locate a hidden encryption key on a victim computer.

In DPA: DPA is an attack tool that enables us to quantify several power traces for each operation and retrieve the hidden key from the protected board.

2.2.1 Machine Learning Algorithm for Classification

To classify the power traces data in this article, we used a classification algorithm called extreme machine learning (Supervised Machine Learning).

2.2.2 Principal Component Analysis

PCA analysis is used to minimize noise in power traces. This is because collecting power traces from a protected device is typically very large. As a result, we choose PCA analysis as the best pre-processing method for feature selection. It splits the main variances data into principal components data for specifica data.

3 .

3.1 Proposed Chaotic Maps—A Brief Overview

The purpose of using ELM with counter measurement is explained in detail in this section. In this study, the Python 3.5 interface with the Extreme Learning Machine (ELM) algorithm model (Fig. 1) is used to train power trace labels to detect a SCA attack. The Elliptical Curve Cryptography is used as a reconfigurable architecture with a chaotic countermeasure to reduce the overhead region of the chip in shown in Fig. 2. To capture traces using UART as a program running on the CPU to extract

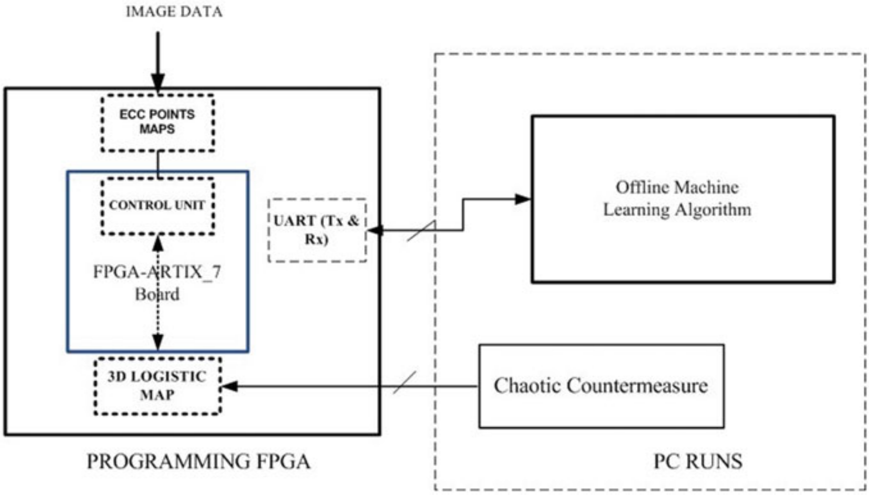


Fig. 1 Co-design of proposed work-hardware and software perspective using ARTIX-7 board

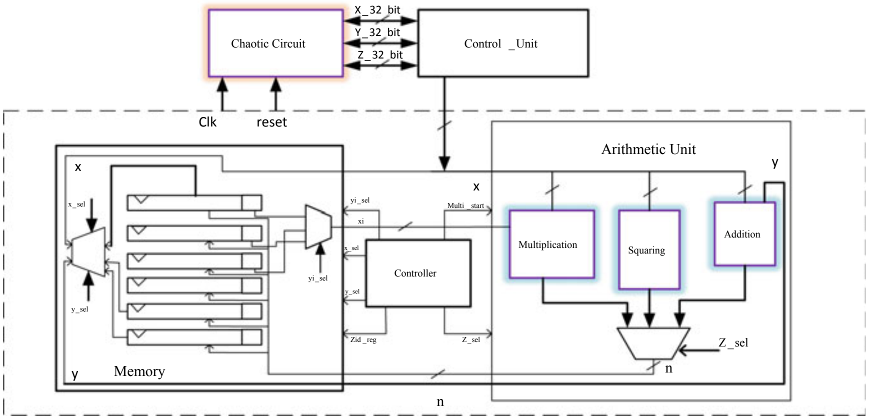


Fig. 2 Overall implementation of proposed architecture

hidden information and converts them to an.h5 or.CSV filefor further data analysis. This format is used as input to the Extreme learning algorithm to do further processing to detect an attack. Here attack means that we are trying to change hamming intermediate bits, so that the data can change, we can quickly recover the confidential information.

In previous work [8], it can easily retrieve the sensitive information from secure board even it is encrypted using ECC, but in case of proposed work, it includes the sensitive information via chaotic countermeasure using 3D logistic maps for further prevention of the attacks. This paper includes 3D Lorentz logistic maps equation as given below equations

$$\mathrm{d}x/\mathrm{d}t = s(y - x) \tag{1}$$

$$\mathrm{d}y/\mathrm{d}t = -xz + gy \tag{2}$$

$$\mathrm{d}z/\mathrm{d}t = -gx + yd \tag{3}$$

The above equation for $S = 10$ and $g = 20$, $d = 35$ different values are assigned (Figs. 3 and 4).

a. For Initial condition $s = 10$, $g = 20$ $d = 35$
b. For other initial condition $s = 15$, $g = 23$ $d = 37$

To produce the key with elevated randomness, the above chaotic equation with the initial conditions is used. Any point of the ECC that is given as the inputs are diffused with the newly generated keys. For 'N' times for the diffusion process, newly created keys are formulated. The 'D' vector is structured in the E matrix after the development of the new key, and the length of the E matrix is scaled to input data

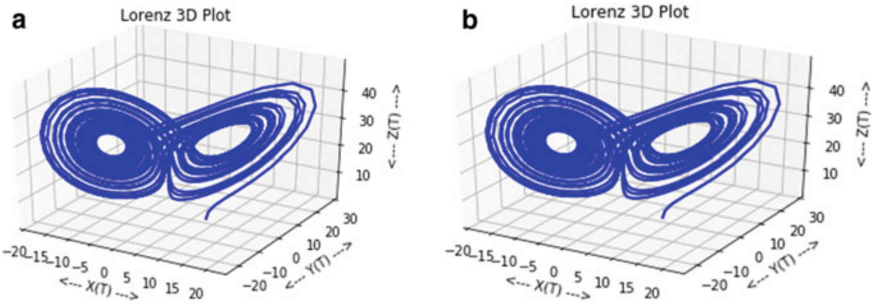


Fig. 3 Projection on X–Y plane−chaotic characteristics for the above equations

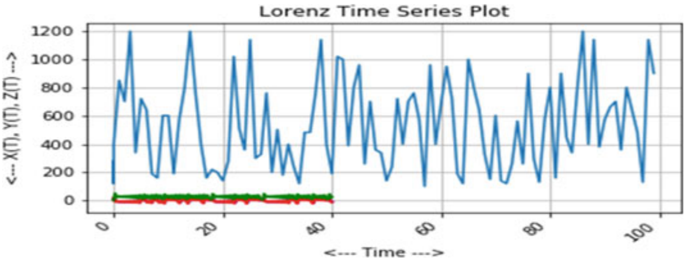


Fig. 4 Nonlinear characteristics of the proposed system designs

streams to prevent problems with data aliasing. The general mechanism of diffusion used in the proposed approach is as follows:

$$\alpha = \sum D(i) \bmod 256 \,\text{where}\, i = 0,\ 1,\ 2,\ 3 \ldots 256 \tag{4}$$

$$\beta = E_i + \alpha + D(i) \bmod 256 \,\text{where}\, 0,\ 1,\ 2,\ 3 \ldots M \tag{5}$$

where α is the diffusion constant and β is the diffusion process.

3.2 Extreme Machine Learning Algorithm

Using Python numpy and MATLAB, we can analyze function formation of specific critical data by taking various power traces from a secure board and classifying them into labels (using bespoke code). For classification, ELM algorithm is a single hidden layer feedforward neural network to overcome training issues, used weka and orange3 tool for training and testing [15, 16]. The ELM architecture consists input/output layer and concealing neurons with random weights and bias values. 'C' is the output sheet, 'K' is the consecration nodes, and 'M' is the input matrix. These neuron values are commonly given as diverse range matrixes. Additionally, random weights are created as W_{ij} and bias values as β_{ik}matrix [14] (Fig. 5)

From Fig. 6, circle number 1 is taken power traces from the FPGA board and translated to labels to observe the attack technique. In circle number 2, we are going to separate the training set and test set to evaluation further analysis. And finally, the

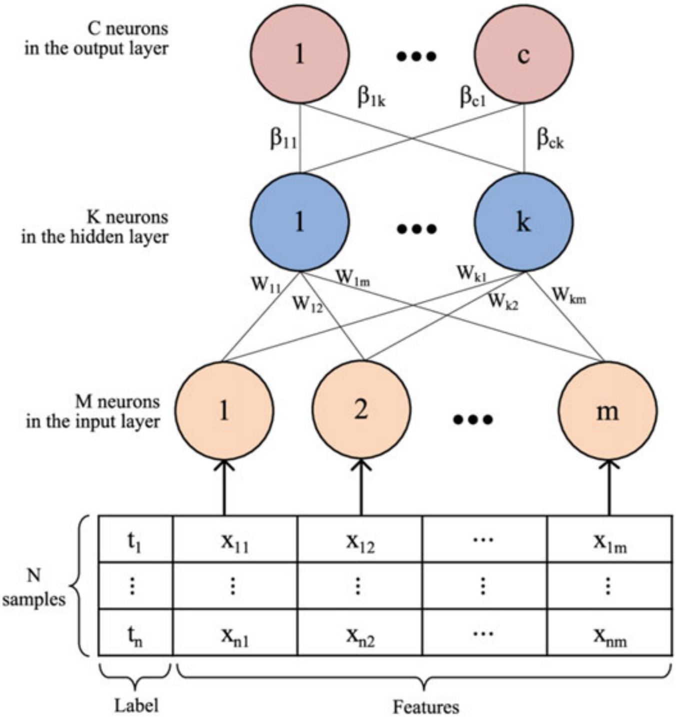


Fig. 5 ELM architecture flow diagram

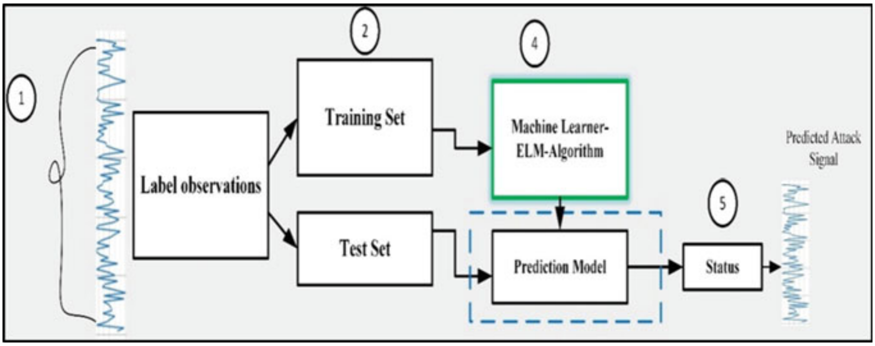


Fig. 6 ELM algorithm to predict the sensitive information

circle number 4 to evaluate critical information power traces using an ELM algorithm to determine an attack signal. We used 70% training data and 30% testing data in this study. We have used about 250 data sets.

3.3 Proposed Architecture

Figure 7 shows the proposed architecture for the proposed chaotic countermeasure methodology. The fixed point implementation is preferred over the floating point arithmetic operations for efficient operations. The initial conditions were taken as 10, 20, and 35 for producing the pseudo randomness of the attacked bits. In the proposed architecture, each stage of operation is applied with the high stages of pipelining which increases the speed of operation, since we have used the divide conquered optimization technique to effectively implement 3D chaotic maps on the hardware. In this optimization, separation in multiplication of lower and upper bits will take place simultaneously followed by the cumulative addition of all bits. This architecture is implemented in DRAM-based mechanism which will be reducing the area and power overheads.

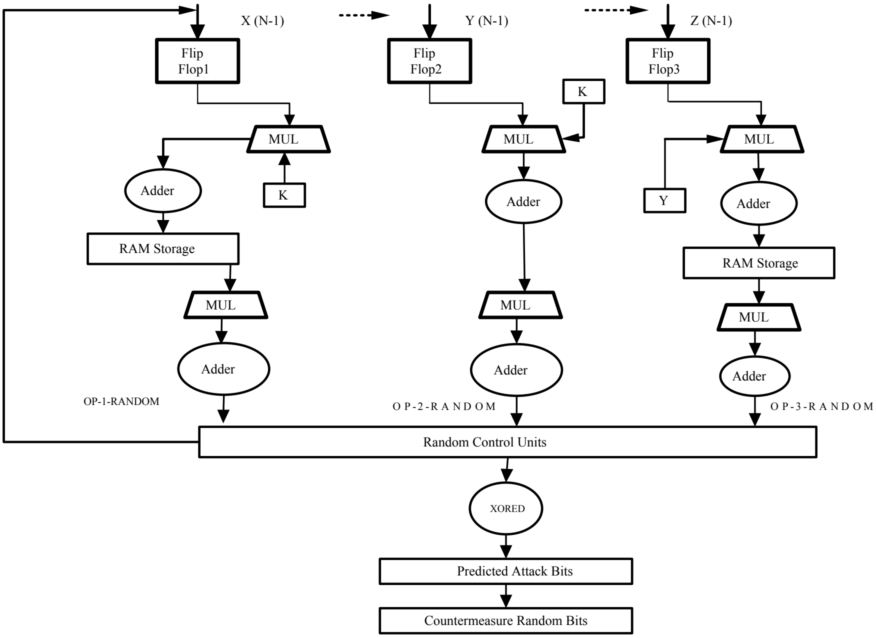


Fig. 7 Proposed architecture for the 3D chaotic logistic maps

Table 1 Illustration of FPGA EDGE board used for experimentation

Sl.No	Specifications	Features
01	Processor	Cortex-A9
02	Number of ports and pin configuration	5—PMOD connectors
03	Number of UARTS	USB host connected to ARM PS
04	No of IoT transceiver support	02(WIFI)/BLE
05	Memory	Micro SD card slot

4 .

4.1 *Experimental Setup*

We need to catch power traces from the 128-bit ECC encryption keys installed on the FPGA running at 450 MHz to perform our experiments. Various power traces were used to enforce the light weight chaotic program on FPGA, and attacks were simulated on the power traces. The features of the board which is used in the paper have been listed in Table 1.

4.2 *Simulation and Synthesis Results*

We used the XILINX VIVADO tool sets to simulate the chaotic random bits for simulating the suggested 3D chaotic program. In addition, we used the Python-based frameworks to document the power traces and used the manipulated attack bits as the inputs to the device in conjunction with the input attack bits (Figs. 8 and 9).

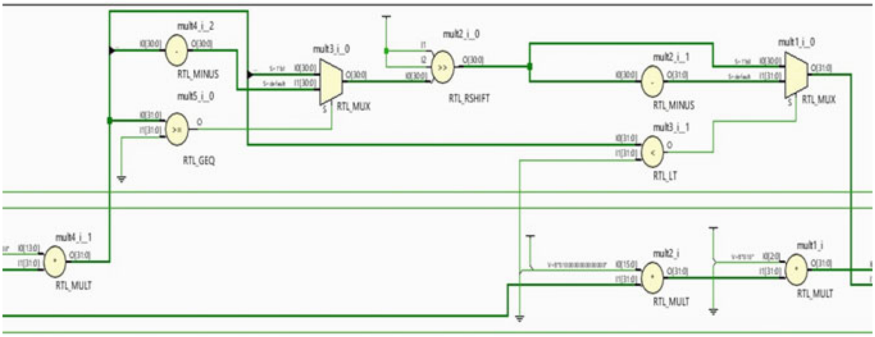
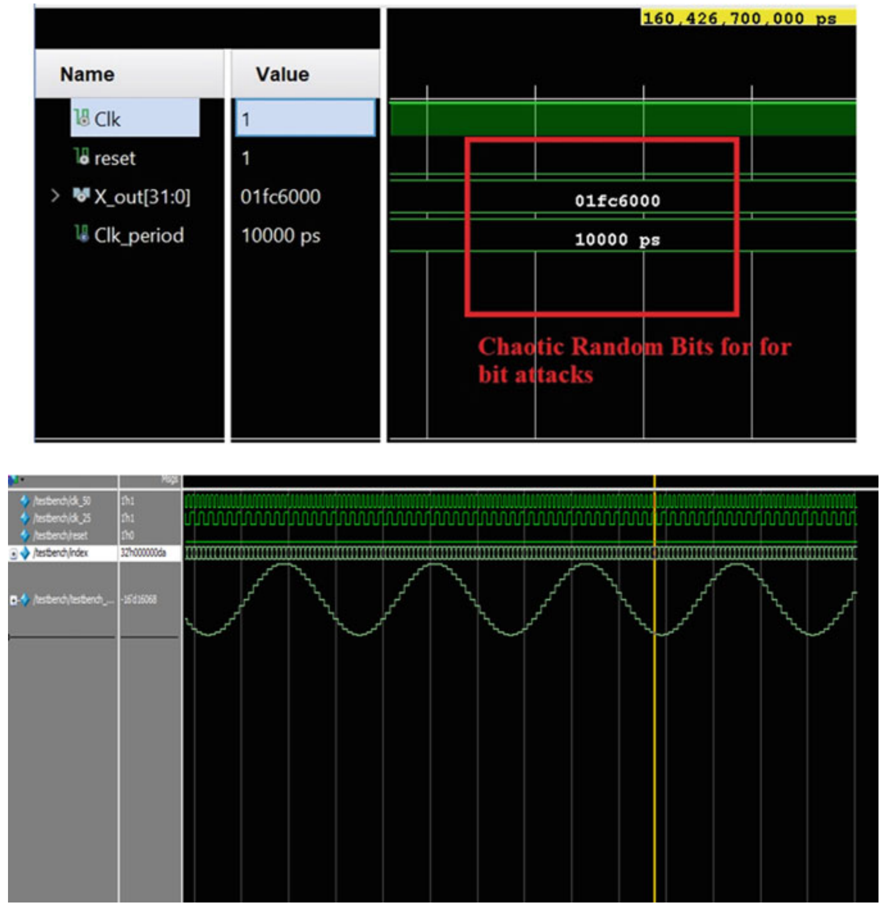


Fig.8 FPGA architecture for the proposed chaotic methodology



Name	Value
Clk	1
reset	1
X_out[31:0]	01fc6000
Clk_period	10000 ps

Chaotic Random Bits for for bit attacks

	Msgs
/testbench/clk_50	1'h1
/testbench/clk_25	1'h1
/testbench/reset	1'h0
/testbench/index	32'h000000da
/testbench/testbench_...	-16'd16068

Fig. 9 Simulation of chaotic random bits for the attacked bits which was detected in the experimentation

Also we have compared the proposed chaotic architecture with the other existing countermeasure methodologies such as WDDL, IVS, and inductive AES in terms of area, power, and frequency. Table 2 illustrates the comparative analysis between the all architectures.

Table 3 shows the comparative analysis for the randomness testing between the two framework, one without chaotic countermeasure and other one is with chaotic countermeasure. The key's continuity is examined for its randomness who's the mathematical expression is given by

$$P = \text{erfc}(|V(n)(\text{obs}) - 2n\pi(1-\pi)|)/2.828n\pi(1-\pi) \tag{6}$$

Table 2 Comparative analysis between the proposed architecture and existing architecture

Methodologies used	CMOS/FPGA	Area overhead (%)	Power overhead (%)	Frequency overhead	Design nature
WDDL	CMOS	210	270	0.023%	Complex
IVR	CMOS	250	210	3.33%	Complex
IAES	FPGA	350	250	17.1%	Complex
Hamming distance-based methodology	FPGA-Sakura boards	1	31	0	Simple
Proposed chaotic countermeasures	FPGA-Artix-7 boards	< 1	15–16	0	Simple

Table 3 Comparative study of randomness testing within the two systems

Sl. no	Image details	Test	ECC (without chaotic)	ECC-Chaotic countermeasure
1	Mammogram Images [12]	Frequency monobit	0.06377221	0.346729
		Run test	0.054252	0.564792
		DFT test	0.0672829	0.567820
2	Diabetic retinopathy ımage[12]	Frequency monobit	0.0545112	0.78490
		Run test	0.0324122	0.678390
		DFT test	0.0213421	0.536389
3	MRI brain ımage [12]	Frequency Monobit	0.0458590	0.549032
		Run test	0.0345678	0.456340
		DFT test	0.0456789	0.567390

where in this test, $V(n)$ (Obs) identicated the faster oscillations (Oscillations is considered as the switch from one to zeros) in which it occurs when there is lot of changes in the bit streams.

Table 3 presents the comparative analysis of test results between the proposed chaotic ECC architecture. As per the NIST guidelines [13], to show the high randomness in the bits, the value of P should be greater than 0.01. The randomness was demonstrated by both the system, but the proposed architecture showed strong randomness relative to the current ECC architecture. The disorderly secure ECC countermeasure approach is therefore more fitting for protecting the SPA in an IoT network during medical image transmission.

From the table, it can be seen that the proposed chaotic application has less over head in terms of power, area, and frequency when compared with other existing algorithms since because the proposed functional part is dependent on the AES architecture. Figure 10 shows that the proposed ELM module, as opposed to SVM, NB, and other methods, provided the best solution. And it is 20% more efficient than the 'NB' algorithm.

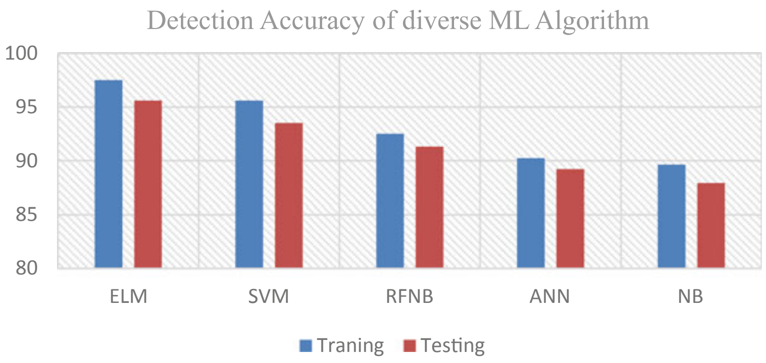


Fig. 10 Analysis of proposed with different machine learning modules

5 Conclusion

The paper proposes the FPGA implementation of chaotic countermeasures for side channel attacks in128-bit ECC with chaotic circuits. The total system is designed with no impact on frequency, power, and area which finds its suitability for cryptographic applications. Also the pseudo random nature of duplicating the bits in the network proves to be more vital and considered to be more resistant to the attacks in the networks. Meanwhile, the 3D logistic chaotic maps can be improvised to 5D or even to 7D maps which may leads even to the prevention of the side channel attacks in the network for future work. And also we planning for cloud-based analysis for side channel attack using machine learning algorithm.

References

1. D. Hwang, K. Tiri, A. Hodjat et al., AES-based security coprocessor IC in 0.18-µm (2006)
2. M. Doulcier-Verdier, D. Jean-Max, F. Jacquesm et al., Aside-channel and fault-attack resistant AES circuit working on duplicated complemented values. ISSCC Digest of Technical Papers (2011), pp. 274–275
3. A. Moradi, A. Poschmann, S. Ling, C. Paar, H. Wang, Pushing the limits: a very compact and a threshold implementation of AES'. Adv. Cryptol. EUROCRYPT 69–88 (2011)
4. M. Kar, A. Singh, M. Sanu et al., Improved power-side-channelattack resistance of an AES-128 core via a security-aware integrated buck voltage regulator. ISSCC Digest of Technical Papers (2017), pp. 141–142
5. W. Shan, X. Fu, Z. Xu, A secure reconfigurable crypto IC with countermeasures against SPA, DPA and EMA. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **34** (7), 1201–1205 (2015)
6. W. Shan; S. Zhang; Y. He, Machine learning-based side-channel-attack countermeasure with hamming-distance redistribution and its application on advanced encryption standard. Electron. Lett. **53**(14), 7 6 (2017)

7. D.A. Silva, E.B. Pereira, E.G. Nepomuceno, Implementation of the logistic map with FPGA using 32 bits fixed point standard. in *XIII SimposioBrasileiro de AutomacaoInteligente—SBAI2017*, Porto Alegre, Brazil. In Portugues, Fri, 11 Aug (2017)
8. A. Pande, J. Zambreno, Design and hardware ımplementation of a chaotic encryption scheme for real-time embedded systems. in *An Effective Framework for Chaotic Image Encryption Based on 3D Logistic Map*, Security and CommunicationNetworks (2018)
9. M. Zhao, G. Edward Suh, FPGA-based remote power side-channel attacks. in *2018 IEEE Symposium on Security and Privacy* (2018)
10. A. Srivastava, P. Ghosh, An efficient memory zeroization technique under side-channel attacks. in *32nd International Conference on VLSI Design* (2019)
11. A. Singh, N. Chawla, J.-H. Ko, Energy efficient and side-channel secure cryptographic hardware for IoT-edge nodes. IEEE Internet of Things J. **6**(1), (2019)
12. http://www.eng.usf.edu/cvprg/Mammography/Database.html
13. https://csrc.nist.gov/projects/random-bit-generation/documentation-and-software
14. B. Illuri, D. Jose, Design and implementation of hybrid integration of cognitive learning and chaotic countermeasures for side channel attacks. J. Ambient Intell Human Comput (2020). https://doi.org/10.1007/s12652-020-02030-x
15. https://www.cs.waikato.ac.nz/ml/weka/
16. L. Punitha, K.N. Devi, D. Jose, J. Sundararajan, Design of double edge-triggered flip-flop for low-power educational environment. Int. J. Electri. Eng. Educ. 2-s2.0–85070403539 (2019)
17. M. Baskar, T. Gnanasekaran, Developing efficient intrusion tracking system using region based traffic impact measure towards the denial of service attack mitigation. **14**(7), 3576–3582 (2017)

Autonomous Scheduling and Distributed Graph Routing Algorithm (ASDGRA) for Hybrid Wireless Sensor Networks

Najmuddin M. Maroof and Mohammed Abdul Waheed

Abstract The technology of wireless sensor–actuator networks (WSANs) is widely employed in the applications of sensor networks due to its wireless nature, and it does not involve any wired structure. The wireless systems that are battery-driven can easily reconfigure the existing devices and sensors efficiently in the manufacturing units without employing any cable for power operation as well as for communication. The wireless sensor–actuator networks that are based on IEEE 802.15.4 consume significantly less power. These networks are designed and built cost-effectively by considering the capacity of battery and expense so that they can be employed for many applications. The application of a typical wireless autonomous scheduling and distributed graph routing algorithm (ASDGRA) has illustrated the reliability of employing its basic approaches for almost ten years, and it consists of the accurate plot for routing and time-slotted channel hopping therefore ensuring accurate low-power wireless communication in the processing site and officially declared by the controversial statements associated with the government of Greek experiences fourth industrialization. There is a huge requirement for sensor nodes to link via WSAN in the industrial site. Also, reduced computational complexity is one of the drawbacks faced by the existing standards of WSAN which is caused because of their highly centralized traffic management systems and thereby significantly improves the consistency and accessibility of network operations at the expense of optimization. This research work enables the study of efficient wireless DGR network management and also introduces an alternative for ASDGRA by enabling the sensor nodes to determine their data traffic routes for the transmission of data. When compared to the above two physical routing protocols, the proposed technique can drastically improve the performance of a network, throughput, and energy consumption under various aspects. Energy harvesting (EH) plays a significant role in the implementation of large sensor devices. The requirement for subsequent employment of power sources is eliminated by the efficient approach of energy harvesting and thereby

N. M. Maroof (✉)
Department of Electronics and Communication Engineering, Khaja BandaNawaz College of Engineering, Kalaburagi-4, Gulbarga, Karnataka, India

M. A. Waheed
Department of Computer Science Engineering, V.T.U.P.G. Centre, Kalaburagi, Karnataka, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_15

providing a relatively close- perpetual working environment for the network. The structural concept of routing protocols that are designed for the IoT applications which are based on the wireless sensor has been transformed into "energy-harvesting-aware" from the concept of "energy-aware" because of the development in the energy harvesting techniques. The main objective of the research work is to propose a routing protocol that is energy-harvesting-aware for the various sensor networks. A novel algorithm for routing called autonomous scheduling and distributed graph routing algorithm (ASDGRA) has been developed and significantly improved by incorporating a new "energy back-off" factor. The proposed algorithm when integrated with various techniques of energy harvesting enhances the longevity of nodes, quality of service of a network under increased differential traffic, and factors influencing the accessibility of energy. The research work analyses the performance of the system for various constraints of energy harvesting. When compared to previous routing protocols, the proposed algorithm achieves very good energy efficiency in the network of distributed WSN by fulfilling the requirements of QoS.

Keywords Wireless sensor network · Distributed computing · Autonomous control · Routing algorithm · Mobile wireless sensor network

1 Introduction

Development in the technology of senor has paved the way for the design of low-powered and relatively small, sensors that are well-furnished with programming ability, efficiency in detecting various parameters, and competency in communication that are wireless. Since the sensor technology is cost-effective, the network incorporates several hundreds of sensors and thus improving the efficiency, area availability, and data precision. In obscure and undeveloped areas, the networks of the wireless sensor provide necessary information or data regarding common ecological factors, remote systems, and so on. When compared to wired communication, the network of wireless sensor offers many advantages like simplicity in designing a network (minimizing initial cost overhead), high speed (a network with relatively small sensors can be allocated over a wide area), fault tolerance (malfunctioning of one node do not impact on the network functioning), self-oriented (the reconfiguring ability of node itself), and some of the intrinsic problems faced by wireless sensors are limited bandwidth, data transmission, that is, error-free, interference-free, and so on. Since cell phones are the most widely used wireless nodes, they use only specific batteries to draw the energy and do not require any constant supply of power. Therefore, this reduces the total energy accessible to the nodes. Furthermore, these wireless nodes find it hard to replace both the sensor nodes and battery packs in few areas; therefore, it is essential to maximize the durability of networks by placing a set of new nodes that can recharge the entire area [1]. A pre-defined implementation is required to identify the nodes that are not working and preferentially substitute them by reducing a few network benefits. An optimum sensor system must possess *location responsiveness and addressing that is based on the attribute.* One more essential aspect of sensors is that they should respond instantly to significant environmental variations such as an application that are based on time. The receiver must be given information regarding other remaining nodes that possess small delay and thus ensuring efficient utilization of bandwidth in the wireless media. As a result, data-centric protocols that have data accumulation efficiency, consistently allocating power dissipation, reduced energy to maintain network durability, and eliminating the constraint of a single node (excluding BS) are essential for the wireless sensor networks. As discussed in the previous paper [2], the conventional network protocols are not applicable for wireless communication that is described for MANETs. Recently, a data transmission protocol which is energy-efficient named LEACH has been presented [3], and based on the data obtained by BS, the hierarchical clustering is achieved. However, to minimize energy, the cluster head (CH) and many nodes are frequently varied by the BS. The cluster head receives data from the sensors, analyses, and then transmits to BS. The consumption of energy is evenly allocated by arbitrarily rotating the CH if not the cluster head that is nearer to BS will not allow data to be transmitted and power dissipation compels them to perish quickly when compared to other nodes. Constant re-clustering is done by the BS to allow another active node to function as cluster head when one of the CH because for some reason is unable to interact with its node members or with the BS. The information regarding how a node is established and on what basis the cluster head must be chosen is addressed in [4].

The common drawback noticed was how to resolve the queries of the users and in what way the required data is routed. The majority of the existing protocols acknowledge that a sensor collects the information regularly from the system and when a query arrives the protocol reacts to it. In LEACH [3], the cluster head receives the information continuously, and after the process of clustering, the BS receives the information to store the data. In sensor networks, particularly the applications which are based on time are not focused specifically. The sensor networks must specify the final users to dynamically deal with energy proficiency, precision, and time taken for responding. In this work, we mainly aim at advancing a routing protocol that is efficient and a detailed request managing process that satisfies the above requirements.

The major issue faced in decentralized IoT is the propagation of cost-effective data which is illustrated by much research work [5]. Several research works have been accomplished in the area of data network aggregation [6], without any compression loss and with compression loss [7] (the main concern is to improve the efficiency of energy by minimizing the total bits to be transmitted), as well as enhancement of various objects present in the wireless communication [8]. Particularly, the specific level of compression, quality of signal trading, and quality of service are allowable for a long-lasting network in several applications of the Internet of things. In this research work, we visualize the networks of IoT that inevitably switch their activity to various sources, their positions, rate of transmission (i.e., encoding of source), distribution flow systems, and quality of service required by the application. Therefore, we propose the application to overcome the drawbacks of distributed learning techniques, reviewing source compression, and distribution flow issues. In general, we collectively describe the issues related to loss data compression at the input side and a successful routing path is established toward the data distribution center of the Internet of things (the gateway of the Internet of things is also known as a sink). This helps in analyzing the basic difference between the efficiency of the distortion rate at the inputs and the outlay required for transmitting the necessary data and addressing the issues of distribution flow. The primary objective is to interpret exactly how much computation has to be carried out near the input with the help of a few lossy data compression algorithms. Thus, the compressed data is processed successfully and communicated via a plot of a network by collectively handling compression and routing. Make sure that the compression is intrinsically combined with the efficiency of interconnected network constraints, their destinations, total inputs and their locations, and the communication node's potency. While identifying the drawbacks, we arrived at the concept of possessing transmission overheads in contrast to the interference of signal and thereby enabling the study of an ideal cost-distortion area (allocative efficiency) of a network.

The main difference between the existing and the proposed work is that no hard-coded protocol standards are used to ensure the quality of source compression and process the data collected. However, these responses generally arrive as alternatives for the process of distributed learning which contains the iterative transmission process among the local signals and the nodes [9]. The nodes do not possess any information regarding the entire network process or status of network standardization, and also, they have no idea regarding the detailed structure of the network. Instead, local communications enable the design of a distributed system to integrate the process of optimization worldwide. An initial data centralized system is introduced in the research work of [10]. Several research works have been carried out in the field of distributed WSNs, namely source encoding, forwarding, and enhancement techniques for multi-objective. Few research works are concerned with encoding of data at the input [11], and some works enable the accumulation of data at the arbitrator nodes while forwarding the data to the destination node. The paths for routing can be predetermined by the advanced computation [12] or determined with the help of a functional approximation process that requires accumulation and data forwarding network [11]. Since the system implements encoding at the input, the researchers are concerned with combined data encoding and collection of data by consuming the energy computed for data encryption in contrast to the energy utilized for the communication. The strategy discussed above is analogous to the proposed one, and the primary distinction is that the paths for routing and encoding standards at the inputs are simultaneously determined and are executed as the process of deep learning. The drawbacks of combined routing and encoding are introduced by employing the theory of Lyapunov optimization in [13]. On the other hand, the path for routing is already defined and is not implemented with the encoding procedures. The research work of explores a conceptual framework for data collection networks along with data encoding techniques, where every individual node preprocesses the data collected before delivering the information to the local network. The proposed work mainly concentrates on the cost-effective encoding and transmission scheduling for the network with a single hop and considering encoding and communication expenses under a rescheduling limit.

Many research works utilize spatial correlation for data accumulation examples [13]. These works mainly concentrate on the information generated from the temporally correlated inputs and forwards the same information to the destination. Encoding inevitably leads to in-network accumulation, and researches are conducted to explore the difference between routing and accumulation. The issue of routing was not considered by the analysis of the distributed method for compressive sensing, for example, illustrated in [10]. For the multi-objective enhancement, there exist several research works. For example [7, 8], mainly concentrates on various issues of the target. The combined optimization and routing are not taken into account even though the main aim is energy reduction. In [12], the proposed algorithm corresponds to the compression of the data source in the initial section of the research work, further routing is carried out by employing a flow-based model. We implement a heterogeneous network where the input nodes possess various detection and transmission abilities; however, it is impossible to accumulate the data flow from various sources. The network of multimedia sensors presents a few examples of these types of the network [3]. In the current research work, we propose a combined optimization method for routing and encoding. The proposed algorithm relies on ADMM resulting in a completely allocated system.

A heterogeneous distributed-edge framework has been specifically formed by many public and private networks by offering support to the application of the Internet of things. The network operators present a network function virtualization (NFV) that distinguishes the operations of the network from specific hardware by operating the features in adaptable software such as virtual network functions (VNFs) that are operated on specific hardware [14]. The network providers are offered a unique chance that is associated with software-defined networking (SDN), to set up the architecture to satisfy the requirements of a specified application [15]. Recently in fundamental cloud conditions, the research works mainly concentrates on the drawbacks of network function virtualization. One of the disadvantages involved in cloud computing is end-to-end dormancy which is due to the impact of the physical gap between the cloud service providers and a heterogeneous distributed-edge framework that has been specifically formed by many public and private networks by offering support to the application of Internet of things. The network operators present a network function virtualization (NFV) that distinguishes the operations of the network from specific hardware by operating the features in adaptable software such as virtual network functions (VNFs) that are operated on specific hardware [16]. The network providers are offered a unique chance that is associated with software-defined networking (SDN), to set up the architecture to satisfy the requirements of a specified application [17]. One of the disadvantages involved in cloud computing is end-to-end dormancy which is due to the impact of the physical gap between the cloud service providers, and an efficient technique is developed for the ideal decision-making policy and enhance the advantages for a long period. The level of accuracy and recurring is minimized by combining the efficiently handling feature of deep learning with reinforcement learning which is having decision-making capacity [18]. Deep learning is considered as one of the key technologies in achieving self-adjusted SFC enhancement because of the benefits of self and virtual learning. Cloud computing that is generally associated with software-defined networking and network function virtualization is employed in various application of IoT to guarantee the efficiency of the service provider. A security framework which is based on cloud computing is introduced in [19] is employed to safeguard the software-defined vehicular applications of IoT. The blockchain nodes examine the real-time applications of the Internet of things so that suspicious traffic and its behavior are updated by the blockchain and thus assuring data security and cost optimization. The attack of distributed denial of service (DDoS) has a significant effect on the exponential increase in the suspicious devices in the IoT applications. The model of decentralized secure DDoS collaboration solution (Co-chain-SC) [20] was introduced by Z. Abou et al. to overcome the above issue by employing smart convention to the model. The main advantage of the Co-chain-SC is that the block chain enables privacy and cost-effective distributed collaboration among various software-defined networking to reduce the attacks. Sharma et al. introduced a new software-defined-cloud architecture consisting of three layers [21] which are based on blockchain to overcome the drawback of security for outsourcing data and hence creating trust between consumer and the service providers. The request services and public setup is examined by the device layer. The distributed resources and enabling data operation are analyzed by a cloud layer. The computational resources are carried to the edge of the IoT network which is based on blockchain and software-defined networking in the fog layer. The architecture of the distributed cloud which is based on the block chain overcomes the drawback of privacy and decentralization. A unique structure combined with SDN that is used for the application of IoT is introduced by Pourvahab et al. in [20]. To assure secure synchronization between various SDN controllers for a network H. Tang et al. in [19] addressed a consensus protocol that is based on the block chain for software-defined industrial Internet. Blockchains are widely used in many distributed applications because of their service trustworthy, cost-effective, and transparent nature [14].

2 Related Work

The most widely considered technique in the research work of wireless mesh and sensor networks is routing. The routing protocol that directs any routing path in the network of the wireless sensor to an individual or many defined base stations is termed as CTP [6]. Some of the applications of CTP are research, training, and an industrial process. The development of RPL [7] has been ensured by the perception of CTP. High efficiency is not achieved by the routing protocols such as CTP and RPL which are based on the tree because it is unable to generate the data set routes described in the wireless HART. Since packet loss is the main issue of these two protocols, they are not applicable for the basic applications of industries. Therefore, to overcome the above problem, a multipath routing protocol is introduced (e.g., [11–15]) which improves the efficiency between the sender node and receiver node by enabling cluster head disjoint or connection disjoint. However, the energy utilization and load in the traffic are maintained between different network nodes by the introduction of routing protocol which is based on multipath RPL (e.g., [16–19, 21]). When compared to other protocols, the data set routing mentioned in wireless HART ensures high efficiency by considering the significant level of routing latency for the network of TSCH.

The application of a typical wireless HART has illustrated the reliability of employing its basic approaches for almost ten years and ensuring accurate low-power wireless communication in the processing site. A set of algorithms were proposed by Han et al. [20] and Wu et al. [22] to generate the data set path in a unified manner whereas, with the application of the Bellman–Ford algorithm, Modekurthy et al. introduced the generation of data set paths in a distributed manner [23]. In contrast to these techniques, a basic distributed routing protocol that is based on RPL is designed to produce and work with the data set paths. The overall solution for the network is obtained by designing a strategy of transmission scheduling which is operated just above the designed protocol. The analysis of transmission scheduling is considered as the prime concern for the applications of wireless HART networks which are based on time [24–27]. The above-discussed scheduling methods are unified alternatives that are intended to operate on a data-centric network manager with data set routing operation. Also, researches have been made for the development of RPL networks with distributed scheduling [10, 27–32]. Consider an example of Orchestra in RPL networks where the nodes can determine their schedules which are introduced by Duquennoy et al. [10]. To ensure the proper utilization of IPv6 which facilitates load above the IEEE 802.15.4e TSCH networks is regulated by the operating group of 6TiSCH developed by IETF [8]. On the other hand, the experimental results of our analysis illustrated that the network operating with RPL experiences more recovery time and reduced efficiency when node malfunction and external interference take place. Synchronous transmissions [22–27] are considered as one more area of research. But for handling the synchronous transmission, there is a necessity for a unified node in this transmission.

Michelusi et al. [28] presented an algorithm for energy-opportunistic weighted minimum energy (E-WME) concerned with the routing techniques of "energy-harvesting-aware" and thereby determining the individual sensor node overhead by employing the rate of energy harvesting and the energy available. A routing algorithm for randomized minimum path recovery time (R-MPRT) is introduced in the research work of [28]. Therefore, the cost measured for a system can be expressed as the amount of energy consumed by the packet processing node (also known as energy packet) to the rate of energy harvesting. By considering the specified cost metric, the best optimal path to the destination is determined by the node. Further, the destination receives the data packet through a low-cost connection from the source node. An algorithm for R-MPRT is introduced by Hasenfratz et al. in [29] which is developed by making use of the energy remaining at the node rather than using the energy harvesting rate to attain an optimal solution. The decision of routing is considered depending on the cost measurements that consist of energy consumption and harvested energy ruined because of overloads which are explained in [13]. The algorithm for distributed energy-harvesting-aware routing (DEHAR) is presented in [30]. This algorithm determines the best optimal path to reach the destination by employing the energy available at every individual node and hop counts of the node. The perception of the energy harvesting algorithm employs a local charge for every individual node that is constantly rationalized and is inversely related to the energy available at the node.

A routing protocol for energy-harvesting-aware ad-hoc on-demand distance vector (AODV-EHA) is introduced by Gong et al. in [31]. The algorithm of AODV-EHA deals with wireless sensor networks by acquiring the benefits of the previous ad-hoc on-demand distance vector protocols. This algorithm calculates the best optimal path with low-cost transmission by making use of efficient energy harvesting techniques. The protocol required for smart energy harvesting routing (SEHR) is presented in [32]. The algorithm of SEHR initially considers the energy accessibility at the node, expected rate of energy, and approximate the energy collected from the renewable sources (e.g., solar and radiofrequency) during the process of selecting the path. A routing protocol for energy-harvesting-aware that is based on the topology control scheme is introduced in [33]. Therefore, to speed up the topology of the network, the energy-harvesting-aware protocol uses the method of game-theoretical to evaluate the status of energy and extracting capacity of every individual node. The power consumption of the node is collectively evaluated by it or by the adjacent node energy resources and further examines the energy produced and extracted at every individual node with various time intervals. Although routing algorithms that are discussed above minimize the energy consumed and increase the durability of nodes, still these algorithms possess some drawbacks. The feature of energy harvesting introduced in [10] and [12] will not consider the actual amount of energy extracted. The concept proposed in [12] and [21] employs perpetual replacement rates for all the nodes present within the network. Therefore, almost all existing algorithms are unable to interact with the probabilistic feature of the resources that are renewable because of the inadequate extraction of energy. One more drawback is the implementation of a single energy source to obtain energy. It is indicated that renewable energy sources are supported by peak/off-peak intervals, climatic changes, seasons, and day and night rotations. Thus in real-time, the application of a single energy source may not be feasible to increase the longevity of the node. As a result, a novel routing algorithm for energy-harvesting-aware [34] is developed by considering various types of energy extracting algorithms. In the case of the stratified environment, the overall data is incapable of adjusting to the differences in the energy status of the sensor nodes (e.g., the remaining, utilized, and collected energy levels). Thus, it is required to create tables for routing by making use of local data in a decentralized manner for the applications of the Internet of things [35].

The framework of efficient routing which is known as directed diffusion is presented by Intanagonwiwat et al. [36] and is utilized for the networks of the sensor. This framework illustrates the concept of data-centric along with the application of data input transmission and encoding. The algorithm of hierarchical clustering mainly focuses on distributed activity, the requirements for communication that are asymmetric, and energy consumption in the networks of the sensor is addressed by Estrin et al. [37]. Jiang et al. in [38] introduced a routing protocol for the networks of mobile ad-hoc and is termed as a cluster-based routing protocol (CBRP). The nodes of a network are split into non-overlapping and interesting nodes that in a decentralized manner with a two-hop diameter. On the other hand, the CBRP protocol is not appropriate for the sensor networks that are energy-limited. LEACH is a hierarchical clustering algorithm presented by Heinzelman et al [39].

3 Proposed Methodology of WSN for Effective Scheduling and Routing

In wireless, routing protocol has become a major role in terms of power consumption, packet delivery ratio and transmission of packet and its packet scheduling, but in the development in wireless communication, there are new protocols such as collection tree protocol (CTP) for better improvement in latency and throughput and routing protocol for low lossy network and power (RPL^{2}NP) which is based on IPv6 IEEE standard discussed in [6–8]. These two advanced protocols are the replacement of routing centralization and scheduling algorithms in WSAN's in industry.

These protocols are combined together by considering the advantages of both and it is named as autonomous scheduling and distributed graph routing (ASDGR) which will take care of automatic routing and transmission of packet between source and destination in a distributed mode. The following are the main contribution of this research work.

1. Development of low power and low lossy networks routing protocol with help of RPL^{2}NP which will operate on graph routing and scheduling through minimum latency and high throughput.
2. Design of two scheduling approaches to compute automatic transmission based on the routing graph. The first approach is to find minimized distance between source and destination for end-to-end latency and eliminates the conflicts between other packet, and it is to minimize the traffic in real-world scenario.
3. The proposed RPL^{2}NP is an oriented distance-based routing protocol for development of low-power IPv6 network. The working principle of RPL^{2}NP is as follows.

The scheduling of packet and transmission is totally based on Internet access point (IAP) and each node has rank and its rank allocation is purely based on distance to the destination using cost function, i.e., excepted transmission count (ETC) and then packet is forward toward route to neighbor node. The generation of routes by RPL^{2}NP is not in graph route initially because every node has only one preferred "head" in the whole network to use many head's, suppose those head's equally preferred and have same or identical rank in the network and then there are choice to get collapse or interference occurring within the network, and it leads to lose of packet.

To overcome this issue, the modified RPL^{2}NP routing network assigns two preferred heads to each node at a time as default routes and forms the routing graph in the network as per the following specifications.

3.1 Directed Routing Graph (DRG)

It forms the routing among all nodes without coexists which are wanted to communicate with other for packet transmission. This routing, i.e., all selected links for routing orient toward the neighbor or terminates at the destination or access points for ensuring that data or messages should be delivered safely to the destination without any coexist in the graph.

3.2 Best Head and Second Best Head Selection

This selection alterative solution for avoiding of coexists or interfering of packet, in the network has best head and second head, the best head is to locate on the main path from node to access point with shortest distance from the destination node and second head has another best shortest distance path from the same destination node to serve as backup routing so that packet delivery ratio can be increased.

3.3 Allocation of Rank

Every node has a rank, and all access points allocate their own rank; initially, it is "1" and based on the best head's; initially, assigned rank is updated by increasing by "1".

3.4 Weighted ETC

The cost function of weighted ($\mathrm{ETC_w}$) is the node to measure the distance from the access point by using two routings based on Eq. (1)

$$\mathrm{ETC_w} = W_1^{*}\mathrm{ETC_{acc}} + W_2^{*}\mathrm{ETC_{accs}} \tag{1}$$

where $\mathrm{ETC_{acc}}$ is the distance accounted ETC from access point with the help of best head. $\mathrm{ETC_{accs}}$ is the distance accounted ETC from access point with the help of second head. $\mathrm{W_1}$ and $\mathrm{W_2}$ are two best heads weighted, and these are given in Eqs. (2) and (3).

$$W_1 = 1 - \left(1 - \frac{1}{\mathrm{ETC_{bh}}}\right)^2 \tag{2}$$

$$W_2 = 1 - \left(\frac{1}{\mathrm{ETC_{bh}}}\right)^2 \tag{3}$$

where $\mathrm{ETC_{bh}}$ represents ETC between the source node and its best head. As per standard wireless communication discussed in [9], the transmission of first packet through best head and the retransmission of second packet through second best head which is backup route. So $\mathrm{W_1}$ is probability of successfully delivered packet at time of first two transmission attempts, and $\mathrm{W_2}$ is probability of unsuccessful attempts fail. All the nodes present in the network are broadcast their own ranks periodically to join into transmission mode, and based on it, the ranks are allocated. After allocation of ranks as best head and second head, "joined callback message" sends to the selected best head and second best head, and also, it informs bout the selection to all other nodes.

Algorithm 1 Autonomous Scheduling and Distributed Graph Routing (ASDGR)

```
                Input             : $R_{id}$, $N_{id}$
                Output            : Updated router table ($R_{table}$)
                Initialization    : $R_{table}$=NULL,
                $ETC_w(N_{id}) = Rank(N_{id}) = \infty$
Condition 1:    if $N_{id}$= $R_{id}$        then
                /////Initialize the access point
                        Compute Rank=1 and $ETC_w$=0;
                        Broadcast about join-in messages;
                end
Condition 2:    if Rank ($N_{id}$) =$\infty$ & ~ $N_{id}$= $R_{id}$   then
                /////Allow receiver to receive the first join-in message from i
                        Compute $ETC_{acc}$ ($N_{id,}$ i) =ETC ($N_{id,}$ i) +$ETC_w$ (i);
                        Compute sender message as its best head;
                        Compute $ETC_{min}$ =$ETC_{acc}$ ($N_{id,}$ i);
                        Compute Rank ($N_{id}$) =Rank (i) +1;
                        Transmit joined message callback;
                end
Condition 3:    if Rank (~$N_{id}$) =$\infty$ & ~ $N_{id}$ = $R_{id}$   then
        /////Receiver, receivers the non-first message join-in from I
                        Compute $ETC_{acc}$ ($N_{id,}$ i) =ETC ($N_{id,}$ i) +$ETC_w$ (i);
                If   $ETC_{acc}$ ($N_{id}$, i) <$ETC_{min}$ then
                        Compute it as best head as the second best head;
                        Compute sender message as best head;
                        Compute $ETC_{min}$ =$ETC_{acc}$ ($N_{id,}$i);
                        Compute Rank ($N_{id}$) –Rank (i)+1;
                        Transmit as joined callback message;
                end
Condition 4:    if $ETC_{acc}$ ($N_{id,}$ $S_{bh}$)>$ETC_{acc}$ ($N_{id,}$ i)>=$ETC_{min}$ and Rank
(i)<Rank($N_{id}$) then
                        Compute sender, sends message as second best head;
                        Transmit joined message as join back;
                end
                        $ETC_w$ ($N_{id}$)=$W_1$*$ETC_{acc}$($N_{id,}$ $S_{bh}$)+$W_2$*$ETC_{acc}$($N_{id,}$ $S_{bh}$);
                        Broadcast message as join-in;
                end
Condition 5:    if Receive joined message callback then
        Update the router table $R_{table}$ and sender message is added as sub router;
                end.
```

The proposed distributed graph is shown in algorithm 1, and it starts with access point to form the routing graph; it routes toward the access point. Before network starts, the graph initializes the rank to 1 and ETC_w to 0; therefore, the network starts broadcasting join-in message. The remaining nodes compute their rank and ETC_w to infinity. Suppose any node receives the join-in messages from any another nodes, it opts its best head and second head purely based on accounted ETC routing table values and then it computes its rank by raising its best heads rank by 1. After updating of heads ranks, the node starts broadcasting join-in message to other nodes.

The ETC initializations between any two nodes are decided purely based on strength of received signals (SRS). In this work, we have set $\text{SRS}_{\text{min}} = -75\text{dBm}$ and $\text{SRS}_{\text{max}} = -90\text{dBm}$, when SRS value is more than -90dBm, then ETC is set to 1 otherwise ETC is set to 5, and the ETC can be scaled randomly between 1 and 5. In case, there is transmission error occurs between range 1 and range 5, and it can be measured by using Eq. (4).

$$\text{ETC} = \text{ETC}_{\text{old}} * \beta + q * (1 - x) \tag{4}$$

where ETC_{old} is the ETC value applied before maximum error occurs, q is error coefficients, and β is weight factor between 0 and 1.

Figure 2 shows the data paths examples for packet transmissions that have three access point (AP) and six field nodes. The dash lines show the ETC values with links. Whenever network starts the packet transmission, the three $\text{AP}_{1,}$ $\text{AP}_{2,}$ and AP_3 are starting broadcasting their ETC_{w} values and ranks to neighbor nodes. #3 selects AP_3 as its best head and AP_1 as its second best head, the selection of head's is based on ETC_{acc} values because $\text{ETC}_{\text{acc}}(3, \text{AP}_3)$ is greater than $\text{ETC}_{\text{acc}}(3, \text{AP}_1)$. Similarly, #4 selects AP_2 as its best head and AP_3 as second best head since ETC_{acc} (4, AP_2) is greater than ETC_{acc} (4, AP_3). Therefore, the rank of both #3 and #4 is 2, and these are starts broadcasting their ranks as join-in message to neighbors. To avoid loops, the #3 and #4 are not selected the link between #4 and #5. Based on connectivity among selected heads and neighbor nodes, the routing graph is generated and it is shown in Fig. 1. The solid lines are denoted the major path (primary), i.e., #8 → #6 → #3 → #AP_1, and the dash lines denote the backup routes (#8 → #7, #7 → #4, #4 → #3, #3 → $AP_1$and#5 → AP_3) (Fig. 3).

Slots allocation for application is shown in Fig. 1.

There is choice of attempting multiple transmissions through scheduling for each and every packet with the help of major path and backup path. Therefore, the transmission and reception of packet or schedule are purely depending on their unique id which is assigned to each and every node. This entire id's are generated as integer byte and stored in LUT and mapped as MAC address. The allocation of slots (s) is given in Eq. (5).

$$s = N * (\text{Node}_{\text{id}} - M_{AP}) - N + \beta \tag{5}$$

where N is no of attempts for transmission for each packet, M_{AP} is no of access points, Node_{id} is neighbor node id from routing table, and β is sth slot in the application for βth transmission attempts. To increase throughput of the ASDGR in WSN, the hybrid WSN which includes Dijkstra's algorithm, minimum tree spanning, and localized minimum spanning are incorporated for finding of the shortest path between source and destination. In these hybrid algorithms, mainly depending on message and node id's which are based on locally best decision of each and every node and it has its own information and this will be shared with neighborhood to find the shortest path in the graph in terms of best and second head's. By reducing of distance among the source and destination nodes, the throughput drastically increased and minimized

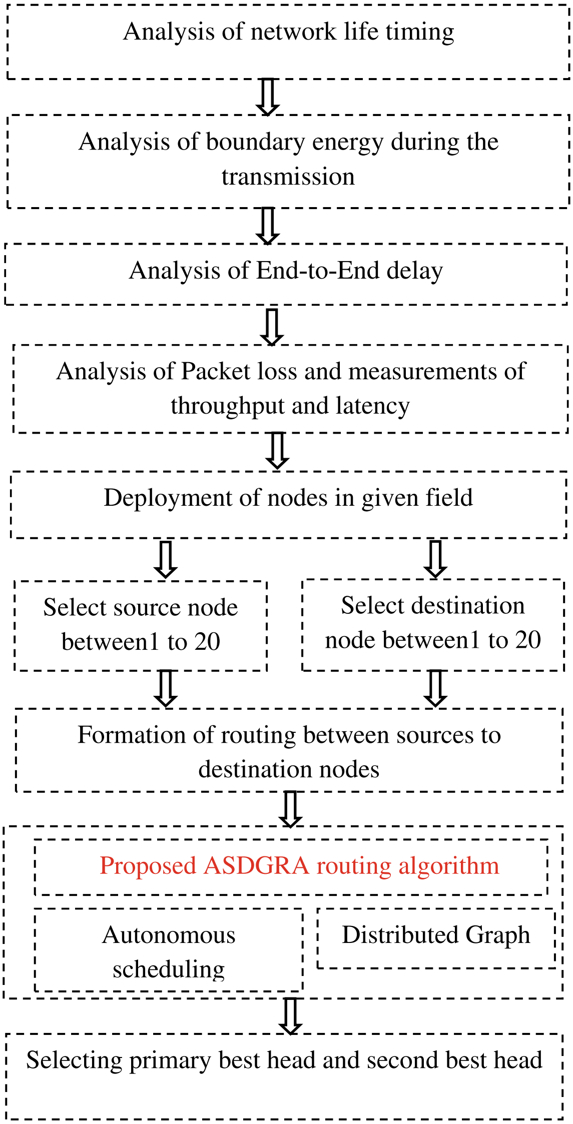

Analysis of network life timing
Analysis of boundary energy during the transmission
Analysis of End-to-End delay
Analysis of Packet loss and measurements of throughput and latency
Deployment of nodes in given field
Select source node between1 to 20
Select destination node between1 to 20
Formation of routing between sources to destination nodes
Proposed ASDGRA routing algorithm
Autonomous scheduling
Distributed Graph
Selecting primary best head and second best head

Fig. 1 Working flow diagram of proposed scheduling and dynamic routing graph using ASDGRA for high throughput and low latency

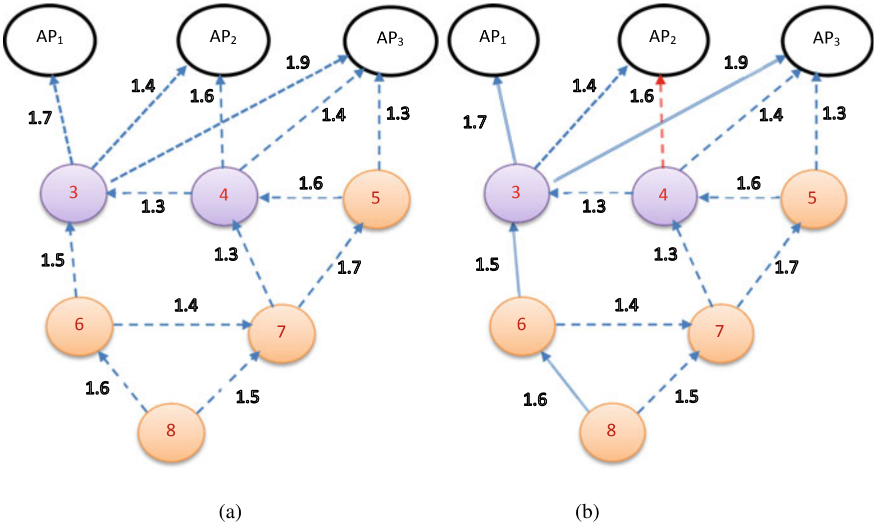


Fig. 2 Proposed created network topology for three access points and six field nodes and their transmission paths and directions. **a** Created network topology **b** Routing graph for best and second head's

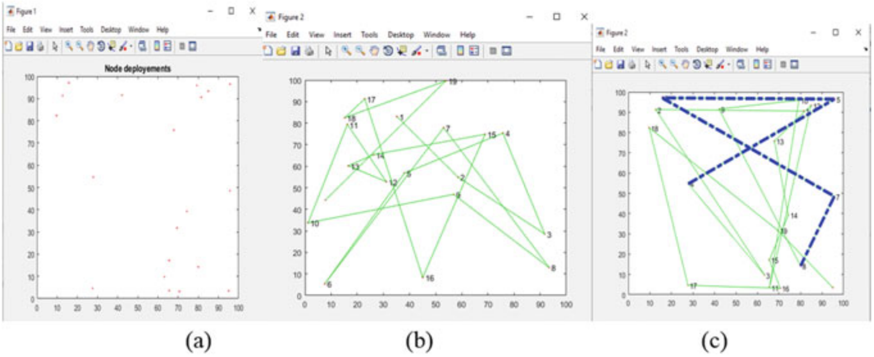


Figure 1
File Edit View Insert Tools Desktop Window Help
Node deployements
Figure 2
File Edit View Insert Tools Desktop Window Help
Figure 2
File Edit View Insert Tools Desktop Window Help
(a)
(b)
(c)

Fig. 3 Proposed wireless sensor network, nodes deployment and path establishment between source node and destination node. **a** 20 nodes deployment **b** Connection establishment among all nodes **c** Path establishment between source and destination nodes

the number dead nodes as shown in Fig. 5. With the help of Dijkstra's algorithm, the problem of maximum distance between source and destination is minimized and step-by-step process is shown in algorithm 2.

Algorithm 2 Shortest path identification between source and destination nodes

```
Step 1: Parameters initialization
Iteration number=0, distance=any number in positive infinity and data set[i]=0, where i=0,1,2,3……..n-m).
Notations: $r_n$=radius of transmission
              sv= starting of the node
              $n_r$=node relay
              $r_n$=Receiver node
Step 2: When cos(e)[sv] [i[<r then data set[i]=1,(i=0,1,2,3……m)
          Compute cos[i][j] for distance between source node (i) and destination node (j)
Step 3: If data set [$r_n$]=1 then
          Compute distance =cos(e)[sv][ $\mathbf{r_n}$]
                              else
                                        go to step 4
Step 4: All values of i are belongs to {data set [i]=1}, $n_r$ [i]=1, data set[j]=1 when j also belongs to {cos(e)[ $n_r$[j]<k} for all values of j=0,1,2,3………..n)
                              else
Record the distance that falls in data set [$r_n$]=1 and compute the distance between nodes is cos(e)[sv] $n_r$[0]<$\sum_{j=0}^{i-1} \cos(e)\ [sv]$ [nr[(j + 1) cos(e) [nr (i(sv))]
end.
```

3.5 Measurement of Energy in WSN

After the establishment of paths among the heads and other field nodes, the energy is measured per packet transmission as per the following specifications and plotted the obtained energy for without ASDGR and with ASDGR.

Sinks: sink. $x = 1.5$*WIDTH and sink. $y = 0.5$*HEIGHT.

Number of Nodes in base station area is $n = 20$.

Probability of a node is $p = 0.2$

Battery capacity is $E_o = 0.1$, ETX = 50*0.000000001, ERX = 50*0.000000001.

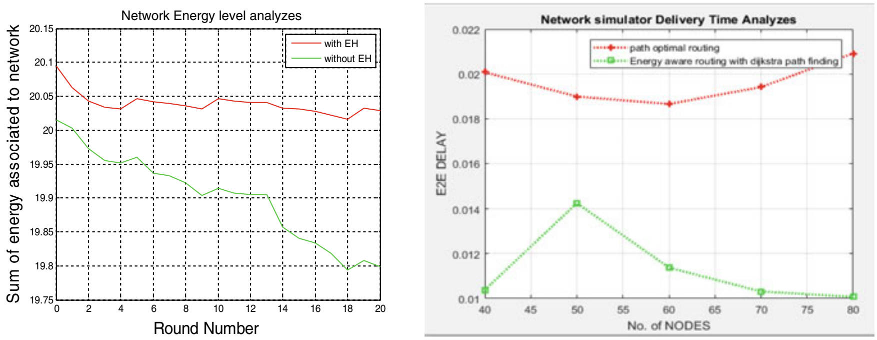
Transmission energy E_{fs}= 10*0.000000000001 and E_{mp} = 0.0013*0.000000000001.

Data Propagation Energy is EDA = 5*0.000000001.

Threshold for transmitting data to SINK are $h = 100$ and $s = 2$.

4 Results and Discussions

Figure 5 shows the performance analysis between different parameters and their optimization. The ASDGR is able to minimized the end-to-end delay during the transmission of packets from source node to destination node; Figure 5a shows the optimization of delay between proposed ASDGR algorithm and existing algorithm, the red color shows the end-to-end delay for existing algorithm, and yellow color shows the proposed algorithm delay; it is concluded that the delay of proposed communication for packet transmission is optimized 14% as compared to existing work. Figure 6b shows the number of packets lost during the packets transmission and due to effective scheduling and formation routing table, the losses are able to minimize and compared to existing results as shown in [12] (Fig. 4).



(a)

(b)

Fig. 4 Energy consumption per round per packet transmission between with ASDGRA-EH and without ASDGRA-EH

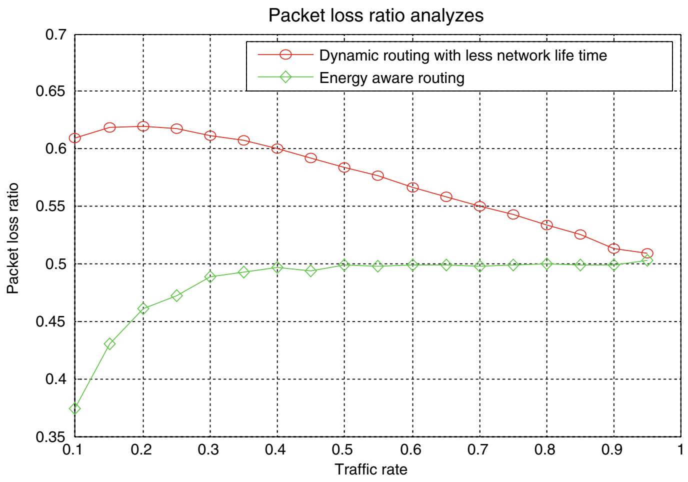


Fig. 5 Performance analysis of end-to-end delay, packet loss ratio with and without of dynamic routing **a** end-to-end delay **b** Without dynamic routing and with dynamic routing

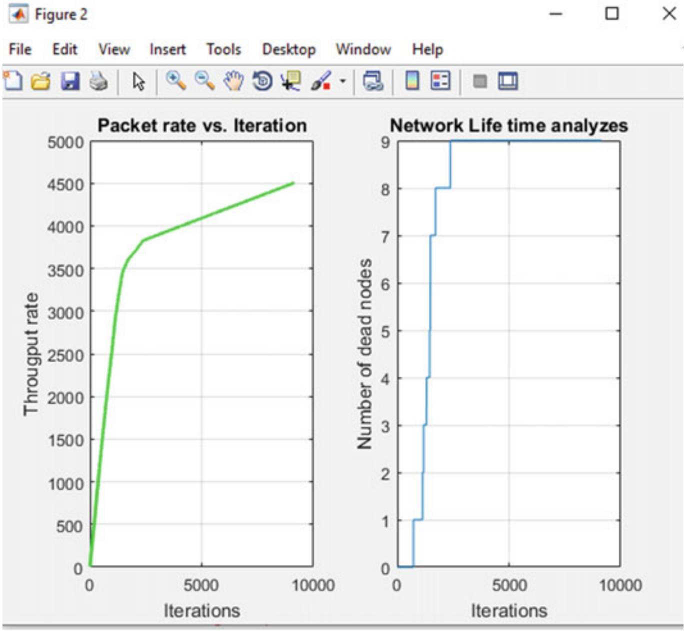


Fig. 6 Performance analysis of throughput and number of dead nodes with respect to lifetime of network of ASDGR during transmission of packets

5 Conclusion

The contribution of this paper is on improvising current WSAN and WSN networks for wireless communications and increasing their scalability through effective routing and scheduling to enhance visibility and predictability of wireless network operation in WSN. This paper decentralizes the organization of the board in wireless ASDGR and presents the primary circulated diagram steering and self-governing booking arrangement that permits the field gadgets to process their own chart courses and transmission plans. The figures show the synopsis of contrasts among existing directing and booking calculations like DiGS/DiGS-CD contrasted with proposed wireless ASDGR. Test results from two physical test beds and a huge scope recreation show our answer that gives a huge enhancement for network dependability, dormancy, energy proficiency, and disappointment resilience under elements, basic properties for modern applications, over cutting edge at the expense of somewhat higher force utilization and longer organization in statement. In this paper, we have likewise explored the issues of energy proficiency and QoS in a consolidated way for heterogeneous WSN networks within the sight of three energy the executive's methods: to address the issues of varieties of traffic burden, and energy accessibility conditions. We have then built up an "energy ease off" instrument, to be coordinated into WSN sensors for ASDGRA. The ASDGRA calculation can be executed in any

IEEE 02.15.4 standard-based WSN applications with the least alterations. Reproduction results have exhibited that our proposed calculation essentially improves energy effectiveness while fulfilling the QoS prerequisites. The outcomes likewise show that the organization of the crossbreed setup with different fuel sources is an effective, compelling, and pragmatic answer for at the same time improves the energy-proficient and QoS issues just as to expand the lifetime of sensors in heterogeneous WSN networks.

Acknowledgements I would like to express my gratitude to my guide Dr. Mohammed Abdul Waheed for his guidance during the preparation of this article. I extend my deep sense of gratitude to Visvesvaraya Technological University, Belagavi, Karnataka, India.

References

1. A. Manjeshwar et al., APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. (IEEE, 2002), pp. 1530–2075/02
2. A. Manjeshwar, D.P. Agrawal, TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. in *1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing,* April (2001). https://doi.org/10.1109/IPDPS.2001.925197
3. W. Heinzelman, A. Chandrakasan, H. Balakrishnan, UAMPS ns code extensions. http://wwwmtl.mit.edu/research/icsystems/uamps/leach
4. W.B. Heinzelman, Application-specific protocol architectures for wireless networks. PhD thesis, Massachusetts Institute of Technology June (2000)
5. N.A. Pantazis, A. Spiridonos, S.A. Nikolidakis, D.D. Vergados, Energy-proficient directing conventions in remote sensor networks: survey. IEEE Commun. Surv. Tut. **15**(2), 551–591. https://doi.org/10.32628/CSEIT20635
6. D. Zordan, B. Martinez, I. Villajosana, M. Rossi, On the exhibition of lossy pressure plans for energy compelled sensor organizing. ACM Trans. Sensor Newt. **11**(1), 1–34 (2014)
7. N. Cao, E. Masazade, P.K. Varshney, A multi objective improvement based sensor choice strategy for target following in remote sensor organizations. in *Proceedings Sixteenth International Conference Information Combination* (2013), pp. 974–980
8. M. Centenaro, M. Rossi, M. Zorzi, Joint improvement of lossy pressure and transport in remote sensor organizations. in *Proceedings IEEE Globecom Workshops*, Washington, DC, USA, Dec. (2016), pp. 1–6
9. M. Rossi et al., Distributed learning algorithms for optimal data routing in IoT networks. IEEE Trans. Signal and Inform. Processing Over Netw. **6**, 2373–776X (2020). https://doi.org/10.1109/TSIPN.2020.2975369
10. P. Giselsson, S. Boyd, Direct intermingling and metric determination for Douglas–Rachford parting and ADMM. IEEE Trans. Autom. Control **62**(2), 532–544 (2017)
11. A. Biason, C. Pielli, A. Zanella, M. Zorzi, Access control for IoT hubs with energy and constancy requirements. IEEE Trans. Remote Commun. **17**(5), 3242–3257 (2018)
12. F. Iutzeler, P. Bianchi, P. Ciblat, W. Hachem, Unequivocal union pace of a conveyed rotating course technique for multipliers. IEEE Trans. Autom. Control **61**(4), 892–904 (2016)
13. S. Javaid, H. Fahim, Z. Hamid, F.B. Hussain, Traffic-mindful blockage control (TACC) for wireless multimedia sensor organizations. Multimedia Tools Appl. **77**(4), 4433–4452 (2018)
14. S. Guo et al., Confided in cloud-edge network resource management: DRL-driven service function chain orchestration for IoT. IEEE Internet Things J. (2019). https://doi.org/10.1109/JIOT.2019.2951593,2327-4662(c)IEEE

15. H. Hawilo, M. Jammal, A. Shami, Organization work virtualization aware orchestrator for administration work affixing position in the cloud. IEEE J. Selected Areas Commun. **37**(3), 643–655 (2019)
16. M.M. Tajiki, S. Salsano, L. Chiaraviglio, M. Shojafar, B. Akbari, Joint energy productive and QoS-mindful way allotment and vnf position for administration work tying. IEEE Trans. Netw. Serv. Managem. **16**(1), 374–388 (2019)
17. J. Shi et al., Distributed graph routing and scheduling for industrial wireless sensor-actuator networks. IEEE/ACM Trans. Netw. **27**(4), 1063–6692 (2019)
18. X. Chen, Z. Zhu, J. Guo, S. Kang, R. Proietti, A. Castro, S.J.B. Yoo, Utilizing blended system gaming to acknowledge motivation driven vnf administration chain provisioning in representative based flexible optical inter datacenter networks. IEEE/OSA J. Opt. Commun. Netw. **10**(2), A232–A240 (2018)
19. C. Mouradian, S. Kianpisheh, M. Abu-Lebdeh, F. Ebrahimnezhad, N. T. Jahromi, R.H. Glitho, Application segment arrangement in nfvbased cross breed cloud/mist frameworks with portable haze hubs. IEEE J. Selected Areas in Commun. **37**(5), 1130–1143 (2019)
20. N.M. Maroof, M.A. Waheed, An efficient information retrieval and routing using AODV on TDMA in WSNs. Int. J. Adv. Sci. Technol. **28**(7), 107–117 (2019)
21. L. Gu, D. Zeng, S. Tao, S. Guo, H. Jin, A.Y. Zomaya, W. Zhuang, Decency mindful powerful rate control and stream planning for network utility boost in network administration chain. IEEE J. Selected Areas in Commun. **37**(5), 1059–1071 (2019)
22. T.D. Nguyen et.al, A distributed energy-harvesting-aware routing algorithm for heterogeneous IoT networks. IEEE Trans. GREEN Commun. Netw. IEEE 2473–2400 (2018). https://doi.org/10.1109/TGCN.2018.2839593
23. P.K. Sharma, M. Chen, J.H. Park, A product characterized haze hub based disseminated block chain cloud engineering for IoT. IEEE Access **6**, 115–124 (2018)
24. F. Ferrari, M. Zimmerling, L. Mottola, L. Thiele, Low-power remote transport. in *Procedings SenSys* (2012), pp. 1–14
25. F. Ferrari, M. Zimmerling, L. Thiele, O. Saukh, Productive organization flooding and time synchronization with polished. in *Proceedings SenSystems*, April (2011), pp. 73–84
26. N.M. Maroof, M.A. Waheed, A. Heena, Mechanism for congestion detection and handling in wireless sensor networks—a survey. IJIRCCE **5**(4), 260–265 (2017). ISSN (Online): 2320–9801ISSN (Print): 2320–9798M
27. M. Doddavenkatappa, M.C. Chan, B. Leong,Sprinkle: Fast information dispersal with useful impedance in remote sensor organizations. in *Proceedings NSDI* (2013), pp. 269–282
28. M. Doddavenkatappa, M.C. Chan, P3: A functional parcel pipeline utilizing simultaneous transmissions for remote sensor organizations. in *Proceedings IPSN*, (2014), pp. 203–214
29. N. Michelusi, M. Zorzi, Optimal adaptive random multi-access in energy harvesting wireless sensor networks. IEEE Trans. Commun. **63**(4), 1355–1372 (2015)
30. T.D. Nguyen, J.Y. Khan, D.T. Ngo, An adaptive MAC protocol for RF energy harvesting wireless sensor networks. in *Proceedings of GLOBECOM'16* (2016)
31. D. Hasenfratz, A. Meier, C. Moser, J.J. Chen, L. Thiele, Examination, correlation, and enhancement of directing conventions for energy reaping remote sensor organizations.in *Proceedings of SUTC'10*, Jun (2010), pp. 19–26
32. G. Martinez, S. Li, C. Zhou, Wastage-mindful steering in energy harvesting remote sensor networks. IEEE Sens. J. **14**(9), 2967–2974 (2014)
33. P. Gong, Q. Xu, T.M. Chen, Energy collecting mindful directing convention for remote sensor organizations. in *Proceedings of CSNDSP'14*, Jul (2014), pp. 171–176
34. J. Bai, M. Fan, J. Yang, Y. Sun, C. Phillips, Keen energy collecting steering convention for WSN based E-wellbeing frameworks. in *Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare* (Mobile Health '15) (2015), pp. 23–28
35. Q. Tan, W. An, Y. Han, Y. Liu, S. Ci, F.- M. Shao, H. Tang, Energy collecting mindful geography control with power transformation in remote sensor organizations. Ad Hoc Netw. **27**, 44–56 (2015)

36. D. Estrin, R. Govindan, J. Heidemann, S. Kumar, Next century challenges: scalable coordination in wireless networks. in *Proceedings of the fifth Annual ACM/IEEE International Gathering on Mobile Computing and Networking* (MOBICOM) (1999), pp 263–270
37. T. He, K.W. Jawline, S. Soh, On remote force move and max stream in battery-powered remote sensor organizations. **4**, 4155–4167 (2016)
38. W. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocols for wireless micro sensor networks. in *Proceedings of Hawaiian International Gathering on Systems Science*, January (2000)
39. N. Michelusi, M. Zorzi, Optimal adaptive random multi-access in energy harvesting wireless sensor networks. IEEE Trans. Commun. **63**(4), 1355–1372l (2015)

Janus Antenna Array Design for Doppler Navigation System

J. Aswin, Radhika Priyavardhini, V. Hrishitha, Konda Harsha, and M. Jayakumar

Abstract This paper has successfully developed the JANUS antenna configuration in navigation systems for real-time applications such as airborne vehicles and ships. Doppler shift in the frequency of the received signal and differential magnitude provides valuable data for determining the status of the vehicle in air as well as for landing. The advancement of computational techniques, high speed processors and availability of various beamforming configurations make the JANUS doppler navigation system highly accurate. The main focus of this paper is to design a 16-element annular ring microstrip patch antenna array for its application in doppler-based aeronautical radio navigation system [ARNS] in the frequency band of 13.25-13.4 GHz. The desired operating frequency of the antenna is 13.3 Ghz, where the substrate used for the patch is Rogers RT Duroid 5880 with a relative permittivity of 2.2 and a dissipation factor, tanδ of 0.0004. The antenna was designed and simulated on high frequency structure simulator [HFSS], and parametric sweeps were applied to find optimum results for single element and for each of the two, four and sixteen element array systems, respectively.

Keywords Microstrip patch antenna · Annular ring · Antenna array · Aeronautical radionavigation system · Janus

1 Introduction

In the modern world, airborne vehicles such as helicopters, aeroplanes, unmanned aerial vehicles and drones with state-of-the-art avionics on board to make it easier and safer to control and operate them. In the selected military and exploratory applications that use unmanned aerial vehicles/systems (UAS), the airborne entity is controlled via a human operator on ground based on the visuals transmitted from cameras on

J. Aswin (✉) · R. Priyavardhini · V. Hrishitha · K. Harsha · M. Jayakumar
Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: m_jayakumar@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_18

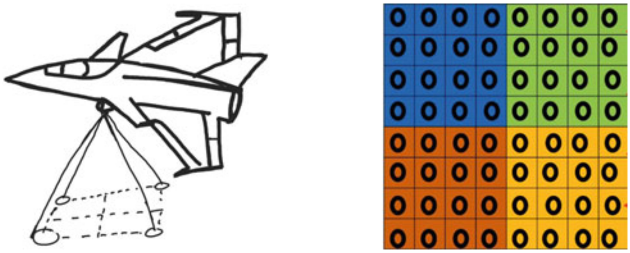
board the vehicle. Night flying or low level flying in hilly regions, in bad weather conditions or navigation in regions of low visibility due to weather, clouds or a foggy environment, however, are prone to human errors and can be dangerous for the vehicle. An emergency situation in the given conditions such as landing or navigation through uneven terrain can have a high probability of the vehicle crashing.

Various research has gone into developing guidance systems that can help navigate through unknown terrain with poor visibility. Antennas have been designed and applied in both, air-borne vehicle communication system to ensure safety of flights as mentioned in [1]. TFR (terrain-following radar) is a military aerospace technology that enables a low-flying aircraft to maintain a relatively steady altitude above ground level by examining the terrain ahead of the aircraft and providing guidance to the pilot and/or the aircraft flight control systems [2].

Current research into terrain-following algorithms such as the ones presented in [3–5] is disadvantageous as these algorithms require an elevation map of the area over which the flight is to navigate, based on which it calculates the most efficient routes in advance that the UAS will follow and, hence, lack dynamicity of operation. T. Templeton et al. [6–9] have also proposed vision-based terrain following systems incorporating various available methods and algorithms including efficient terrain mapping, 3d reconstruction of terrain from visuals, motion detectors, etc. These systems though efficient in stable environments with clear visual range cannot be utilised in the aforementioned scenario. Moreover, vision-based systems require the hovering vehicle to be very stable to form error free and clear images. Large airborne vehicles such as military drones or manned aerial vehicles that require assistance in navigation, however, may not be stable due to various factors such atmospheric drag, engine vibrations and air turbulence experienced during flight.

A radar-based navigation system that may be analysed dynamically to give directions can be useful in such circumstances as the constraint of low visibility does not affect the performance of a radar system. Projects have been undertaken such as the one mentioned in [10], aimed at designing, manufacture and testing of a synthetic aperture radar-based terrain observation system suitable for installation aboard a miniature UAV. Alternatively, an airborne doppler navigation system using the Janus configuration can also be beneficial in classifying the terrain underneath the UAS. The Janus doppler radar navigation system uses four antenna beams as shown in Fig. 1a; two fore and two aft, on the two sides of the ground track, to compute the aircraft velocity vector referenced to the terrain by measuring the doppler shift of the ground echo from the beams. The beams may transmit in pairs or sequentially, depending on the system design. The Janus configuration can be achieved by a system of closely packed phased array antennas that produces a pencil shaped beam which can be rotated in any desired direction by using certain beamforming algorithms or beamformer chips. These navigation systems can be installed in aircraft (helicopters, as well as certain aeroplanes) and are used for specialised applications such as continuous determination of ground speed and drift angle information of an aircraft with respect to the ground [11].

Saltzman and Stavis [12] had designed a dual beam planar antenna array for Janus doppler navigation system and have discussed about the depth/height of the



(a) Janus Configuration of Antenna Beam

(b) Janus Antenna Array

Fig. 1 Airborne Janus Doppler navigation system

antenna being a major factor for aircraft installations. Taking this into consideration, we propose designing the Janus configuration using microstrip patch antennas given their compact nature and ease of fabrication.

2 Annular Ring Configuration

Microstrip patch antennas are popular for primary spaceborne applications. These antennas consist of a metallic patch on a grounded substrate. The metallic patch can take various configurations. They are low profile, conformable to planar and non-planar surfaces, simple and inexpensive to fabricate using modern printed circuit technology, mechanically robust when mounted on a rigid surface and are very versatile in terms of resonant frequency, polarization and impedance [13]. Satellite launch vehicles, missiles and aeroplanes demand antennas to be fully conformal in order to reduce the drag related effects with the surrounding atmosphere [14]. These characteristics make microstrip patch antenna ideal for our desired application.

Of the various configurations of microstrip antennas available, annular ring configuration is selected for designing the array. For a given frequency, the size is substantially smaller than that of the circular or rectangular patch when both are operated in the lowest mode. In application of arrays, this allows elements to be more densely placed, thereby reducing grating-lobe problem. In an annular ring microstrip patch antenna, the mean circumference of a ring is equal to the guided wavelength of the microstrip used. For any given frequency, the mode corresponding to $n = m = 1$ (TM_{11} mode) has the minimum mean radius of the ring and is hence known as dominant mode. Input impedance of the ring operated in TM_{11} mode is considerably higher, whereas the impedance bandwidth is smaller, in comparison with the other patch shapes, as well as the separation of resonant modes can be controlled by varying the ratio of outer to inner radii [15, 16].

The resonant frequency for the annular ring is determined by following equations from [15, 16]:

$$f_{\mathrm{nm}} = \frac{k_{\mathrm{nm}} c}{2\pi \sqrt{\varepsilon_r}} \tag{1}$$

$$J'_n(\mathrm{kb}) Y'_n(\mathrm{ka}) - J'_n(\mathrm{ka}) Y'_n(\mathrm{kb}) = 0 \tag{2}$$

where,
f_{nm} is the resonant frequency
ϵ_r is the relative permittivity
a and b are inner and outer radius, respectively
K_{nm}/k is the roots of the characteristic Eq. (2), $Jn(X)$ and $Yn(X)$ are Bessel functions of the first and second kind, order n, respectively, and the prime denotes first order derivative with respect to X.

Assuming $C = b/a$, Eq. (2) takes the form

$$J'_n(CX_{\mathrm{nm}}) Y'_n(X_{\mathrm{nm}}) - J'_n(X_{\mathrm{nm}}) Y'_n(CX_{\mathrm{nm}}) = 0 \tag{3}$$

where,

$$X_{\mathrm{nm}} = K_{\mathrm{nm}} a \tag{4}$$

Substituting the equivalent of K_{nm} in terms of X_{nm} in (1), we arrive at (5).

$$f_{\mathrm{nm}} = \frac{X_{\mathrm{nm}} c}{2\pi \sqrt{\varepsilon_r} a} \tag{5}$$

To account for the fact that a small fraction of the field exists outside the dielectric, it is customary to use an effective permittivity ϵ_e, where ϵ_e is given by the formula:

$$\varepsilon_e = \frac{1}{2}(\varepsilon_r + 1) + \frac{1}{2}(\varepsilon_r - 1)\left(1 + \frac{10t}{W}\right)^{\frac{-1}{2}} \tag{6}$$

where,
W is $b - a$
t is the thickness of the dielectric substrate.

The value of f_{nm} is taken as 13.3 Ghz based on recommendations from [11]. X_{nm} is found by solving (3). An initial a (inner radius) and b (outer radius) are found by solving (5). The radii values are then substituted in (6) to find ϵ_e. The ϵ_e value is then substituted in place of ϵ_r in (5) to find effective inner and outer radii.

The feeding technique is an important factor that determines how efficiently the voltage is transferred to the antenna which in turn affects the antenna's radiation characteristics [17]. Among the various available feeding techniques including the microstrip line feed, coaxial probe feed, aperture and proximity coupling methods, the microstrip feed method is chosen as it is a simple conducting strip whose dimensions can be varied. Further for design of the planar antenna array, corporate feed structure is utilised because of its simplicity in producing multiple levels of binary power distribution resulting in each element of the array receiving equal power from the source [18].

3 Antenna Design and Simulation

From [19], it was inferred that in a large array configuration, and the minimum distance between two adjacent elements should be at least 0.5 * λ_g to achieve high directivity. Solving in the dominant mode ($n = m = 1$) TM$_{11}$, (3) was solved by varying the ratio (C) between outer and inner radius from 1.1 to 3 in steps of 0.1 by programming them in python to calculate X_{nm}. X_{nm} values were then substituted in (5) with ϵ_r value as 2.2 to calculate the radius.

From the calculated radii values, it was observed that for the values of C above 1.9, the area of no two adjacent elements overlapped when placed linearly at a distance of 0.5* λ_g. Hence, the ratio of 1.9 between the outer and inner radius was chosen.

Owing to the small size of the antenna and the fact that the edge of the antenna was circular, it was necessary for a narrow feed to be connected to it. The width of feedlines for various impedances was calculated, and the feedline of 200 Ω characteristic impedance having a width 0.1615 mm was chosen. The antenna was connected to a quarter wave transformer via a 200 Ω line of length $\lambda_g/4$, and a parametric sweep was applied on the width of the quarter wave transformer.

As mentioned in [15], the measured values based on (5) can have a 3% deviation from the calculated resonant frequency. The calculated radii values for 13.3 Ghz were simulated on HFSS, and it was observed that majority of the results resonated at 13.4 Ghz. Hence, the values were calculated again for 13.2 Ghz as f_{nm}, and the sweep showed majority of the values resonating at 13.3 Ghz. Corresponding calculated values are given in Table 2. As shown in Fig. 2, the annular ring with dimensions as mentioned in Tables 1 and 3 was designed. The optimum result was found at the quarter wave width of 2.96 mm.

It is known that an aggregate of radiating elements (array) in an electrical and geometrical arrangement can produce the desired results as the radiation from all

Table 1 Antenna parameters

Parameter name	Values
Substrate material	Rogers RT duroid 5880
Substrate thickness	1.52 mm
Relative permittivity	2.2
Dissipation factor	0.0004
Operating frequency	13.3 Ghz
λ_g	15.207 mm

Table 2 Calculated values

f_{nm} (Ghz)	Ratio b/a	X_{nm}	a (mm)	a^{eff} (mm)	b (mm)	b^{eff} (mm)	Simulated resonance (Ghz)
13.2	1.9	0.6994	1.7056	1.8952	3.2406	3.6010	13.3
13.3	1.9	0.6994	1.6927	1.8813	3.2163	3.5745	13.4

Fig. 2 Single antenna element

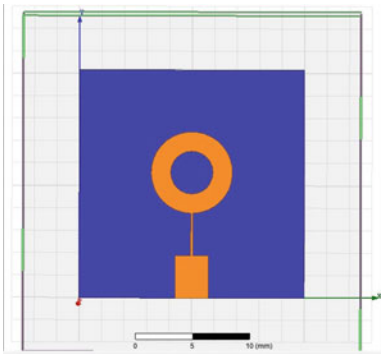


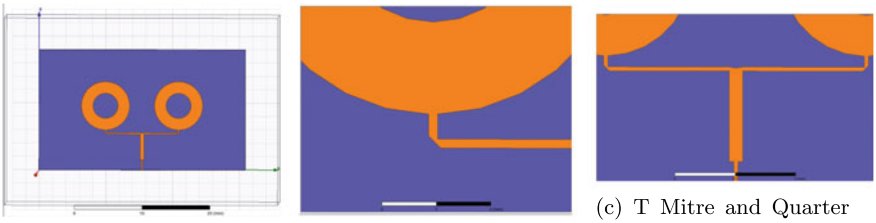
Table 3 Parameters of single element

Parameter name	Values (mm)
Inner radius a	1.8952
Outer radius b	3.601
Quarterwave transformer length	3.801
Quarterwave transformer width	2.96
Feedline length	3.801
Feedline width	0.1615

the elements interfere together to give a maximum radiation in a particular direction [13].

To proceed into developing an array system, a linear array of two elements as shown in Fig. 3 was designed. Unlike the single element design, to achieve symmetry in further design which will involve mirroring the linear array, the feedline was bent in *L* shape where it met the antenna, and the two elements receiving power through the same source was facilitated with the use of a *T* junction to divide power. A quarter wave transformer of 141.421 Ω was placed at the *T* junction to convert 100 Ω impedance at the junction back to 200 Ω.

From [17], it can be inferred that discontinuities in the feed lines can lead to induction of parasitic reactance and capacitance that affects the impedance matching of the feed which causes more signal to be reflected back to the source. To compensate for the discontinuities and to ensure reduction in the amount of signal that will be reflected back towards the source, the L shaped bends and the T junctions are mitred

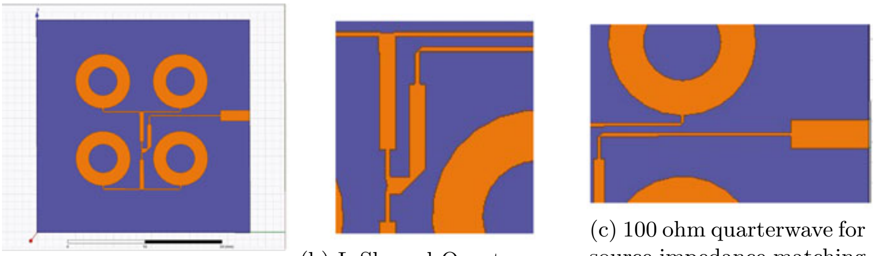


(a) Linear Array

(b) L Bend Mitring

(c) T Mitre and Quarter wave Transfomer

Fig. 3 2 element array design



(a) 2x2 Configuration

(b) L Shaped Quarterwave

(c) 100 ohm quarterwave for source impedance matching

Fig. 4 4 Element planar array

based on equations from [17, 20]. The structure was then simulated on HFSS, and a parametric sweep was applied for finding optimum performance by varying the vertical length of the L bend (represented by variable JAV in Table 5) and the inter-element spacing.

Verifying that the linear array could resonate at the desired operating frequency over a varied range of interelement spacing, a planar array of four elements was designed by joining a mirrored version of the linear array to itself. A quarter wave transformer in L shape at the centre of the 2 × 2 configuration was used to convert the impedance back to 200 Ω and a 100 Ω quarter wave transformer was used at the edge of the substrate to match the impedance to 50 Ω. The same is shown in Fig. 4.

An array of 16 elements was designed by using four subarrays of 2 × 2 configuration and joining them with quarter wave transformers. A quarter wave transformer from the centre of the 4 × 4 configuration was used to convert the impedance to 200 Ω, and a 100 Ω quarter wave transformer was used to match the structures impedance to the source impedance of 50 Ω as shown in Fig. 5.

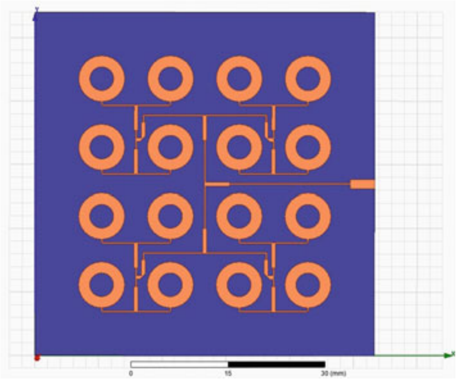


Fig. 5 16-Element planar array

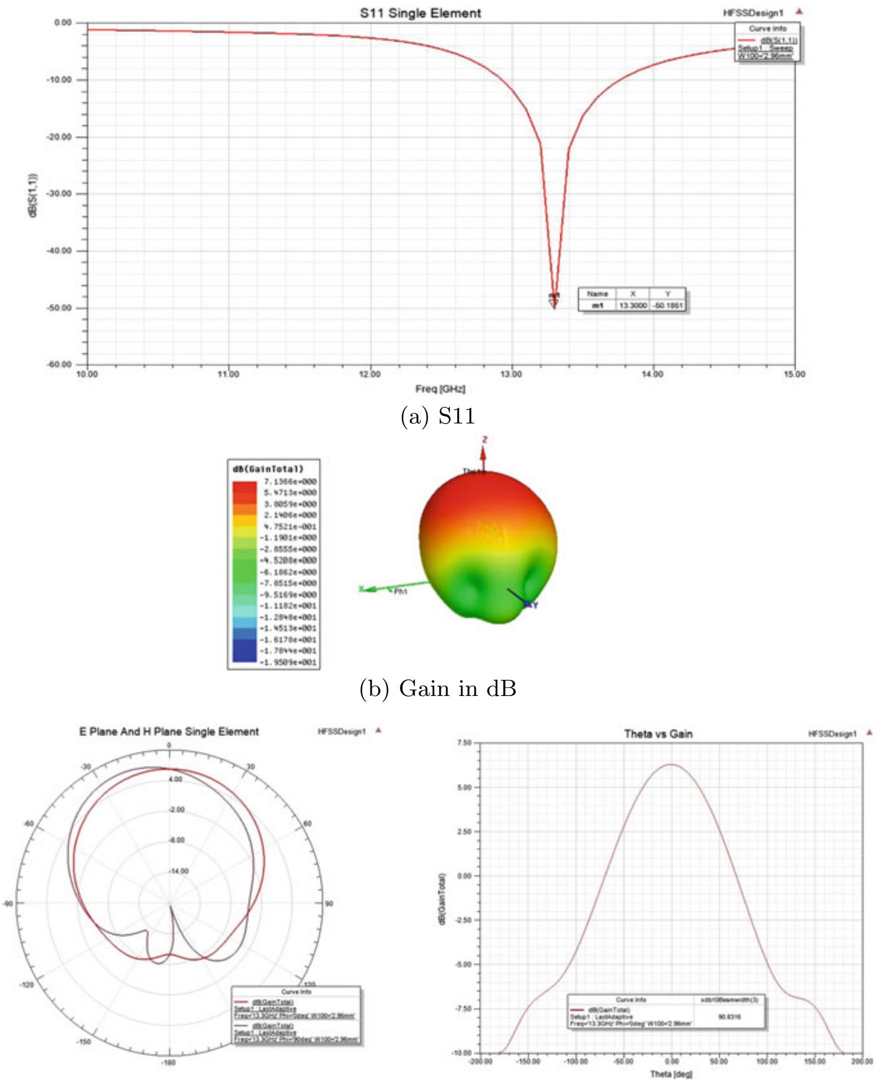
4 Results and Discussion

Highest gain value was found at an inter-element spacing of $0.71^*\lambda_g$ with a value of 15.63 dB. It can be assumed that the usage of four such elements in the Janus configuration as shown in Fig. 1b should give a gain of more than 20 dB as four times the power corresponds to a 6 dB increase in gain value.

Any return loss below −20 dB is an acceptable value for operation. The 16-element array giving −33.58 dB can be thus considered not just sufficient but a very good value for operation.

The HPBW value has reduced to 27.33 degrees which is narrow compared to the singe element that had an 90.83 deg HPBW (Figs. 6, 7, 8 and 9; Table 4, 5, 6 and 7). Proceeding into the Janus configuration, it can be expected that the HPBW value would further reduce as it would make use of four such sub-arrays controlled by a state-of-the-art beamformer chip such as the Adar1000 [21] for operation, thereby improving resolution of the navigation system drastically and the radiation pattern thus obtained with a high gain value, and low HPBW value would look like a pencil beam.

In planar antenna arrays, a serious problem of coupling between antenna elements occurs due to the presence of surface waves. If mutual coupling is strong, a large portion of the power fed into one antenna element will be coupled to the other neighbouring antenna elements rather than radiating to free space and, thus, reducing antenna gain, operational bandwidth, the radiation efficiency [22]. The system design can be enhanced for better performance by application of various solutions one of which has been discussed in [23]. The usage of signal processing techniques such as range migration correction or radar angle superresolution algorithm as mentioned in [24] can be used to improve the terrain following capability of the navigation system.



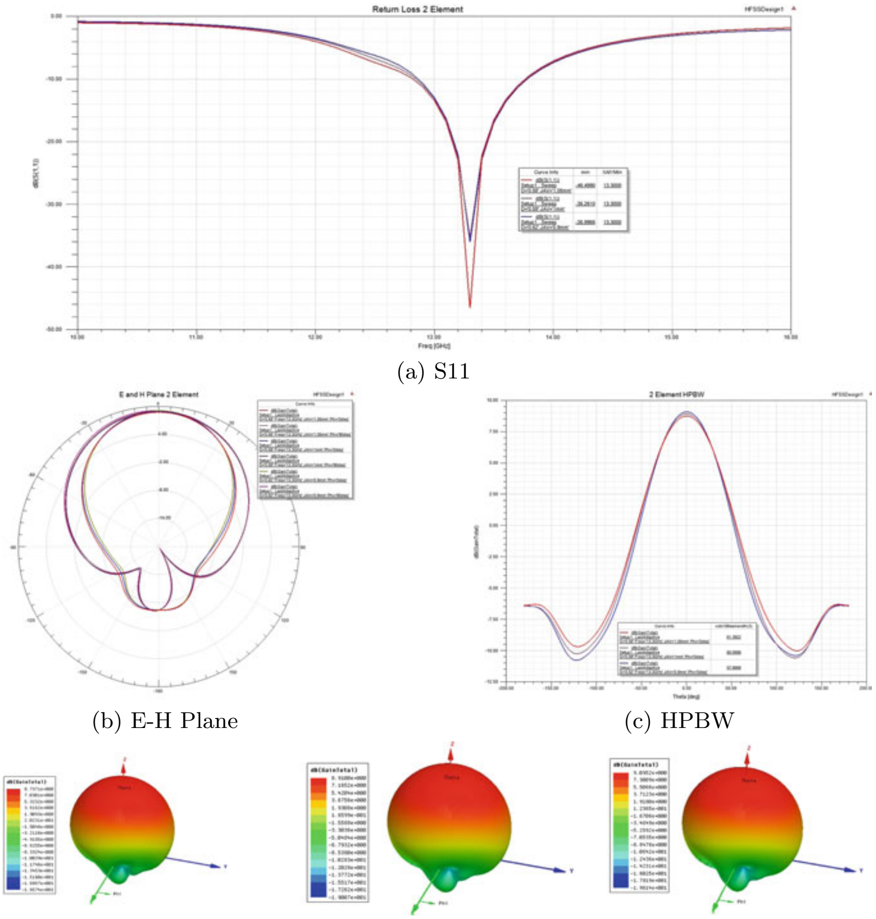
(a) S11

(b) Gain in dB

(c) E-H Plane

(d) HPBW

Fig. 6 S11, Gain, [E, H] Plane and half power beamwidth of single element



(a) S11

(b) E-H Plane

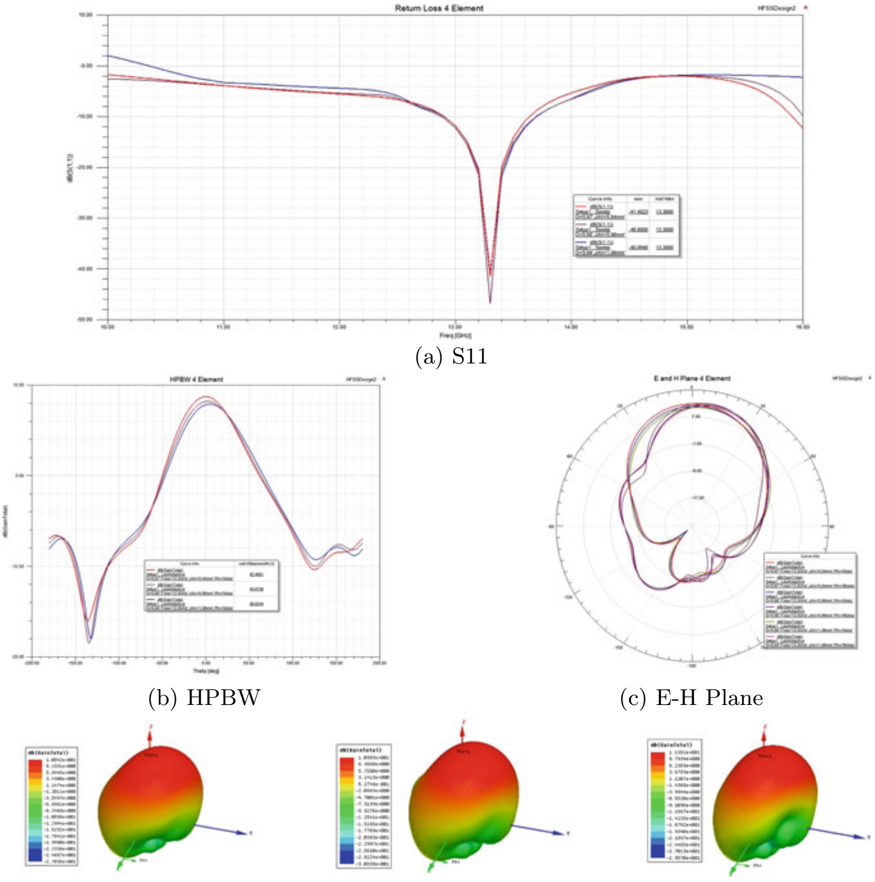
(c) HPBW

(d) 0.58 λ_g Gain

(e) 0.59 λ_g Gain

(f) 0.62 λ_g Gain

Fig. 7 S11, $[E, H]$ Plane, half power beamwidth and gain of two element array



(a) S11

(b) HPBW

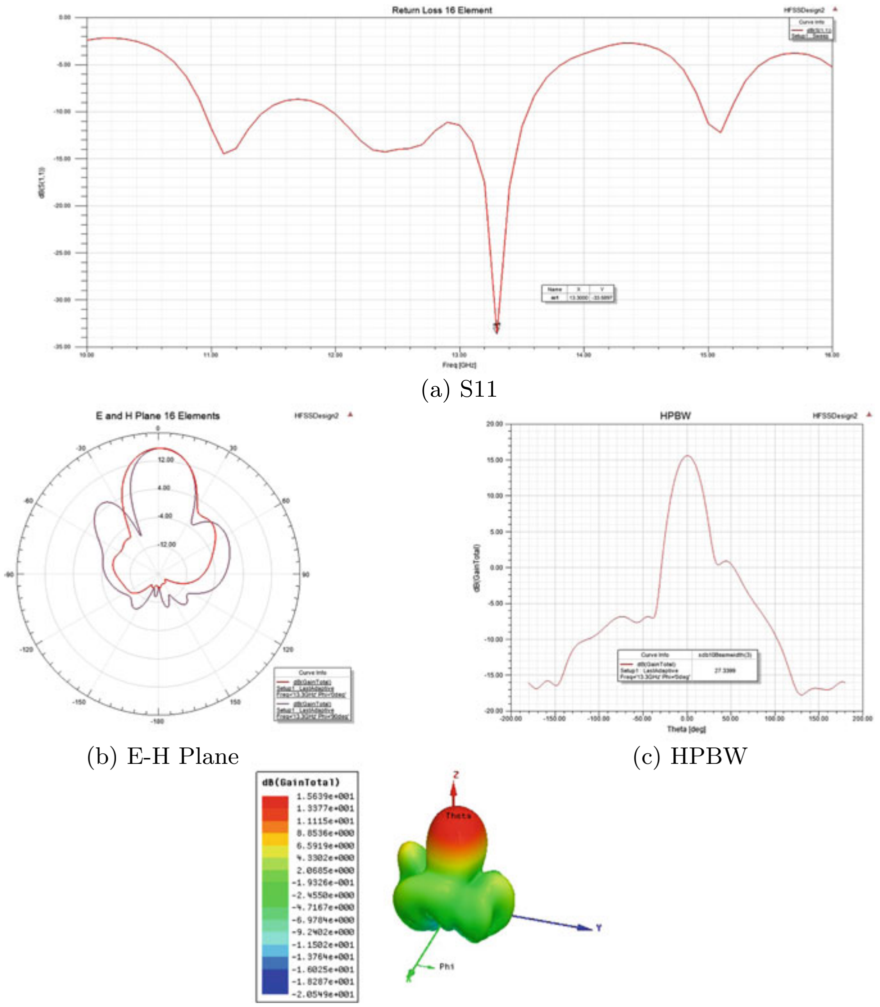
(c) E-H Plane

(d) 0.64 λ_g Gain

(e) 0.65 λ_g Gain

(f) 0.67 λ_g Gain

Fig. 8 S11, $[E, H]$ Plane, half power beamwidth and gain of four element array



(a) S11

(b) E-H Plane

(c) HPBW

(d) Gain in dB

Fig. 9 S11, $[E, H]$ plane, half power beamwidth and gain of sixteen element array

Table 4 Parameters of single element

Return loss (dB)	Gain (dB)	HPBW
−50.181	7.13	90.83

Table 5 Parameters of two element linear array

Interelement spacing	Return loss (dB)	Gain (dB)	HPBW	JAV (mm)
0.58 * λ_g	−46.49	8.737	61.39	1.05
0.59 * λ_g	−35.26	8.910	60.05	1
0.62 * λ_g	−35.99	9.095	57.66	0.9

Table 6 Parameters of four element planar array

Interelement spacing	Return loss (dB)	Gain (dB)	HPBW	JAV (mm)
0.64 * λ_g	-40.89	10.54	58.82	1.06
0.65 * λ_g	-46.85	10.98	54.67	0.98
0.67 *λ_g	-41.45	11.35	52.48	0.54

Table 7 Parameters of sixteen element planar array

Interelement spacing	Return loss (dB)	Gain (dB)	HPBW
0.71 * λ_g	−33.58	15.63	27.33

5 Conclusion

As proposed by this project, a 16-element annular ring microstrip patch antenna array for application in airborne vehicles and ships has been designed. The antenna was modelled and successfully simulated on high frequency structure simulator (HFSS). The proposed Janus configuration-based doppler navigation system would use four different 16-element arrays, that were designed as part of this project. The final configuration is shown in Fig. 1b where each quadrant is a 16-element array.

The antenna array system, thus, designed would have four different input sources. Varying the phase and power input to each source would allow the beam generated by them to steered into the required direction. This functionality can be easily achieved in today's world with the availability of high-speed processors and state-of-the-art beam former chips.

Thus, instead of using four high gain Janus arrays, a single phased array controlled by a beamformer chip can be used to steer the beam in the required direction. Fast switching by the beam former chip can be utilised to create four beams required in minimal time such that it appears that there are always four beams present. The current dimensions of the 16-element array are 53 mm × 53 mm × 1.52 mm. The Janus configuration would have a size of 106 mm × 106 mm × 1.52 mm. Thus, the proposed navigation system would not only be low profile, and cheap but also would require less place to setup aboard an UAV.

References

1. S. Anand Selvin, K. Bharathi Ravi, Darwino, G.S., R. Harishankar, R. Jitha, M. Jayakumar, Design and studies on non-planar conformal patch antennas for air-borne vehicles (2009)
2. P. Hai-yang, L. Shun-an, A special terrain-following system based on flash LIDAR, in *2012 IEEE International Conference on Mechatronics and Automation* (Chengdu, China, 2012), pp. 653–657. https://doi.org/10.1109/ICMA.2012.6283219
3. R. Samar, A. Rehman, Autonomous terrain-following for unmanned air vehicles. Mechatronics **21**(5), 844–860 (2011)
4. A. Kosari, H. Maghsoudi, A. Lavaei, R. Ahmadi, Optimal online trajectory generation for a flying robot for terrain following purposes using neural network. Proc. Inst. Mech. Eng. Part G **229**(6), 1124–1141 (2015)
5. S.J. Asseo, Terrain following/terrain avoidance path optimization using the method of steepest descent, in *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, vol.3 (Dayton, OH, USA, 1988), pp. 1128–1136. https://doi.org/10.1109/NAECON.1988.195148
6. T. Templeton, D. H. Shim, C. Geyer, S. S. Sastry, Autonomous vision-based landing and terrain mapping using an MPC-controlled unmanned rotorcraft, in *Proceedings 2007 IEEE International Conference on Robotics and Automation* (Rome, Italy, 2007), pp. 1349–1356. https://doi.org/10.1109/ROBOT.2007.363172
7. F. Ruffier, N. Franceschini, Visually guided micro-aerial vehicle: automatic take off, terrain following, landing and wind reaction, in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, vol.3 (New Orleans, LA, USA, 2004), pp. 2339–2346. https://doi.org/10.1109/ROBOT.2004.1307411
8. B. Hérissé, T. Hamel, R. Mahony, F. Russotto, A terrain-following control approach for a VTOL unmanned aerial vehicle using average optical flow. Auton. Rob. **29**(3–4), 381–399 (2010). https://doi.org/10.1007/s10514-010-9208-x
9. M. Garratt, J. Chahl, Vision-based terrain following for an unmanned rotorcraft. J. Field Rob. **25**(4–5), 284–301 (2008). https://doi.org/10.1002/rob.20239
10. P. Kaniewski, C. Leśnik, W. Susek, P. Serafin, Airborne radar terrain imaging system, in *16th International Radar Symposium (IRS)*, vol. 2015 (Dresden, Germany, 2015), pp. 248–253. https://doi.org/10.1109/IRS.2015.7226215
11. Characteristics and protection criteria for radars operating in the aeronautical radionavigation service in the frequency band 13.25-13.40 GHz. https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2008-1-201402-I!!PDF-E.pdf
12. H. Saltzman, G. Stavis, A dual beam planar antenna for Janus type doppler navigation systems, in *IRE International Convention Record*, vol. 1958 (New York, NY, USA, 1958), pp. 240–247 (1958). https://doi.org/10.1109/IRECON.1958.1150681
13. C.A. Balanis, A. Theory, *Analysis and Design*, 3rd edn. (John Wiley, Hoboken, NJ, 2005)
14. T.S. Cousik, K. Ameer Banu, H. Harsha Pillai, M. Jayakumar, Dependence of radius of cylindrical ground plane on the performance of coplanar based microstrip patch, in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (Kochi, India, 2015), pp. 675-679. https://doi.org/10.1109/ICACCI.2015.7275688
15. R. Gharg, B. Bhartia, *Ittipiboon, Microstrip Antenna Design Handbook* (Artech House, Inc, 2001)

16. J.R. James, Hall P.S. (eds.), *Handbook of Microstrip Antennas* 2nd edn (Peter Peregrinus, London, U.K. 1989)
17. I. Bahl, M. Bozzi, R. Garg, *Microstrip Lines and Slotlines* 3rd edn (Artech, 2013)
18. K. DurgaRao, K.N. Pillai, Design array antenna using different feeding technique in HFSS, *International Conference for Emerging Technology (INCET)*. (Belgaum, India, 2020) pp. 1–4. https://doi.org/10.1109/INCET49848.2020.9154127
19. N.H.M. Adnan, I.M. Rafiqul, A.H.M.Z. Alam, Effects of inter Element Spacing on large antenna array characteristics, in *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)* (Putrajaya, Malaysia, 2017), pp. 1–5. https://doi.org/10.1109/ICSIMA.2017.8311993
20. R.J.P. Douville, D.S. James, Experimental study of symmetric microstrip bends and their compensation. IEEE Trans. Microw. Theo. Tech. **26**(3), 175–182 (1978). https://doi.org/10.1109/TMTT.1978.1129340
21. ADAR1000 Datasheet. https://www.analog.com/en/products/adar1000.html#product-overview
22. A.B. Abdel-Rahman, M. Aboualalaa, Improving isolation between antenna array elements using lossy microstrip resonators, *2019 13th European Conference on Antennas and Propagation (EuCAP)* (Krakow, Poland, 2019), pp. 1–4
23. U. Krishnan, A. Muralidharan, B. Krishnan, S. K. Menon, Isolation enhancement of patch antenna array using open loop resonator, in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (Bangalore, India, 2018), pp. 1612–1616. https://doi.org/10.1109/ICACCI.2018.8554841
24. W. Jiang, Y. Huang, J. Wu, W. Li and J. Yang, A new approach for terrain following radar based on radar angular superresolution. Lect. Notes Electr. Eng. 223–231 (2015). https://doi.org/10.1007/978-3-319-08991-1_23

Emergency Medical Services Using Drone Operations in Natural Disaster and Pandemics

R. Anand, M. S. Muneshwara, T. Shivakumara, M. S. Swetha, and G. N. Anil

Abstract The recent innovations in Unmanned Aerial Vehicles (UAV) have the potential to revolutionize the healthcare sector especially in the domains of medical service delivery and transportation. In order to accomplish such task, drones are employed to deliver healthcare products such as drugs and medical kits to the patients without any physical contact. This method reduces the total time taken for the delivery of the drugs. By implementing the proposed unmanned systems, inaccessibility would no longer pose a threat to the delivery of drugs. The main aim of this paper is to develop the drone based service delivery idea with a particular emphasis on healthcare. Here, an android application has been developed to monitor the status of the drug delivery and provide an efficient, accurate as well as fast delivery of drugs.

Keywords Battery · Drone · Flight controller · Flying robot · Healthcare · Unamed aerial vehicles · Quadcopter

R. Anand (✉) · M. S. Muneshwara · M. S. Swetha · G. N. Anil
Department of Computer Science and Engineering, BMS Institute of Technology and Management, Bengaluru 560064, Karnataka, India
e-mail: anandor@bmsit.in

M. S. Muneshwara
e-mail: muneshwarams@bmsit.in

M. S. Swetha
e-mail: swethams_ise2014@bmsit.in

G. N. Anil
e-mail: anilgn@bmsit.in

T. Shivakumara
Department of Master of Computer Application, BMS Institute of Technology and Management, Bengaluru 560064, Karnataka, India
e-mail: shivakumarat@bmsit.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_19

1 Introduction

Technological advancements are not only changing the world around but also driving the wireless technology to next generation. A drone is an aerial vehicle that is controlled by the operator; the more the number of motors and propellers utilised, the greater the thrust produced. A drone is an autonomous flying robot that is used in a variety of areas such as military, sports, arts etc. Another significant use area is in the healthcare industry. The main function of using a drone is to deliver drugs during emergency situations [1].

Drones or unmanned aerial vehicles have improved the technology to the point that they have become a suitable transportation mode for drug deliveries. Drone deliveries are spotlighted because of three main advantages such as autonomous operation, avoidance of traditional traffic network, and velocity [2]. Using drones to deliver drugs can be cheaper and faster than other traditional methods. The main focus that comes when delivering drugs is the reliability of drone in the scheduling phase in order to increase the network reliability without changing the vehicle's hardware or software without adding any other extra costs.

The light composite materials and lithium batteries enable efficient flight of the drone and also they can fly further without requiring any repeated recharge. The status of the drug delivery can be known by an android application, which notifies whether the medicines are delivered or not. The application notifies when the medicines are out for delivery. This helps in tracking the deliveries.

The advent of technology has brought colossal changes in everyday life [3]. From simple tasks to advanced ones, the completion of almost every errand is directly or indirectly dependent on technology. A humble world without the use of advanced technology is a thing of the past and is not feasible to implement today. In such times where all sectors and domains have made tremendous progress, a niche has been created which needs to be filled by improvements made in the healthcare sector. Healthcare is one of the most essential sectors in the world which influence the mortality and comfort level of humans to great levels.

Traditional medicine and its implementation can now be deemed redundant and even obsolete to a certain level. New methods need to be developed to meet the needs of the day. The high population growth leaves humans vulnerable to tons of diseases including the ones already prevalent and new ones which arise due to the growth of new germs.

The healthcare sector is a burdened and overloaded sector as most of the work is done by humans themselves. The number of people trained to work in this sector is quite low leading to a high disparity between need and demand. Such a scenario needs a well-oiled mechanism that does away with quirks and gives way to accurate way of dealing with issues. Reduction of human influence will bring about such a change as human errors can be avoided by using high functioning technology [4].

Drones or Unmanned Aerial Vehicles can be utilized in the healthcare sector for multiple purposes. Delivery of drugs to intended recipients accurately is one such task. Feeding in the location of the recipients and their details for verification during delivery avoids the common human error of delivering to the wrong location. This increases the speed of delivery as well as the accessibility of drugs [5]. Drones can also be used to spray disinfectants in case of rampant infections as witnessed by the Covid-19 pandemic.

2 Literature Survey

The drone delivery models for healthcare source from the 50th hawaii international conference on system sciences, this helped us get a view on the drone delivery network models which were developed using an approach that simultaneously involved the concepts of orthodox land transportation. [6] The major methodology introduced in this field were drones that also used mobile technology which aided developing countries to be more advanced in the healthcare sector by implementing delivery to remote locations. The major advantages found were these models facilitated more timely, efficient and economical healthcare delivery.

A budget constraint to make the deliveries more affordable and cost effective measures were undertaken. Faster response would prevent medical trauma and save lives. Though certain challenges were faced where the actual road network was not explicitly modelled [7]. Obstacles such as high mountain range on a drone path are neither accounted nor integrated in the location model.

The importance of usage of digital image objects detection as well as recognition system using artificial neural networks and drones. It implements object identification in terms of person identification using face as a key of object processing. [8] This developed system was successful enough to acquire knowledge about the features of the analysed object which the designers failed to predict. The main aim was to detect various objects in various circumstances, example as different lighting and presence of interface [9]. The knowledge of just the operating algorithm is not enough but knowledge of the hardware is also needed for implementation of object recognition. Multi-thread processing and conversion of serial to parallel computation is a must for efficient algorithms.

Communication and coordination of drone networks which proposed high level architecture to be used to design the multi-UAV systems which consisted of on-board sensors, embedded processing, sensing, coordination and communication followed by networking blocks. [10] The major focus was on the interaction between the design blocks. Also the flight dynamics quadcopter, position and orientation of the UAV have greater impact on the communication links. This system is best used for managing disaster scenarios with strong interdependence between design blocks as well as efficient evaluation methods makes a huge advantage. Though the difficulty lies in measuring the level of interdependence between the design blocks with overall performance of the UAV.

Design of the quadcopter for aerial view and organ transportation using drone technology in which the UAV system implements a quadcopter emergency organ or drugs delivery via clear airways instead of crowded roadways using sonars to avoid obstacles and collisions. Designated recharging points are placed to recharge drones for long flights. The quadcopter has capability of carrying payload, the system also archives necessary lifts and provides surveillance of the terrain [11]. One of the main advantages is transmission of real time video or audio from inaccessible areas. Short battery life can be extended by placing recharging points to charge drones. Though the recharging points must be in the path of UAV [12].

A drone based wireless power transfer and communication platform in which drones are made cost effective for both wireless power transfer and communication [13]. The main aim is to make a light weight, cost effective, easy to implement and commercially available. This helps in collecting data from remote sensors. Data from remote areas or inaccessible areas which have weaker signals can be accessed [14]. Transfer of complex data that transmits about 1 Mbps without usage of any batteries or solar power. Though the voltage needed and noise formed may cause hindrance and to transfer data IQ modulation support is needed to transfer high data rates [15].

Optimization models created using linear regression to optimize the drone delivery schedule [16]. The goal was to develop optimization models that came up with the optimal drone depot locations and computed the delivery path. A parcel delivery system was developed for a multi-UAV system. The computational time for solving the operational planning model is computer by developing a variable pre-processing algorithm and utilising primal and dual bound generation methods [17]. The optimal solution pertains to the least number of drones required for delivery and their flight path keeping in mind their safe return in correspondence to battery life. The result obtained portrayed a linear relationship between BCR and payload where BCR is the Battery Consumption Rate which proves that the amount of payload affects the battery life of a drone [5].

The Acceptance of drone technology in healthcare forms another major criteria. The usage of the drones in healthcare especially the acceptance of these drones in hospitals [18]. Delivery drones can be very handy in healthcare fields. The emergency drones have high efficiency, high economic feasibility and have higher potential to save lives. The emergency drones are faster, more capable and have payload capability [19]. Though emergency drones have no safety protocols or social conduct parameters [7].

3 Proposed System

3.1 Problem Statement

The healthcare sector is such a crucial element of society that it decides the fate of billions. It has the power to cause catastrophes if mishandled and save lives if used appropriately [20]. A small progress can potentially save millions of lives while if left under-developed can lead to the death of those millions due to negligence and ailments that could have been otherwise easily treated [21].

The healthcare sector in spite of dedicated workers, is still plagued by issues that arise outside its scope and domain. The issues have a say over the availability and efficiency of the healthcare system and can potentially cause a lot of damages if left unchecked [22].

These issues include the failure of in-time delivery of medical treatment and drugs to patients who require it [23]. This may arise due to heavy traffic, road obstructions or shortage of delivery personnel. A lot of people who suffer from ailments which can be treated if aid is available under certain time limits have to suffer the consequences of not getting this required treatment [24]. For example, a patient may need medications for asthma urgently and the usage of inhalers would easily give relief to the patient. However, if this medication is unable to reach the patient in time, the patient may suffer from complications.

There are also issues that arise due to needy patients residing in inaccessible geographical locations. These may be due to physical barriers, curfews, lock downs and so on. In such times, humans cannot possibly deliver drugs to these areas causing patients to lose out on treatment [25]. Additionally, during rampant infections and pandemics, areas may need to be sanitized and sterilized to prevent future infections. Traditionally, it is done by health care workers but then they too run the risk of catching these infections [26]. Such issues need to be resolved to improve not just the healthcare industry but also the general health and well-being of the society to keep all the citizens safe and healthy.

3.2 System Architecture

The autonomous navigation for flying robot is the working principle for the architectural diagram. The flight controller receives the input signals from the device [27]. The input signals are in the form of digital signals. The flight controller gives the output to the motors directly (Fig. 1).

The measurements are taken from the sensors continuously and adjustments are made to the speed of each rotor so that the balance is maintained and to keep the body level stable [28].

To make the quadcopter be perfectly balanced the adjustments are done autonomously using a sophisticated control system [29]. A quad copter has four controllable degrees of freedom: yaw, roll, pitch, and altitude. Adjusting the thrusts of each rotor can be controlled by each degree of freedom (Fig. 2).

The speed of the regular rotating motors can be turned up by Yaw-Turing left and right, by taking away power from the counter rotating by taking away the same amount of power that is put on a regular motor has no extra lift since the counter torque is very less [30]. The speed in increased by on the motor and lowering on one side makes an roll-tilting left and right which can be controlled [31]. Pitch is moving up and down is controlled the same way as roll, but using the second set of motors [32].

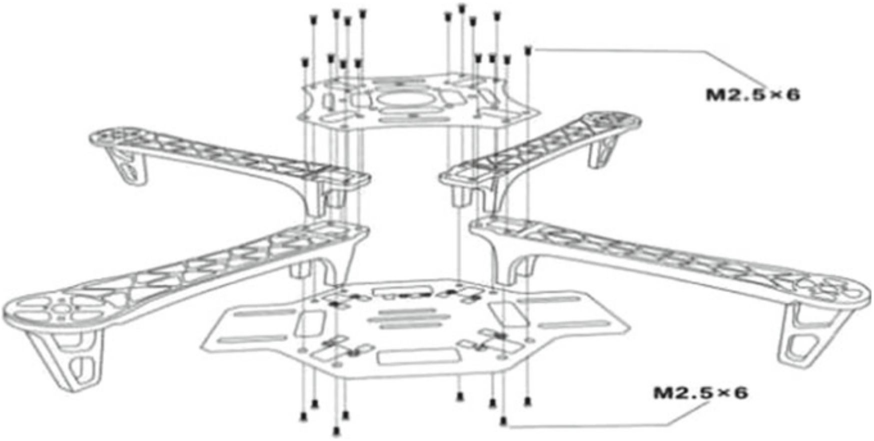


Fig. 1 System architecture of quad copter

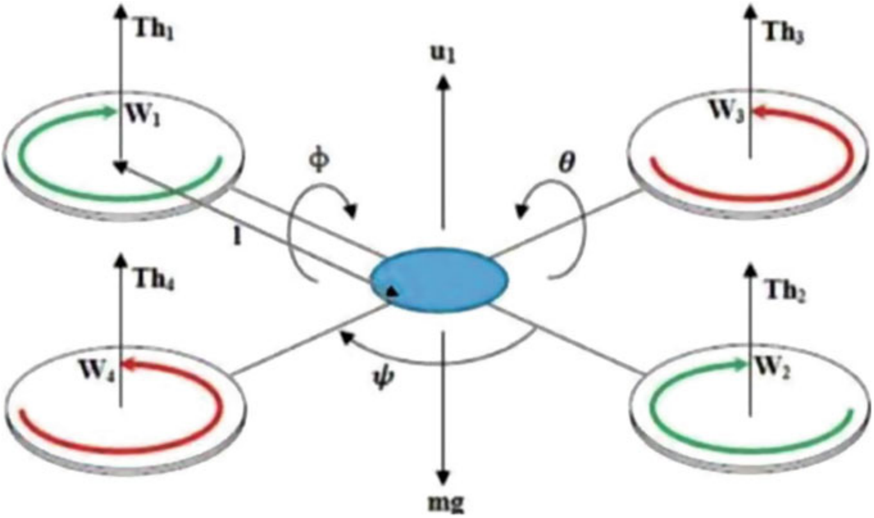


Fig. 2 Yaw-Turning left rotate

The thrust generated by the propellers is responsible for this type of movement of the quad copter. Thrust is directly proportional to the altitude. This is because the thrust is simply the downward force generated by the propellers (Fig. 3).

The Pitch is the main reason for the quadcopter motion [33]. The quadcopter moves in front and back direction because of the thrust of the forward and backward pair of motors (Fig. 4).

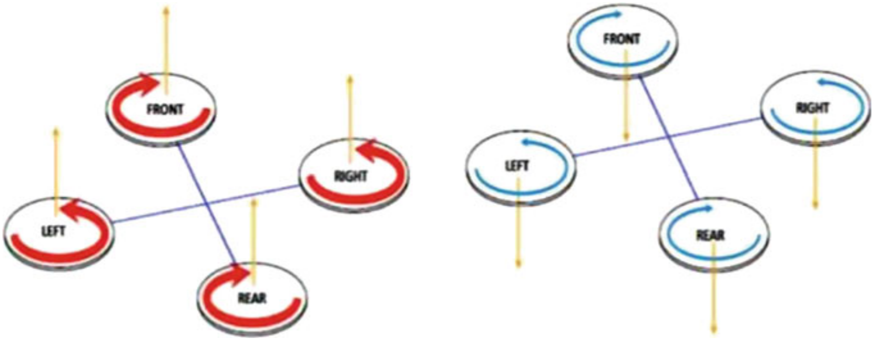


Fig. 3 Pitch axes quadcopter movements

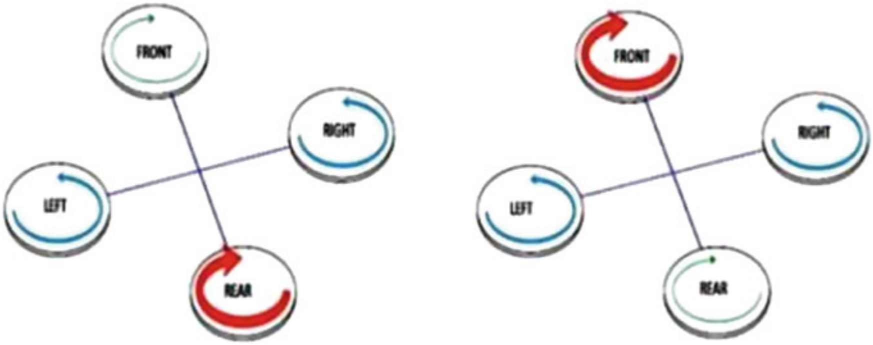


Fig. 4 Quad copter rotating direction

4 System Implementation

The implementation of the system is carried out in a sequence of steps. The first step involves block diagram analysis. This step encompasses all activities from developing a synopsis and abstract that depict the purpose of the system to coming up with a blueprint or block diagram of the required components for the successful completion of the propose system [34].

The next step involves the hardware development part. This step includes the selection of appropriate microcontrollers or processor system boards, IDE and debugging tools. Additional components and peripherals as required are selected [35]. Once this selection is done, the hardware assembly is done using apposite circuit designing. A sample program is developed and run. The testing of this sample program is a preliminary check to see if all components are up and running.

The next step is the development of software code to run the peripherals of the system. This code is different from the main logic code of the system and is involved with only the operation of the peripheral devices. A successful run of this code symbolizes that all peripheral devices are successfully operable in their befitting manners.

Logic development is the next step in the process. This step involves the formulation of logic depending on the aims and purposes of the system. A code that executes the established logic is developed. This core code is responsible for the successful deployment of the propose system in attaining its intended aims and goals.

The last step is the final testing of the system. This is done after all hardware, software and logical implementations are done. This is performed to check if the system successfully meets all its targets and is able to perform all its established goals. After the final testing is done, the performance of the system is analysed. The result analysis is documented and a corresponding report is created.

4.1 Android App Development

An android application is developed to monitor the flight of the drone and assess the deliver y of the drugs. The app gets notified whenever the drone takes off for a new develop and the status of the deliver y at the end of the flight. The information stating whether the delivery has been accurate or not is also sent to the app. This app is used by the client who sends in the drone for the delivery. The android application and the drone connected via Bluetooth. This is done by HC-05 which is a serial Bluetooth product consisting of Bluetooth serial interface and Bluetooth adapter. There are two types of Bluetooth serial interface module i.e. industrial level and civil level. The HC-05 belongs to civil level Bluetooth serial interface module and it is a mobile operating system that runs on any platform whether its windows, Linux or Mac OS. As initially developed by Android Inc, a firm later purchased by Google and lately by the open and set alliance. The device i controlled by Google developed java libraries and allowed all the developers to write code in java language Android relies on Linux version 2.6 for not only the core system services but also screw' s like security, memory management, process management, network stack and driver model. Between the hardware and the rest of the software stack the kernels acts as an abstraction layer.

Android not only includes java but also C/C++ libraries are also used by various components of the android item. The android application frame works are exposed to the developers and their capabilities. Some of the core libraries are listed as follows. System C library- a BSD derived implementation of the standard C system library, tuned for the embedded Linux based devices. The Media libraries play an important role in support playback and recording of audio's and video formats and they also include static image files (Fig. 5 and 6).

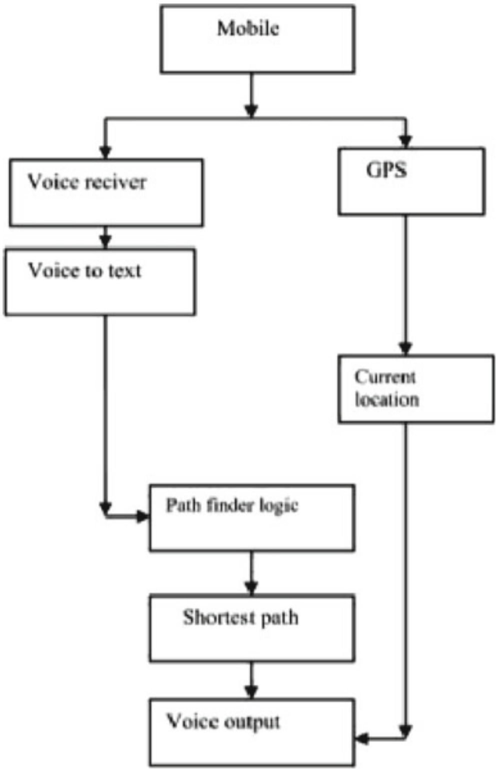


Fig. 5 Flowchart

4.2 Android Speech

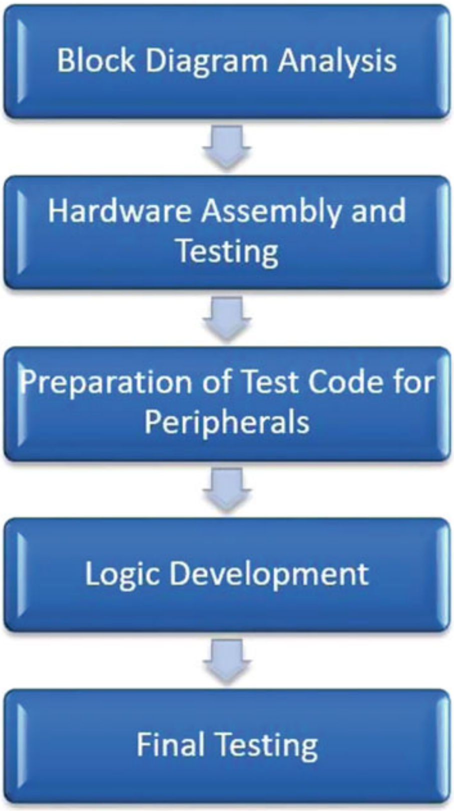
5 Results and Discussions

The drone was subjected to various tests and experiments to check its efficiency and effectiveness. The drone was judged on different parameters that corresponded various characteristics of the quadcopter. It was made sure that the drone met all required standards as to operate without any issues and complete its intended goal of delivering drugs.

The drone's flight performance was monitored first. The drone could successfully lift off from the ground, fly around, hover in space without moving and land as and when commanded. The drone did not drift while hovering. The capability of a drone to hover in air without drifting displays its stability and our drone was deemed stable for use.

The drone met all safety protocols that were issued for commercial purposes. The drone does not pose a risk to the environment, humans or property. The privacy of society is not compromised due to the flight of our drone.

Fig. 6 Steps of Implementation


Block Diagram Analysis
Hardware Assembly and Testing
Preparation of Test Code for Peripherals
Logic Development
Final Testing

The drone can be easily setup after disassembling. This proves useful in case the drone ever malfunctions and needs to be disassembled for repairing purposes. The ease with which it can be reassembled increases the overall ease in setup, operation and repair.

The drone is light in weight and small in size. This makes it easy for the drone to be transported from one place to another deeming it portable. Portability assures that the drone can easily be transported while not in use thus, having no influence over its flight time and mobility. However, the drone performs better during flight. It can enter small spaces and fly a longer time than would have been the case had the size been bigger and the weight higher.

The drone was tested in various climatic conditions like high temperature (afternoons during hot days), low temperature (cold nights), varying humidity(measured on different days) and rainfall. The varying climatic conditions seemed to not have much of an impact on the performance of the drone. This made the drone stable in operation.

The drone responded well to the stress test without faltering unless a high payload was added. The drone could take a load of up to 400 gm without any effect on its

Table 1 Drone characteristics

Weight	2.3 kg
Maximum payload	0.4 kg
Maximum flying time	13 min

performance. However, the payload bearing capability could be increased further to increase the amount of supplies that it could deliver at once. This could be done by using more resilient and strong materials which do not add additional bulk to the weight of the drone.

The drone is free of hazardous chemicals and is safe to touch and use. Its components are all electrically insulated so as not to cause any electrical shock or similar damage. All the features, accessories and functionalities added to the drone work properly. The code of the drone works efficiently to let the remote pilot fly the drone in the required direction. However, the code is not well equipped with safety protocols against potential cyber attacks. The camera attached to the drone performed well and captured frequent images of the route as well as of the recipient at the end of the delivery. The navigation system worked well as the we were able to track the drone's whereabouts during its flight.

The android app connected to the drone worked properly as well. We were able to place orders, track the drone, monitor the flight and receive notifications of the status of the delivery. The app notified the user in case there was a failure of delivery so that the cause of failure could be analyzed and efforts made to correct it.

The drone was not able to communicate over long distances as it was restricted by the limited reach of Bluetooth. The drone when flying over a long distance lost connection with the base.

The battery life of the drone turned out to be limited making it difficult to make it fly over long distances. Backup batteries were added to curb this limitation. However, this increased the overall weight of the drone leading to a reduction in its payload bearing capacity without any hindrance to its flight time. When the payload was maintained at its maximum amount, the maximum flying time reduced (Table 1).

6 Conclusion and Future Enhancement

A drone was able to be successfully developed for the delivery of drugs. The drone is able to complete its intended task without any hurdles or obstacles. It effectively cuts down the time for delivery of drugs, medical equipment and kits. This increases the accessibility of the healthcare system as patients in remote locations can also be reached with this system. Lives are thus saved with a faster and more accurate delivery system. The android application works successfully notifying the user when the drone takes off for a delivery and when it successfully delivers the product. Failure in delivery is also notified. The flight of the drone can be monitored and the drone tracked using GPS. The camera successfully takes pictures along the route and at the time of delivery.

Thus, modifying a drone to mimic a delivery system which can effectively deliver drugs, first aid equipment and blood among other items, unnecessary risks posed to public health can be minimized while improving the quality of medical services. The benefit gained by communities reading in remote areas highlights the importance of the UAV application.

Although, the system proves quite effective in its goals, it doesn't come without any limitations and challenges. The main issue with UAVs is the short flying time which affects the drone built as a part of this paper. Short flying time essentially means that reaching locations far off from the starting point becomes difficult. The flying time can be changed by modifying the drone size, the propeller size, the diameter, the wing span and other components but this would affect the performance of the drone in other aspects.

References

1. M. Torabbeigi, G.J. Lim, S.J. Kim, Drone delivery schedule optimization considering the reliability of drones, in *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2018
2. P. Vijay Kumar, A. Challa, J. Ashok, G. Lakshmi Narayanan, GIS based fire rescue system for industries using Quad copter—A novel approach, in *2015 International Conference on Microwave, Optical and Communication Engineering (ICMOCE)*, 2015
3. D. Pietrow, J. Matuszewski, Objects detection and recognition system using artificial neural networks and drones, in *2017 Signal Processing Symposium (SPSympo)*, 2017
4. E. Yanmaz, M. Quaritsch, S. Yahyanejad, B. Rinner, H. Hellwagner, C. Bettstetter, Chapter 7 communication and coordination for drone networks. Springer Science and Business Media LLC, 2017
5. J.E. Scott, C.H. Scott, chapter 16 Models for drone delivery of medications and other healthcare items, IGI Global, 2019
6. J.E. Scott, C.H. Scott, Drone Delivery Models for Healthcare. HICSS (2017). https://doi.org/10.24251/hicss.2017.399
7. Mobile Multimedia Processing, Springer Science and Business Media LLC, 2010
8. D. Pietrow, J. Matuszewski, Object Detection and Recognition System Using Artificial Neural Networks and Drones. J. Electr. Eng. **6**, 46–51 (2018). https://doi.org/10.17265/2328-2223/2018.01.007
9. S. Nimara, A. Mereu, M. CrisanVida, R. Bogdan, Portable device for remote control of a vehicle, in *2018 26th Telecommunications Forum (TELFOR)*, 2018
10. E. Yanmaz, S. Yahyanejad, B. Rinner, H. Hellwagner, C. Bettstetter, Drone networks: communication. Co-ordination and Sensing. Elsevier. Doi **10**(10), 16 (2017)
11. S. Selvaganapathy, A. Ilangumaran, 30 April 2017, Design of quadcopter for aerial view and organ transportation using drone technology
12. A. Tiurlikova, N. Stepanov, K. Mikhaylov, Wireless power transfer from unmanned aerial vehicle to low-power wide area network nodes: Performance and business prospects for LoRaWAN". Int. J. Distrib. Sens. Netw. (2019)
13. H. Xuanke, B. Jo, M. Tentzeris, Manos, A Drone based wireless power transfer and communication platform, in *IEEE Wireless Power Transfer Conference (WPTC)*
14. C.N. GireeshBabu, M. Thungamani, S.K. Pushpa, M.S. Muneshwara, in *2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*

15. M.S. Muneshwara, H. Vallae, M.S. Swetha, M. Thungamani, Comparison on hyper ledger fabric and hyper ledger composer of block chain technology, in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*
16. T., Maryam, Lim, G.J., Kim, S. Jin, Drone delivery scheduling optimization considering payload induced battery consumption rates. J. Intell. Rob. Syst. (2019).10.1007
17. M.S. Muneshwara, A. Lokesh, M.S. Swetha, M. Thunagmani, Ultrasonic and image mapped path finder for the blind people in the real time system, in *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*
18. K. Mike, S. Roger, Usage and acceptance of drone technology in healthcare
19. M.S. Muneshwara, M.S. Swetha, M. Thungamani, G.N. Anil, Digital genomics to build a smart franchise in real time applications, in *2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT)*
20. E. Frachtenberg, Practical drone delivery. Computer **52**(12), 53–57 (2019)
21. R.P. Anand, R.M. Patil, Health monitoring in aerospace system. Int. J. Inf. Futuristic Res (IJIFR) (2017)
22. M. Alwateer, S. W. Loke, On-drone decision making for service delivery: Concept and simulation, in *Proceedings IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Mar. 2019, pp. 937–942
23. M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, M. Debbah, A tutorial on UAVs for wireless networks: applications, challenges, and open problems. IEEE Commun. Surveys Tuts. **21**(3), 2334–2360 (2019)
24. R. Anand, M. Pushpalatha, R.M. Patil, A social networking for sharing infrastucture resources in the social cloud computing. Int. J. Inf. Futuristic Res. (IJIFR) 2016
25. M.V. Vijaykumar, P. Jagadish, K. Shryavani, R. Anand, Authorized deduplication in hybrid cloud. IJCSN Int. J. Comput. Sci. Netw. 2016
26. D. Schneider, 'The delivery drones are coming.' IEEE Spectr. **57**(1), 28–29 (2020)
27. V. Chamola, V. Hassija, V. Gupta, M. Guizani, A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. IEEE Access **8**, 90225–90265 (2020)
28. M.S. Muneshwara, B.R. Rajendra, Intelligent robot positioning system (IRPS) for tracing the contemporary location. IAETSD J. Adv. Res. Appl. Sci. Sci. J. Impact Factor—5.2 Indexed by: Thomson Reuters' Research ID : H-2404–2017 Volume 4, Issue 1, Indexed Journals SCOPUS SUGGESTED JOURNAL ID:50E4CF07B9886F83 UGC APPROVED JOURNAL – JARAS.
29. M.S. Swetha, S.K. Pushpa, M.S. Muneshwara, T.N. Manjunath, Blockchain enabled secure healthcare Systems, in *2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*
30. J.S. Raj, A novel information processing in IoT based real time health care monitoring system. J. Electron. **2**(03), 188–196 (2020)
31. S. Shakya, L. Nepal, Computational enhancements of wearable healthcare devices on pervasive computing system. J. Ubiquitous Comput. Commun. Technol. (UCCT) **2**(02), 98–108 (2020)
32. F. Semiz, F. Polat, Solving the area coverage problem with UAVs: a vehicle routing with time windows variation. Robot. Auto. Syst. **126** (2020). Art. no. 103435
33. E. Yakıcı, 'Solving location and routing problem for UAVs.' Comput. Ind. Eng. **102**, 294–301 (2016)
34. D.K. Brands, E.T. Verhoef, J. Knockaert, P.R. Koster, 'Tradable permits to manage urban mobility: Market design and experimental implementation. Transp. Res. A, Policy Pract. **137**, 34–46 (2020)
35. K. Dorling, J. Heinrichs, G.G. Messier, S. Magierowski, 'Vehicle routing problems for drone delivery. IEEE Trans. Syst., Man, Cybern. Syst. **47**(1), 70–85 (2017)
36. J.E. Scott, C.H. Scott, Chapter 3 Drone delivery models for medical emergencies. Springer Science and Business Media LLC, 2020
37. M.Y. Arafat, S. Moh, Routing protocols for unmanned aerial vehicle networks: a survey. IEEE Access **7**, 99694–99720 (2019)

Design and Analysis of Digital Beamforming for SATCOM on the Move Based on Specific Geographical Area

R. Neeraj, P. Dinesh, V. Pavan Kumar, N. Umeshraja, and M. Jayakumar

Abstract SATCOM on the move using digital beamforming is one of the latest technological breakthroughs due to the advancement of RF systems and computational capabilities. Digital broadcasting by the satellite is getting momentum when used with beamforming antenna systems. We have selected the specific geographical area and mapped the look angles of the ground station with the satellite while on the move. It was found that for the Indian continent, the look angle varies from 48.7° to 3.8° Azimuth, 52.7° to 59.4° Elevation angles from east west and from 31.03° to 63.9° Azimuth, 45.5° to 69.2° Elevation angles from North to South. The digital beamforming antenna systems with minimum angle and optimised weight factors have been computed and evaluated. In order to maintain line of sight with the satellite, it was found that the minimum computational capabilities with high gain antenna array is essential to realize this proposed system.

Keywords Circular patch microstrip antenna · HFSS · Look angles · Beamsteering · SOTM · Beamforming

1 Introduction

Beamforming is a process of combining the antenna array outputs to produce a beam in a particular direction. Concentrating the beam from an antenna array spatially towards the required direction has many advantages [1]. Using this process gain of the antenna system can be increased enabling long range wireless communication, and it can also filter noise emanating from other directions.

Beam steering can be done mechanically or electronically. The advantages of using electronic beam steering are it is fast to steer the main beam towards the desired direction, and it has fewer mechanical parts reducing the rate of failure of the system.

R. Neeraj (✉) · P. Dinesh · V. Pavan Kumar · N. Umeshraja · M. Jayakumar
Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: m_jayakumar@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_20

The system requires different sub-systems for its functioning, and all the individual sub-systems are explained in this paper. Explaining the concept of beam forming in short, it is simply the process of providing phase shift to each antenna element of the antenna array system in the form of antenna weights, creating a spatial filter and receiving only the required signals coming from a particular direction. Some of the techniques have been explored in [2, 3]. So, the first part of the system is to determine the look angles aiming towards the geostationary satellite from the base receiver system located on Earth.

The coordinates to which an earth station's bore sight must be pointed to communicate with a satellite are called look angles. The two antenna look angles are Azimuth and Elevation angles. The Azimuth angle is measured eastward from geographic north to the projection of the path to satellite on a local horizontal plane at the earth station, whereas the angle of elevation of a satellite is that angle which appears between the line from the earth station antenna to the satellite and the line from the earth station antenna to the earth's horizon. The estimation of look angles was discussed in [4–7].

Next part of the system is the individual antenna element. An inset fed circular patch antenna is constructed and simulated using Ansys HFSS (High Frequency structure simulator). The final part of the system is to calculate the weights of the antenna array elements and steer the main beam according to the look angles determined (Fig. 1).

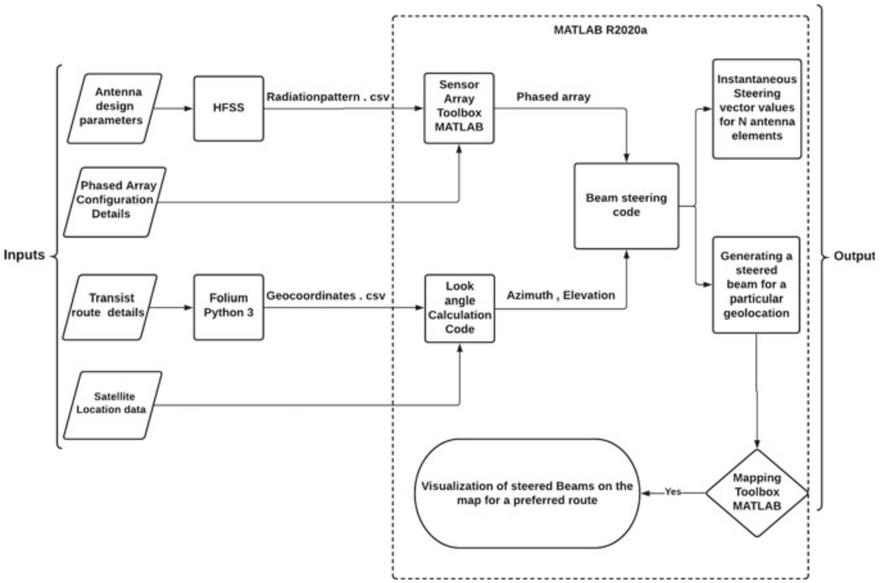

MATLAB R2020a
Inputs
Antenna design parameters
HFSS
Radiationpattern . csv
Sensor Array Toolbox MATLAB
Phased array
Instantaneous Steering vector values for N antenna elements
Phased Array Configuration Details
Beam steering code
Output
Transisit route details
Folium Python 3
Geocoordinates . csv
Look angle Calculation Code
Azimuth , Elevation
Generating a steered beam for a particular geolocation
Satellite Location data
Visualization of steered Beams on the map for a preferred route
Yes
Mapping Toolbox MATLAB

Fig. 1 A concise overview of the project

2 SATCOM on the Move

SATCOM on the Move (SOTM), also known as SATellite COMmunications On The Move, is the phrase used in the context of mobile satellite communication technology where the basic principle behind SATCOM On The Move is that a vehicle equipped with a satellite antenna is able to establish communication with a satellite and maintain that communication even while the vehicle is in motion. When it come to SOTM, the three main requisites for the proper functioning of the technology are Long distance communication capacity, line of sight (LOS) and mobility. Beamforming is the most successful technique that can satisfy all the three requisites [2, 8].

3 Description of Proposed System

The performance of a satellite communications system can be optimized by pointing the direction of maximum gain of an earth station antenna (referred to as boresight) directly at the satellite. To do the above-mentioned process, the concept of beamforming (one of many available techniques) can be utilized. This proposed system aims to construct one such system for the applications of enhanced communications between mobile earth receivers and geostationary satellites as part SATCOM on the move system.

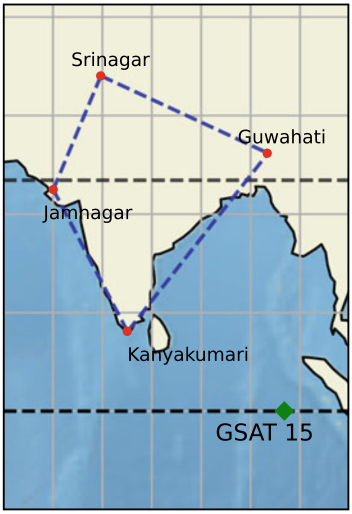
The construction of this system is done in parts, namely antenna element construction, look-angle determination systems and beam steering system. FEM-based EM simulator software—Ansoft's HFSS is used for the construction of individual antenna. Determining look angles can be done using Python, MATLAB or using any other preferred programming language. The beam steering angles for the antenna elements are determined using MATLAB.

3.1 Determination of Look Angles

To determine the look angles, the requirements are the current location of the base receiver system in terms of longitude and latitude and the contour or borders on the Earth within which the base receiver system is expected to be in. For this application, the contour is determined by choosing the cities in the extreme of North, South, East and West directions of India.

The selected cities are Srinagar (North), Kanyakumari (South), Guwahati (East) and Jamnagar (West). These cities are the four vertices of the desired contour as shown in Fig. 2. The contour is required because any geostationary satellite does not have a line of sight to all the parts of Earth. A carefully chosen contour gives an assurance that the geostationary satellite has a line of sight to all the points within

Fig. 2 Four extreme places of intended landscape for Look angle determination



the contour. After the contour is set, look angles of any latitude and longitude within the contour can be determined.

For the model of look angle determination system implemented, the longitude and latitude of the base station are assumed to be known at any given point of time. Now that the latitude and longitude of the GSAT15 are fixed and known, and the look angles can be determined. Before the look angles can be determined, the location of the base station must be made sure to be within the field of view of the satellite which in this case is the GSAT15. The contour chosen, shown in the figure, is well within the field of view of GSAT15. As all the conditions are satisfied, the look angles can be determined from any point within the contour.

The equations for determining the look angles as given by Sudhakaran [9]

$$\cos(\gamma) = \cos(L_e)\cos(l_s - l_e) \tag{1}$$

$$d = r_s\left[1 + \left(\frac{r_e}{r_s}\right)^2 - 2\left(\frac{r_e}{r_s}\right)\cos\gamma\right]^{1/2} \tag{2}$$

$$El = \sin^{-1}\left[\frac{(r_s\cos(\gamma) - r_e)}{d}\right] \tag{3}$$

$$A = \tan^{-1}\left[\frac{\tan|l_s - l_e|}{\sin(\mathrm{L_e})}\right] \tag{4}$$

where L_e—Latitude of earth station,
l_e—Longitude of earth station,
l_s—Longitude of geostationary satellite,

d—Distance between earth station and satellite,
r_e—Distance between centre of earth and earth station,
r_s—Distance between centre of earth and satellite,
El—Elevation angle (in radians),
A—Azimuth angle (in radians).

3.2 Construction of Antenna Element

HFSS is the software used to construct and simulate the antenna element. The constructed antenna element is an Inset fed circular patch antenna. It was decided to use a patch antenna array system as it is compact and easy to construct and manufacture according to the specifications. Circular patch has lesser side beams compared rectangular and triangular patch antennas.

Circular microstrip patch antenna is designed for a target frequency of 12GHz in the Ku band. The intended application for the antenna array and the beamforming is mainly for providing DTH services for vehicular systems link buses, trains, etc. The reason for choosing circular-shaped patch for antenna element is for its lower side bands. The design equation for circular patch in dominant mode TM_{110} is given in [10, 11]. CMPA has been fed using microstrip line with inset feeding technique. Width of microstrip line used is given by Pozar [12]

Using the equation mentioned in [10, 12], parameters of the antenna are optimized in order to achieve optimal performance in 12 GHz frequency. Length and width of the substrate are taken as 11.5 mm. Dielectric material used is RT/duroid 6006 with the dielectric constant of 6.15. Impedance matching is achieved exactly at 12 GHz with higher return loss value which is shown in Fig. 5.

3.3 Processing Data for Beamforming

After the construction of single antenna element and simulating it in HFSS, the radiation pattern from the simulation results is exported to .csv format. This csv file is accessible by MATLAB, and this data is used to get the model of antenna element in MATLAB. This antenna element imported into MATLAB through csv file is used to make an 8 × 8 uniform rectangular array as shown in Fig. 3.

Simulating this array provides the radiation pattern which is shown in Fig. 7. The main beam shown in the radiation pattern can be steered by providing preferred Azimuth and Elevation angles to the beamsteering function, which calculates the weight vector. The beamsteering is performed by multiplying the acquired weight vector with the signal matrix of the antenna array. Now that the phased array has been created, the next step is to assess the variation in the look angles over a sample route, over which base station when mounted on to a vehicle can be expected to travel. Once the sample route is chosen, the routes were plotted and extracted from an open-

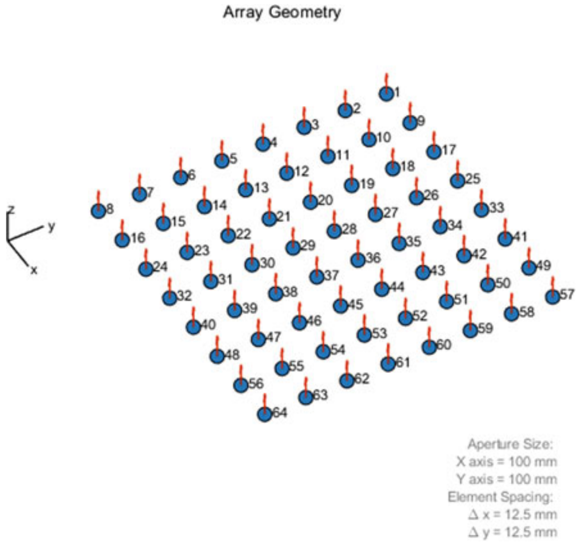


Fig. 3 Schematic view of 8 × 8 phased Array made of Circular Patch Antenna elements

source routing machine (OSRM) where the coordinates of the routes are chosen based on the linearity of the route. All the points are extracted as GPS exchange format and converted to csv file. These extracted points of coordinates are represented by geocoordinates.csv in Fig. 1. This file is then imported into MATLAB and given as an input to the look angle calculation system, which then calculates look angles for every point in the form of azimuth and elevation angle values corresponding to the latitude and longitude values $[\theta, \phi]$.

Finally, after creating the phased array and after obtaining the look angles, the steering vector matrix for a particular value of look angle can be obtained. For an N element antenna, N steering vectors will be obtained for a particular look angle value.

4 Result and Discussion

Azimuth and elevation angles for the beam to point towards the GSAT15 for antennas at respective places are:-
(NOTE - The format followed here is [azimuth, elevation] in Degree)

Kanyakumari: [63.8082143475072, 69.0721113210881]
Jamnagar: [48.618817106880606, 52.69229817576906]
Srinagar: [31.09483352985306, 45.68081412962961]
Guwahati: [3.9452895243989166, 59.35064394125674]

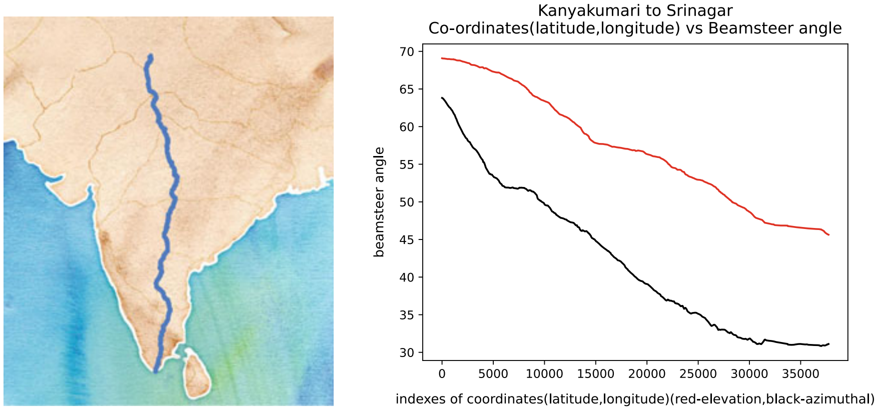


Fig. 4 Shows the change in look angles when moving from Kanyakumari to Srinagar (Map source-folium python package)

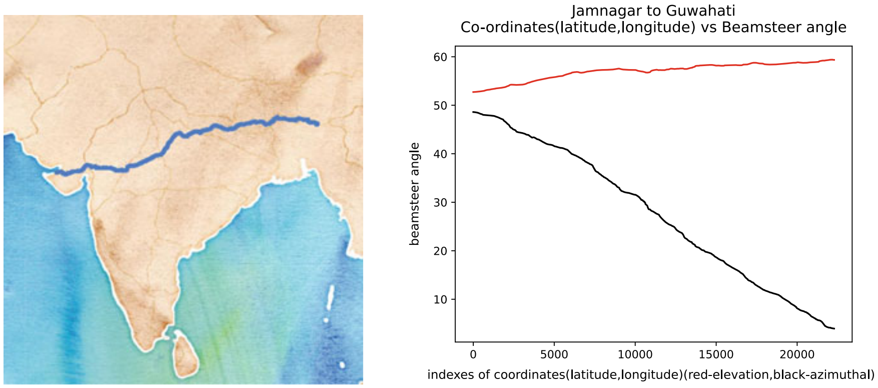


Fig. 5 Shows the change in look angles when moving from Jamnagar to Guwahati (Map source-folium python package)

The Azimuth and elevation angles calculated from these above-mentioned four places tell the range of the angles within which the steering angles are expected to be for the chosen area of operation which is shown in Fig. 8.

Return loss curve of the designed antenna element is shown in Fig. 6. The frequency of operation for this application is 12GHz. The target is to make an antenna array system with a total gain of at least 20 dB. HFSS simulation shows that 8 × 8 array antenna has a gain of 23.99 dB with HPBW of 13°. (Figs. 4, 5, 7 and 4)

One of the key features of our work is that the generated beam can be visualized for the purpose of determining whether the beam that is generated is optimal. Another feature is that the beam can be steered and the entire events can be visualized on a Mapping visualization tool. Using MATLAB and the map visualization tool for four sites on the map, namely Srinagar, Kanyakumari, Guwahati and Jamnagar,

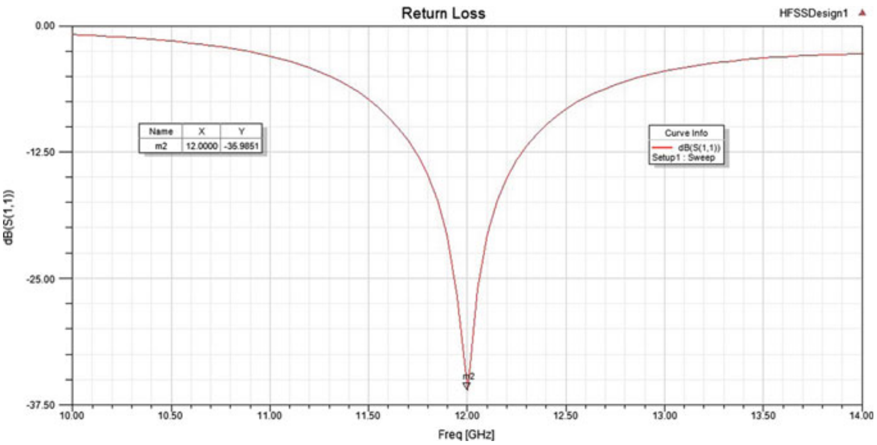


Fig. 6 Return Loss curve of the designed CMPA with the return loss of −35.98 dB at 12 GHz

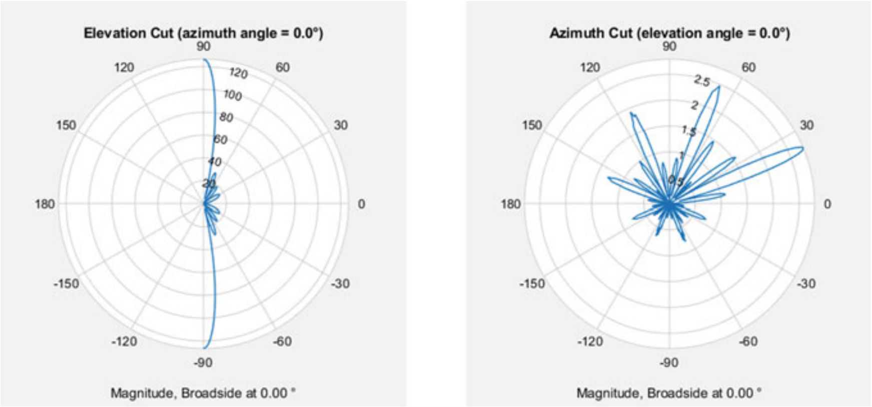


Fig. 7 Radiation pattern of antenna array

the latitude and longitude have been taken, and the steered beam according to the respective look angle values of each of the places has been visualized as shown in Fig. 8.

5 Conclusion and Future Work

A system has been constructed in a simulated environment which is capable of calculating weights to perform electronic beamsteering for the phased array circular microstrip patch antenna with the target frequency of 12 GHz. The antenna weights calculation system can be implemented in single board microcomputers

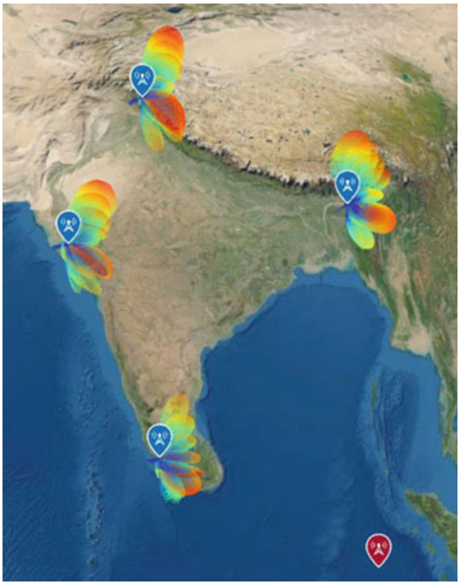


Fig. 8 Four receiver sites with beam steered radiation pattern and the GSAT15 sub-satellite location (Red Blip) over the Indian Ocean at 93.5° E (Map Source - GMTED2010 7.5 MATLAB)

like Raspberry Pi, proving that the system can implemented and constructed onto mobile vehicles, enabling Direct-To-Home (DTH) Broadcasting services and many such services that needs SATCOM On The Move technology. Digital beamforming can prove to be a challenging one, to get results on a real-time basis, as it is computationally expensive.

References

1. J.Lu. Wenhao Xiong, X. Tian, G. Chen, K. Pham, E. Blasch, "Cognitive radio testbed for digital beamforming of satellite communication, Cognitive Communications for Aerospace Applications Workshop (CCAA). Cleveland, OH, USA **2017**, 1–5 (2017). https://doi.org/10.1109/CCAAW.2017.8001885
2. C. Sahana, M. Jayakumar, V.S. Kumar, High performance dual circularly polarized microstrip patch antenna for satellite communication, in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (Bangalore, India, 2018), pp. 1608–1611. https://doi.org/10.1109/ICACCI.2018.8554817
3. S.J. Gopal, J. Ramnarayan, S. Kirthiga, M. Jayakumar, M. Nirmala Devi, R. Gandhiraj, S. Chandra Bera. Capacity analysis of correlated MIMO in GEOSAT downlink land mobile system, in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (IEEE, 2018), pp. 1594–1600
4. D.A. Ogundele, S.Y. Aiyeola, Y.A. Adediran, E.O. Oyedeji, O.F. Oseni, Model validation and analysis of antenna look angles of geostationary satellite, *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, (Zhangjiajie, China, 2012), pp. 509–513. https://doi.org/10.1109/CSAE.2012.6272824

5. S. Tomas, W. David, *Determination of Look Angles to Geostationary Communication Satellites* (National Geodetic Survey, Silver Spring, MD20910, 1994), pp. 115–126
6. E.D. Williams, Basic of Satellite. Antenna Positioning, RF and Communication. Engineering, Steven Water Monitoring. Systems, Inc
7. T. Wayne, *Electronic Communications Systems: Fundamentals Through Advanced*, 4th ed. (Pearson Education, Inc., 2001), pp. 790–800
8. A.J. Singh, M. Jayakumar, Machine Learning based Digital Beamforming for Line-of-Sight optimization in Satcom on the Move Technology, in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (Coimbatore, India, 2020), pp. 422–427
9. G. Sudhakaran, B. Kandipati, G.B. Surya, V.K. Shree, M. Sivaprasad, M. Jayakumar, Evolutionary algorithm based structural optimization for patch antenna design and its performance analysis, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (Udupi, India, 2017), pp. 2189–2192. https://doi.org/10.1109/ICACCI.2017.8126170
10. C.A. Balanis,*Antenna Theory: Analysis and Design*, 2nd ed. (John Wiley & Son, Inc., 1997)
11. P.S. Hall, J.R. James, C. Wood, *Microstrip Antenna: Theory and Design* (Peregrinus, United Kingdom, 1986)
12. David M. Pozar, *Microwave Engineering* (Wiley, Hoboken, NJ, 2012)

Hybrid Encryption Algorithm for Storing Unimodal Biometric Templates in Cloud

Jayshil Dave and M. Gayathri

Abstract The recently developed digital applications are requiring increased security and accessibility, and the cloud has become as the primary source for providing a platform for such applications. The existing major cloud service providers provide a wide range of services to manage a massive quantity of data on a regular basis. Each use case has the same requirement: secure data storage and transmission. Biometric authentication is the most popular form of security in this sector, although it has several drawbacks and limitations in terms of scalability. This paper discusses the advantages and disadvantages of storing biometric data on the cloud. It shows a cloud-based hybrid encryption method for unimodal biometric templates. The algorithm that has been proposed is generic. The overall impact of the existing algorithms and the challenges faced by using them are also briefly discussed.

Keywords Cloud computing · Encryption algorithm · Cloud security · Data security · Biometric templates

1 Introduction

1.1 Cloud Computing Definition

Cloud computing can be defined as a service for establishing on-demand supply of computational resources such as storage, computing power, etc. Cloud is remaining as one of the best testimonials of the twenty-first century. This technology can be used by any and every organization depending on their requirement. The most popular use of cloud computing is in Data Centres. As cloud computing can meet any IT requirements at whatever point it is very advantageous to organizations. Also, the

J. Dave (✉) · M. Gayathri
Computer Science and Engineering, SRM Institute of Science and Technology Chennai, Chennai, India
e-mail: jy2647@srmist.edu.in

M. Gayathri
e-mail: gayathrm2@srmist.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_21

“Pay-As-You-Go” policy for users urges them to invest time and money into cloud. Some of the leading Cloud Provider companies like Amazon, Google, Microsoft, IBM offer a wide range of cloud services for developing, deploying and hosting applications on cloud. Cloud Computing enables support and monitor the different features with immediate assistance of resources and proper documentation.

1.2 Encryption Algorithms

An encryption algorithm can be defined as a computation function, which manipulates the data to convert it to unreadable form. Such algorithms are used to hide the original data and secure it from unauthorized access. Some commonly used algorithms are AES, RSA, DES, Homomorphic, and so forth. Every algorithm performs two operations namely encryption and decryption. Let us discuss some of the majorly used algorithms.

1.2.1 Data Encryption Standard (DES)

DES is a symmetric-key algorithm used for encrypting the desired data. The algorithm makes use of a single key for performing both encryption and decryption. The data is divided into 64-bit block sizes, which goes through 16 rounds of encryption. The input data is called plaintext, which is isolated. The plaintext undergoes numerous operations of replacement DES calculation, which comprises of two changes (P-boxes) and 16 encryption feistel adjusts. The whole activity is separated into three stages.

1. Initially the substitutions revamp the pieces of 64-cycle plaintext. It does not utilize any keys, working in a predefined structure.
2. There are 16 fiestel which adjusts in second stage. Each round uses an alternate 48-bit round key applies to the plaintext pieces to create a 64-bit yield, produced by a predefined calculation. The key is shortened to 48-bits from the 56-bits.
3. At long last stage all the permutations are reversed and the yield is a 64-bit figure text.

1.2.2 Progressed Encryption Standard (AES)

AES is a symmetric-key block cipher which is derived from the subset of the original Rijndael block cipher. AES was established in 2001 and is recognized as the most secure algorithm. It works calculation on bytes instead of pieces, treats 128 bits as 16 bytes. The bytes are divided and arranged into matrix form. AES algorithm is different from DES in the aspect that the matrix size i.e., the number of bytes determine the number of rounds of encryption the input data goes through. Below listed are the number of rounds with respect to the key size:

- 10 rounds for 128-digit keys.
- 12 rounds for 192-piece keys.
- 14 rounds for 256-bit keys.

Significant points of interest of AES over DES are

1. Information block size is 128 pieces.
2. Key size 128/192/256 pieces relying upon form.
3. Most CPUs presently incorporate equipment AES uphold making it quick.
4. Potential keys are 2128, 2192 and 2256
5. Safer than DES

1.2.3 Rivest-Shamir-Adleman (RSA)

RSA is an asymmetric encryption algorithm. It is a well-known deviated key cryptographic calculation. This calculation utilizes different information block size and different size keys. It has uneven keys for both encryption and unscrambling. It utilizes two prime numbers to produce people in general and private keys. These two unique keys are utilized for encryption and unscrambling reason. This calculation can be comprehensively characterized into three phases; key age by utilizing two prime numbers, encryption and decoding.

RSA today is utilized in several product items and it can be utilized for key trade, advanced marks, or encryption of little squares of data. This calculation is fundamentally utilized for establishing secure correspondence and validation upon the open correspondence channel.

1.3 Biometrics

Biometrics is analytical body measurements, which are unique to each human being. Biometrics alludes to the programmed confirmation of a person's physiological or social qualities. Today, biometric authentication is booming and it becomes as one of the most trending and secured forms of authentication. A biometric authentication system comprises of 5 significant parts. They are; Sensor, Feature Extractor, Template Generation, Matching Module and a Verification Module. Biometric templates are registered and verified to confirm client identities. Biometric traits include features like face, fingerprint, iris, vein pattern, voice print, etc.

There are 2 major types of biometric authentication systems:

1. *Multi Modal System:* Multi Modal Systems combine two or more biometric traits. In these kinds of systems either one trait can be fixed and merged with the other to form the final template or both can be different each time for each person. Generally, the former approach is followed where an admin biometric print is merged with the user's template [1].

2. *Unimodal System:* Unimodal systems only use a single biometric trait for registration and verification. Over the years there have been many issues concerning higher recognition and acceptance which make the unimodal systems less effective [1].

We led this survey to discover the different biometric assaults and dangers that have been recorded in existing writing and decide the biometric layout security plans and methods presently being utilized towards making sure about biometric unique mark layouts in a biometric system's information base. Some known type of attacks on the physical sensor and logic are.

1.3.1 Attack at the Underlying System

In this type of attack the attacker can destroy the underlying admin access and tamper all the existing templates thus making the complete authentication obsolete. A unique trademark is introduced by the attacker which would be administered in the the template increasing the acceptance rate.

1.3.2 Attack on the Medium Between the Scanner and the Feature Extractor

In this type of attack the attacker may interfere in the transmission of data from the scanner module to the feature extractor module. Now, the programmer can capture the unique mark and supplant with theirs.

1.3.3 Attack on the Database

In this attack the attacker targets the database storing the public and private keys. If the key lifecycle is not properly maintained then the same key can be used to retrieve all the templates linked with that key.

1.3.4 Attack on the Verification Module

This is the most dangerous attack on a biometric system where the attacker introduces a virus in the system through which the attacker can send commands and produce a false positive giving him/her access to the stored data/template.

1.4 Benefits and Challenges of Cloud Computing

1.4.1 Benefits

Some benefits of Cloud Computing are explained below:

(i) *Diminished Cost:* Cloud registering give office to begin an IT organization with less exertion and beginning expense. Distributed computing administrations are shared by numerous purchasers on the planet. It decreases the expense of administration because of huge number buyers. It charges sum contingent on the use of framework, stage and different administrations; this encourages shoppers to diminish the expense by indicating the specific prerequisites. Organizations can undoubtedly increment or diminishing their interest for administrations as indicated by the presentation of their organization in market.

(ii) *Adaptability and Flexibility:* Cloud figuring can help organizations to begin with a little set up and develop to an enormous condition reasonably quickly, and afterward downsize if essential. Additionally, the adaptability of distributed computing permits organizations to utilize additional assets at top occasions, empowering them to fulfil customer requests. Additionally, distributed computing is prepared to meet any top time necessity by setting up with high-limit workers, stockpiles and so on.

(iii) *Reinforcement and Recovery:* Since all the information is put away in the cloud, backing it up and re-establishing the equivalent is generally a lot simpler than putting away the equivalent on an actual gadget [2]. Likewise, it has numerous strategies to recuperate it from a fiasco; generally proficient and new procedures are embracing by most cloud specialist organizations to meet any sort of calamity. Cloud Providers can get any sort of specialized and other help exceptionally quick than any independently set up associations regardless of their topographical restrictions.

(iv) *Wide organization Access:* Cloud administrations are conveyed through open organization (Internet), it very well may be available whenever anyplace on the planet. These offices can be gotten to by different gadgets, for example, cell phones, workstations, PDAs and so forth with various stages. Purchasers can get to their documents and different applications whenever from anyplace by utilizing their mobiles. This has expanded the pace of embracing distributed computing innovation.

(v) *Multisharing:* Cloud Computing offers administrations by sharing of engineering and different applications over Internet for single and various clients by utilizing virtualization and multi-tenure. With the cloud working in a conveyed and shared mode, different clients and applications can work all the more proficiently with cost decreases by sharing framework.

(vi) *Cooperation:* Major activities or applications are conveying by the exertion of numerous gatherings of individuals cooperating. Distributed computing gives a helpful method to work gathering of individuals together on a typical task or applications in a successful way.

1.4.2 Challenges

Some of the major challenges in Cloud Computing are explained below:

(i). *Protection of information:* Protection of information is key security worry for distributed computing. The greater part of associations feeling more comfort while placing significant information in their site than cloud space. Purchasers don't know concerning area of information, move of date, procedure on cloud, and so forth The greater part of the associations are uninformed of security system actualized by specialist co-ops. Numerous inquiries are emerging by purchasers, for example,

1. Which are the associations sharing administrations.
2. How creation and back-up of records occurring.
3. What befell the erased records?
4. Which kind of buyers can get to information?

(ii) *Privacy of information:* Classification is identified with information security; it guarantees information is obvious to just approved clients. It is exceptionally troublesome because of the virtualization and multi tenure properties that numerous shoppers sharing the equipment, programming all the while in a disseminated network. Secrecy is the duty of specialist co-op. Normal answer for the privacy is encryption. Numerous symmetric and unbalanced calculations are accessible for information privacy, despite the fact that encryption and decoding is the answer for the privacy, there are numerous inquiries are emerging identified with this.

1. Where is encryption and decoding occurring (customer side or cloud side).
2. In what capacity can look through the information in a scrambled structure.
3. What are dangers while moving information from customer to cloud?
4. Any miss utilization of information by specialist organization?
5. Any miss utilization of key by specialist organization?

(iii) *Information Remanence:* Information should be erased from cloud after the life-cycle, or the memory should be reformatted or reused. The reformatting of capacity media doesn't eliminate the recently composed information from the media, yet in addition it tends to be gotten to or recuperated from the media later. No unmistakable standard is accessible for reuse the capacity media. This information remanence makes troublesome the excursion of equipment.

(iv) *Transmission of information:* More often than not information is moving among purchaser and cloud. At first information is sent from customer site to cloud, information is gotten back from cloud to customer after questions during the activity. Encryption is utilized give assurance while the transmission of information. More often than not information is moved without encryption because of part of time is needed for encryption and decoding for every activity upon information.

(v). *Vindictive Insiders:* Pernicious insiders are approved representatives; they delegate for overseeing and keeping up cloud by cloud specialist co-op. These clients now and again take or degenerate the delicate information of associations in the cloud and pass on this delicate data to different associations sharing the same cloud. These pernicious insiders may get instalment for this malignant work. Once in a while administration supplier not ready to make any move against these workers.

2 Literature Survey

This literature survey covers an overview of the observations, outcomes and analysis done by the authors. In this paper a concise outline and correlation of Cryptographic calculations, with an accentuation on Symmetric calculations which should be utilized for Cloud based applications and administrations that require information and connection encryption. However, the paper neglects to suggest the answers for unordered information or different garbage information showing up in the stream [3]. For storing large chunks of biometric data and a system is needed which can handle both ordered and unordered data which is elaborated by the Adaptive Bin Packaging Algorithm. The authors have proposed an efficient storage system for data in the cloud. They have focused on the issues where data at rest and data at transit both are secured using their Adaptive Bin Packaging Algorithm. The data is split into chunks by a splitter module, which protects data from insider threat and the distribution of the chunks across the storage server improves the performance by eventually reducing the time taken to process the chunks which is of lesser size than the original data and therefore it moves to the storage locations in much short time. The proposed algorithm has reduced server load and enhanced performance by a great factor [4]. To overcome cloud security problems the author proposes, a crypto-biometric system the author explains the implementation of fuzzy logic to store the biometric data in AWS platform. Overall implementation gives overcomes certain factors but the author here focuses only on fingerprint data failing to explain the impact on other biometrics and also fails to elaborate the key redundancy issue while storing the templates [2]. An improvement to the former mentioned system, The creator proposed a cloud ID model for rethinking capacity by using matric and bothered terms. To begin with, the accuracy of biometric ID is accomplished. Also, security and effectiveness of the biometric ID conspire is ensured against Level II and Level III assaults [5]. The authors present a watermarking algorithm to protect biometric data. The proposed calculation depends on watermarking, rearranging measure, Hadamard lattice and tumultuous guide. The paper does not specify the storage concerns on adding the watermark which increases per template file size which would increase the overall cost. [6] Contrary to this [7] discusses various IAM policies provided by cloud service providers and the loopholes present currently distinguishing the primary weaknesses in this sort of frameworks and the main dangers [7]. Considering all these above mentioned problems the authors review and discuss the different advantages and significant security difficulties of distributed computing, it likewise features the different cryptographic encryption calculations as the significant arrangement of security challenges [8]. Data Integrity also has been a persistent problem throughout all Cloud platforms, the author has introduced a check measure which lessens the danger of data theft and diminishes the load on resources used. After this the data is encrypted using the Circle Cipher and Pearson's equation [9]. From this literature survey, it can be observed that major concerns when it comes to storing and encrypting the biometric templates are the storage cost and key management.

3 Problems in Existing System

The existing systems have shown great results but their work comes with some flaws which need to be addressed. Some problems persisting in current cloud security systems are first storage complexity, every organisation aims to minimise their cost of resources and achieve maximum output. It has been seen that the existing algorithms all increase the file size thus leading to a larger cost. Secondly current cloud platforms all have their own internal algorithms to encrypt their data as all these platforms have different architecture so there is no generic guideline to be followed causing problems in cross platform applications. Focusing on biometrics unlike passwords, biometric traits cannot be cancelled and reissued (i.e., if a user's fingerprint is compromised it cannot be changed and the user cannot use it in the future for authentication purposes) so environmental conditions and inevitable situations like aging, injury, etc. are not dealt with in unimodal systems. As these issues are not deal no failure compliance SLA is specified by cloud providers thus compromising the reliability of data. Also, misconfigurations of cloud security settings are a main source of cloud information penetrates. Numerous associations' cloud security acts the board procedures are insufficient for ensuring their cloud-based foundation. Unimodal systems currently lack individuality which may result in false acceptance and increasing FAR(False Acceptance Rate). Apart from biometric, current algorithms also face a serious key management issues due to above mentioned difference in architecture and models like IaaS, PaaS, SaaS.

4 Hybrid Model

4.1 *Architecture*

For demonstration of the algorithm a desktop application interface is developed using C# and .Net which runs on the client machine. The application has features constituting user login, user registration, biometric capture and verification divided into modules. Figure 1 shows a layout of the architecture used to test the encryption algorithm.

To protect the data in transit and when stationary the model uses both client-side and server-side encryption. A cloud server running on the three cloud platforms AWS, Google Cloud, Azure each running their Microsoft SQL server for storing the biometric templates and user data. The overall architecture is divided into 3 modules.

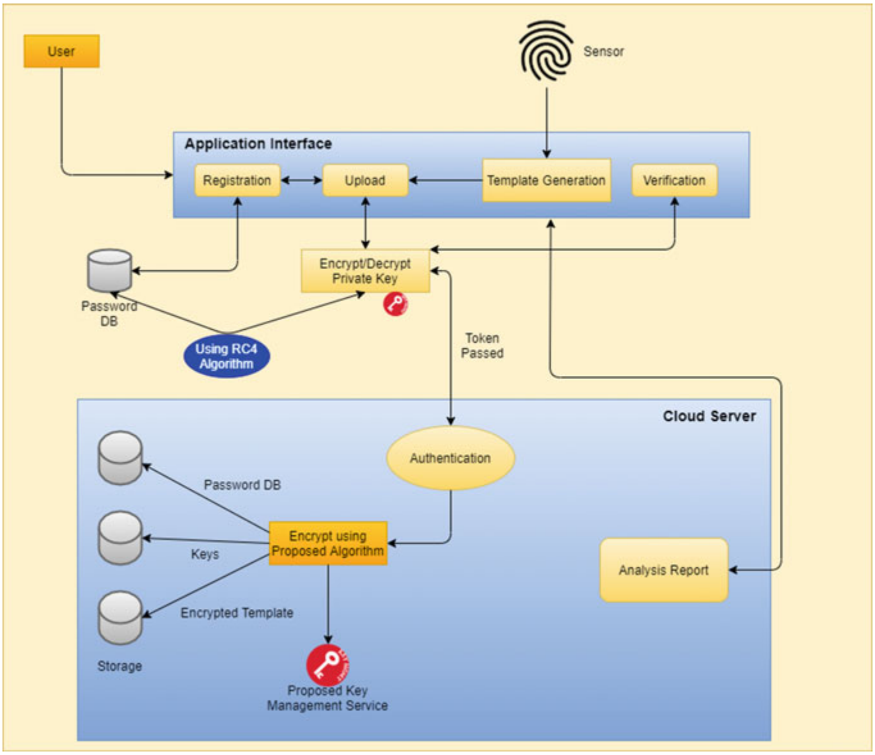


Fig. 1 Hybrid model architecture

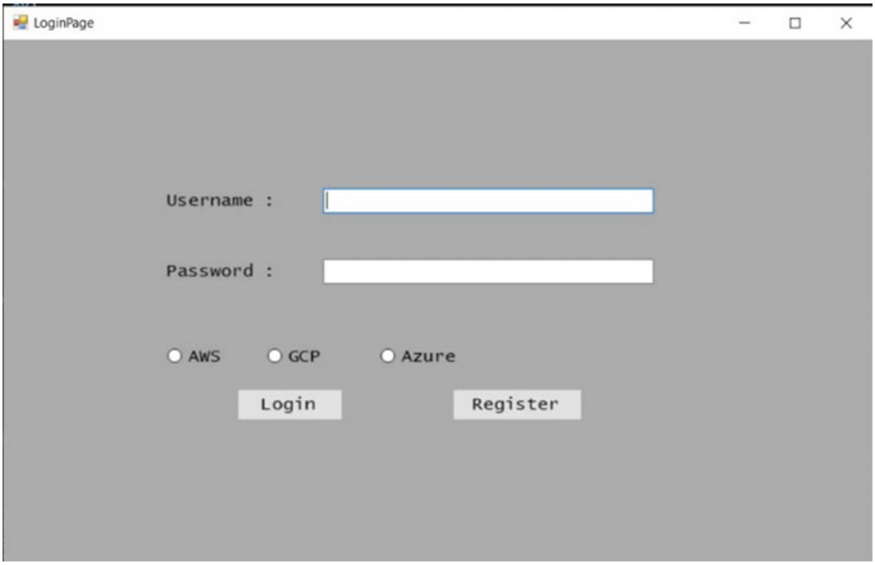


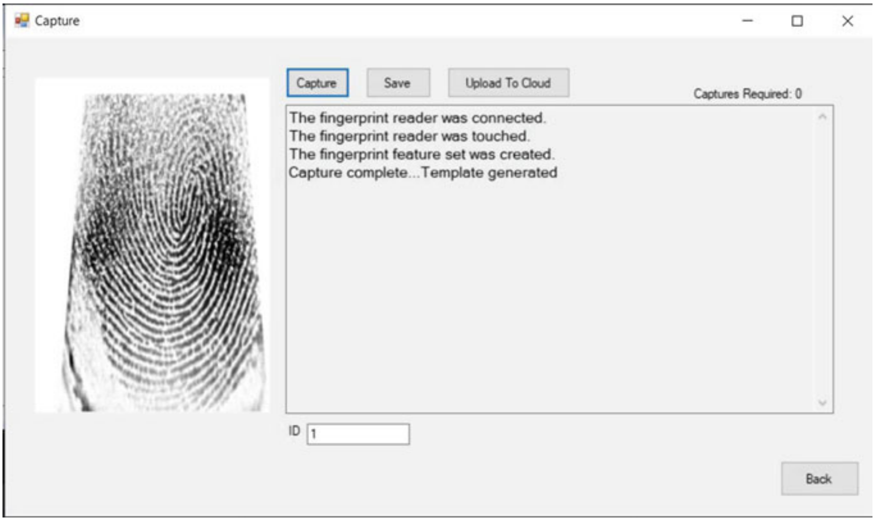
Fig. 2 Login module

4.1.1 Login/Registration Module

This login and registration module is the initial start point of the application. In this window the user has to enter his/her registered username and password. Apart from this the user must also select which cloud service they want to access and also select the biometric which they want to enroll or verify. Fingerprint and Facial biometric data is captured and stored for testing of the algorithm. The system runs an SQL query to the specified cloud server authenticating the user. An important feature here is the client-side encryption of the user password along with the secret key. Former mentioned items are encrypted using the stream cipher RC4. RC4 though being an easily hackable algorithm is much effective on transit data. Figure 2 shows the initial landing window when the application is started.

4.1.2 Capture and Upload Module

Once the user is logged into the cloud server the user has options to either enroll, verify, save, upload biometric templates. The user can either upload an existing. fpt (raw data) file or a bitmap file or they can capture their biometric live using the sensor. During Upload and Verification several metrics such as False Acceptance Rate, Encryption/Decryption Time, Processing Speed as displayed in the console for analysis. Figure 3 shows capture of fingerprint using a live sensor.



The fingerprint reader was connected.
The fingerprint reader was touched.
The fingerprint feature set was created.
Capture complete...Template generated

ID: 1

Fig. 3 Fingerprint capture

4.1.3 Backend Encryption Module

This module basically runs on the cloud server performing encryption/decryption of the biometric template and key management. Unlike other key management systems for example, AWS KMS which is the key management service offered by Amazon Web Services provides server-side encryption of the key only and the service is not offered with AWS VPCs so in private subnets the keys are prone to attack. Similarly, Google's Customer managed encryption keys (CMEK) has limitations, which allows users to only display and manage up to 1000 keys, and the audit logs of the keys are very complex to understand, making it not easy to read by humans. This module only comes into play after template upload to the cloud is complete or template retrieval is requested for verification.

4.2 Supplementary Material

The encryption algorithm uses some permutation and compression matrix which are mentioned in Table 1.

Table 1 Matrix description

P	Permutation Matrix 1	8 × 16 = 128 bit
P^{-1}	Inverse Permutation Matrix 1	8 × 16 = 128 bit
P1	Permutation Matrix 2	4 × 8 = 32 bit
$P1^{-1}$	Inverse Permutation Matrix 2	4 × 8 = 32 bit
CT	Compression Table	4 × 8 = 32 bit

4.3 Algorithm

The proposed hybrid encryption algorithm incorporates some features of block cipher algorithms like DES, AES and Blowfish. Some salient features of this algorithm are it encrypts 64-bit block at a time, the algorithm uses 2 keys one private 128-bit intermediate key and one 64-bit subkey for encryption [10]. Each block goes through 8 rounds of encryption. Figure 4 shows brief flow of the algorithm.

- **Algorithm for Secret Key Generation**

  1. Accept 32-bit password and private key from user

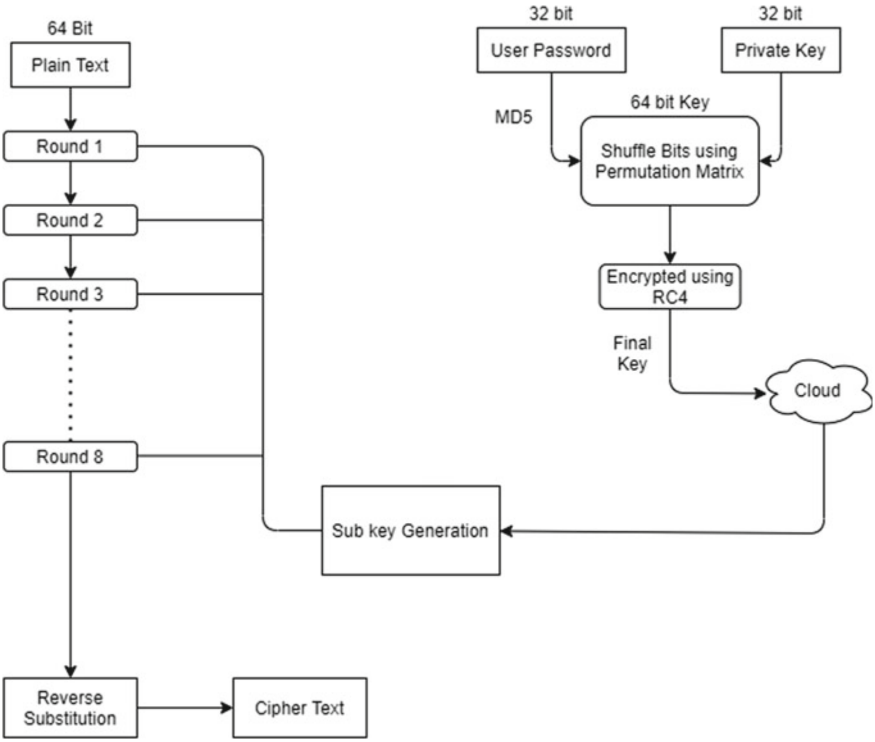


Fig. 4 Encryption algorithm overview flowchart

2. Hash the user password using MD5 hashing technique
3. Combine password and private to generate 64-bit text
4. Shuffle the binary bits of the text using permutation matrix P.
5. Encrypt the 64-bit text using RC4 Algorithm to generate the secret key

- Algorithm for Subkey Generation

$$M = \left(\sum \text{ASCII } * \text{position}\right)\%(\text{length of the template})$$

Each round's subkey is generated using the main private key and the input plaintext itself making it difficult to crack. A value M is generated depending on the ASCII values of the plaintext. This M is added to the plaintext and then the bits are permuted and reversed finally. The final result here generates a 128-bit Key. Still this key is an intermediate key and will not be used entirely for encryption.

- Algorithm for Round Encryption

1. Split 64-bit plaintext block into two 32-bit blocks P_L & P_R, similarly split the 128-bit secret key into 64-bit blocks S_L & S_R.
2. If MSB of plaintext = 1 then round subkey = S_L
3. Else round subkey = S_R
4. Truncate the round subkey for 64-bit to 32-bit using compression table matrix CT
5. Perform XOR of P_R and round subkey
6. Shift bits of P_R to left by 1 bit
7. Perform XOR of P_R and P_L
8. Permutate bits of P_L using P1 matrix
9. Inverse P_L and P_R for next round (Fig. 5).

4.4 Experiment Results and Analysis

The biometric fingerprint or facial data which are captured or uploaded by the user are in raw database format or bitmap format. The data in these files is converted to Base64 encoding and send to the cloud for encryption. Upon testing it was noted that the biometric template files were 2-60 KB in size for fingerprint and facial images were 1–1.5 MB in size.

On converting the raw data into Base64 we get a string of integers. Table 2 shows the workflow of 8 bytes of data going through 8 rounds of encryption. The algorithm incorporates ASCII range of non printable characters also making it impossible to read and find the pattern. Also, after each round depending on the plaintext the subkey is chosen from the secret key so it becomes impossible to predict the next round subkey. Figure 6 shows the input plaintext on the left side, where the private key entered by the user is "major" using that the plaintext is encrypted and the ciphertext in Base64 is displayed on the box to the right.

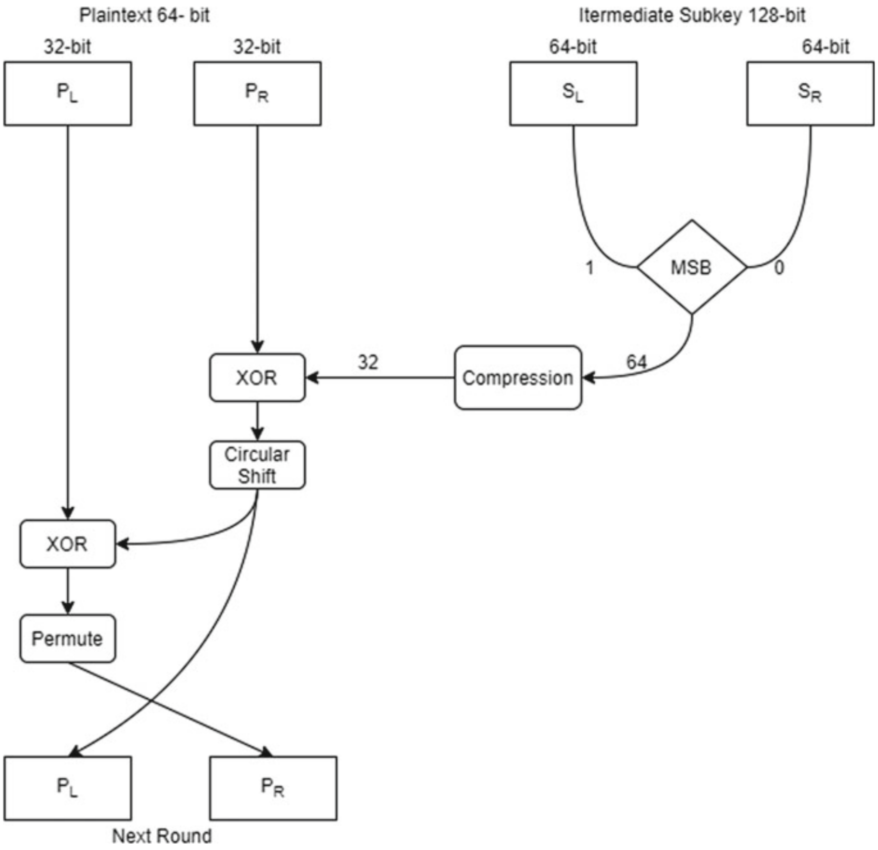
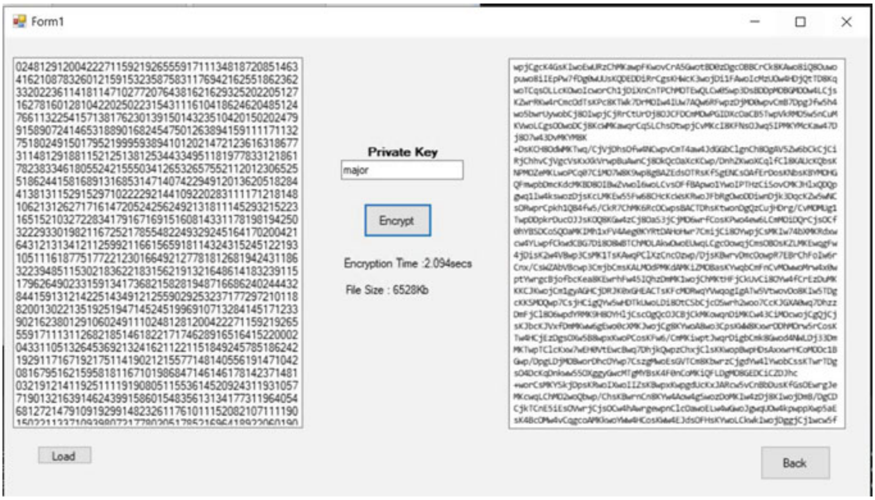


Fig. 5 Round algorithm

Table 2 Encryption process

Round	Plaintext	Subkey	Cipher Text
1	02,347,681	CˆA]|NwsGFk	E0!J«
2	E0!J«	CˆA]|NwsGFk	ioÑ◆b!!⌂
3	ioÑ◆b!!⌂	wsGFkˆ\\{} < |	îĐĐë♀M%
4	îĐĐë♀M%	CˆA]|NwsGFk	ýâl↓B8î
5	ýâl↓B8î	wsGFkˆ\\{} < |	5~t ×
6	5 ~ t ×	wsGFkˆ\\{} < |	Âíïü¿d§Ô
7	Âíïü¿d§Ô	wsGFkˆ\\{} < |	U3^{1}ûùY5▲
8	U3^{1}ûùY5▲	CˆA]|NwsGFk	ÙHoÀ



```
024812912004222711592192655591711134818720851463
416210878326012159153235875831176942162551862362
332022361141811471027720764381621629325202205127
162781601281042202502231543111610418624620485124
766113225415713817623013915014323510420150202479
915890724146531889016824547501263894159111171132
751802491501795219995938941012021472123616318677
311481291881152125138125344334951181977833121861
782383346180552421555034126532657552112012306525
518624415816891316853147140742294912013620518284
413813115291529710222292144109220283111171218148
106213126271716147205242562492131811145293215223
165152103272283417916716915160814331178198194250
322293301982116725217855482249329245164170200421
643121313412112599211661565918114324315245122193
105111618775177221230166492127781812681942431186
322394851153021836221831562191321648614183239115
179626490233159134173682158281948716686240244432
844159131214225143491212559029253237177297210118
820013022135192519471452451996910713284145171233
902162380129106024911102481281200422271159219265
559171113112682185146182217174628916516415220002
043311051326453692132416211221151849245785186242
192911716719217511419021215577148140556191471042
081679516215958181167101986847146146178142371481
032191214119251111919080511553614520924311931057
719013216391462439915860154835613134177311964054
681272147910919299148232611761011152082107111190
150221133710939807217790205178521696418922060190
```

Private Key

major

Encrypt

Encryption Time :2.094secs

File Size : 6528Kb

```
wpjCgcK4GsKIwoEwURzChMKawpFKwovCnA5GwotBDKzDgcOBBCrCk8KAwo81Q8Ouwo
puwo81IEpPw7fDg8wUUsKQDEDDiRrCgsKHwcK3wojDi1FAwoIcMzUOw4HDjQtTD8Kq
woTCqsOLLcKOwoIcworCh1jDiXnCnTPChMOTEwQLCw8Swp3DsBDDpMOBGMODw4LCjs
K2wrRKw4rCmcOdTsKPc8KTWk7DrMODIw4IUw7AQw6RFwpzDjMOKwpvCmB7DpgJfw5h4
wo5bwrUywobCj8O1wpjCjRrCtUrOj8OJCFDCmMOwPGIDKcOaCB5TwpVkRMO5w5nCuM
KVwoLCgsOOwoDCj8KcwMKawqrCq5LChsOtwpjCvMKc1BKFNsOJwq5IPMKYMcKaw47D
j8O7w43DvMKYM8K
+DsKCH8OdWMKTwq/CjVjDhsOfw4NCwpvCmT4aw4JdGGbClgnCh8OgAV5Zw6bCkCjCi
RjChhvCjVgcVsKxXkVrwpBuAwnCj8OkQcOaXcKCwp/Dnh2KwoXCq1fCl8KAUcKQbsK
NPMOZeMKLwoPCq07CiMO7w8K9wp8gBAZEdsOTRsKfSgENCsOAfErDosKNbsK8YMOHG
QFmwpbOmcKdcMKBDBO1Bw2vwol6woLCvsOFfBApwo1YwoIPTHzCiSovCMKJHlxQDQp
gwq1lw4kswozDjsKcLMKEw55Fw68CHcKcwsKRwoJFbRgOwoODiwnDjk3DqcK2w5wNC
sORwprCpkh1QB4fw5/CkR7ChMK6RcOCwpsBACTDhsKtwonDgQzCujHDrg/CvMOMUg1
TwpODpkrDucOJJsKOQ8KGw4zCj8Oa53jCjMO6wrfCosKPwo4ew6LCmMOiDQrCjsOCf
0hYBSOCo5QOaMKIMh1xFV4AwgBKYRtDAHoHwr7CmijCi8OYwpjCsMKIw74bXMKRdxw
cw4YLwpfCkwdCBG7Di8OBwBTChMOLAkwOwoEUwqLCgcOowqjCmsOBOsKZLMKEwqgFw
4jDisK2w4VBwp3CsMK1TsKAwqPClXzCncOzwp/DjsKBwrvOmcOowpR7LBrChFoIw6r
Cnx/CsWZAbVBcwp3CmjbCmsKALMOdPMKdAMKiZMOBasKYwqbCmFnCvMOwwoMrw4xdw
ptYwrgcBjofbcKea8KEwrhFw45lQhzDmMK1wojChMKtHFjCkUvCi8OYw4fCrLzDuMK
KKCJKwojCm1gyAGHCjDRJK8xGHEACTsKFcMORwqYVwqog1gATw5VtwovOo8KIw5TDg
cKKSMODQwp7CsjHCigQYw5wHDTkUwoLDi8OtCSbCjcO5wrh2woo7CcKJGXAWwq7Dhzz
DmFjCl8O6wpdYRMK9H8OYH1jCscOgQcOJCBjCkMKowqnDiMKCw43CiMOcwojCgQjCj
sKJbcKJVxfDmMKwwSgEwo8cXMKJwojCg8KYwoA8wo3CpsKGw8KxwrODhMOrw5rCosK
Tw4HCjEzDgsOXw5B8wpxKwoPCosKFw6/CmMKiwptJwqrOigbCmk8Gwod4NwLDj33Dm
MKTwpTClcKxw7wEHKVtEwcBwq7DhjkQwpzChxjClsKKwop8wpHDsAxowrHCoMODc1B
Gwp/DpgLDjMOBwor0hcOYwp7CszgMwoEsGVTCm8KbwrzCjgdYw4lYwobCssKTwrTDg
sO4DcKqDnkow55OXggyGwcMTgMYBsK4F0nCoMKiQFLDgMOBGEDCiCZDJhc
+worCsMKYSkjDpsKRwoIXwoIIZsKBwpxKwpgdUcKxJARcw5vCnBbOusKfGsOEwrgJe
MKcwqLChMO2woQbwp/ChsKBwrnCn8KYw4Acw4g5wozDoMKIw4zDj8KIwojDm8/DgCD
CjkTCnE5iEsOVwrjCjsOCw4hAwrgewpnClcOawoELw4wGwoJgwqUOw4kpwppXwp5aE
sK4BcOMw4vCqgcoAMKIwoYWw4HCosKWw4LJdsOFHsKYwoLCkwkIwojDggjCj1wcw5f
```

Fig. 6 Sample encryption analysis

Plaintext: 02347681

***Secret Key**:* C^A]lNwsGFk^\\{}<l (128-bit)

The above algorithm takes 3.39 × 10^{-5} s giving a processing time of 5.2 × 10^{-7} s/bit. Let us a take sample file to measure the encryption time and run it through different algorithms,on comparing the same template file with existing algorithms Table 3 shows resulting output file sizes and their respective encryption time.

Compared to currently adapted AES algorithm it is slower due to additional time required to calculate the subkey at the start of each round which in the case of AES the subkeys are generated for each block at start itself, but it solves a major problem persisting today that is storage. One advantage of the proposed algorithm is that the input and output file size remain the same, saving organisations and individual users a lot of storage cost in cloud. Also, the proposed algorithm adds a complex mapping nature, the permutations used are designed in such a way that most of the encrypted characters are non-printable characters, only understandable by the computer.

Table 3 Comparison with existing algorithms

Algorithm	Input file size (KB)	Time (sec)	Output file size (KB)
Hybrid	51	14.1	51
AES	51	8.6	180
DES	51	19.8	130

5 Conclusion and Future Work

We used the hybrid algorithm for both fingerprint and facial biometric templates and based on results of Table 3 we concluded that the hybrid algorithm might lack speed but solves the problem for storage and also increases complexity making it more secure as compared to the existing algorithms. The most important feature of this algorithm is the generation of subkey which is completely dependent on the input of each round making it impossible to predict the intermediate keys. This algorithm being generic to any cloud platform can help increase efficiency of application and act as an internal security layer for them. On thorough comparison with currently used algorithms such as AES and DES it is found that the algorithm is faster than and more secure than DES while it is slower than AES but adds an internal secure layer for protection of data which AES does. For further work the algorithm can be optimised using Galios Counter Modes which help ensure data integrity all round which may solve existing data loss problem in AES. The encryption algorithm can be implemented to various other biometric data such as iris patterns, vein patterns, voice print waves, etc.

References

1. U. Gawande, Y. Golhar, Biometric security system: a rigorous review of unimodal and multimodal biometrics techniques. Int. J. Biometrics. (2018)
2. D. Gonzalez, E. Rua, Secure crypto-biometric system for cloud computing. Int. J. Sci. Technol. (2011)
3. A. Bhardwaja, G.V.B. Subrahmanyamb, V. Avasthic, H. Sastry, Security algorithms issues for cloud computing, in *IEEE*
4. I. Mohiuddin, A. Almogren, M. Al Qurishi, M.M. Hassan, I. Al Rassan, G. Fortino, Secure distributed adaptive bin packing algorithm for cloud storage. Future Gener. Comput. Syst. (2018)
5. O. Nafea, S. Ghouzali, W. Abdul, *Hybrid Multi-Biometric Template Protection Using Watermarking*, Oxford University Press
6. F. Alsolami, B. Alzahrani, Cloud-ID-Screen: secure fingerprint data in the cloud, in *IEEE*
7. K. Hashizume, D.G. Rosado, E. Fernández-Medina, E.B. Fernandez, Analysis of data compliance and storage policies in cloud. J. Internet Serv. Appl.
8. K.V. Nasarul Islam, K.V Mohamed Riyas, Analysis of various encryption algorithms in cloud computing. Int. J. Comput. Sci. Mobile Comput. (2012)
9. S. Gawade, A. Bharti, A. Raj, S. Madane, Biometric authentication using software as a service in cloud computing. Int. J. Eng. Comput. Sci.
10. A. Zli, M.S. Hossain, G. Muhammad, I. Ullah, H. Abachi, A. Alamri, Edge-centric multimodal authentication system using encrypted biometric templates. Future Gener. Comput. Syst. (2018)

Zero-Trust Security Implementation Using SDP over VPN

D. Abhiram, R. Harish, and K. Praveen

Abstract VPNs are widely used among organizations for securing their internal networks. This research demonstrates implementing the VPN environment and possible attack vectors of HTTP traffic over VPN connections. We also acknowledged the importance of software-defined perimeter or SDP by implementing it at a very basic level and mitigating the attacks that could be performed in a VPN environment. This research also focuses on the vulnerable area found in between the VPN client to the web server, where implementation of the MITM attack and cryptojacking using CoinIMP API service was successful. The architecture of software-defined perimeter based on zero-trust security model was studied. The client, controller, and gateway operations which are the three main important modules inside the SDP architecture were implemented using python script into three different virtual machines, respectively. Also tried the same attack on SDP environment and proved that software-defined perimeter is resilient to such attacks. SDP is capable of hiding the servers containing sensitive information from unauthorized users. SDP helps to overcome the MITM attacks and cryptojacking. Theoretical and practical implications of the results are discussed.

Keywords Virtual private networks · MITM attack · CoinIMP · Cryptojacking · Zero-trust security · Software-defined perimeter · Dynamic firewalls

D. Abhiram (✉) · R. Harish · K. Praveen
TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2cys19005@cb.students.amrita.edu

R. Harish
e-mail: r_harish@cb.students.amrita.edu

K. Praveen
e-mail: k_praveen@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_22

1 Introduction

Technology is evolving day by day in our daily lives in the form of mobile phones, Websites, Internet of things, cloud computing, and many more. Undoubtedly, they offer much comfort in accomplishing the tasks with ease. But, do they implement enough security strategies to protect the networks from attacks? Will the security measures implemented currently be enough to avoid security threats? VPN's was opted as a security solution by many organizations. VPNs provide security, privacy, and anonymity by encrypting the connections, creating private networks from public ones, and masking the IP addresses. VPNs provide these security advantages until the VPN server since the encryption and decryption, ends and starts here, respectively. If a web server is protected with VPN, then the encryption of HTTP requests and responses ends at the VPN server. If the request is HTTP instead of HTTPS, then it should travel until the webserver in plain text since they are not encrypted. Many organizations implement VPNs to protect their internal networks and servers to prevent attacks. With the help of this, cryptojacking [7] has been implemented by intercepting the traffic coming out of the VPN server using the mitmproxy tool and HTTP response were altered by injecting the malicious Javascript code which is capable of performing cryptojacking. The SDP [1] is a mitigation strategy that has been proposed to avoid these attacks. SDP is a security model that is responsible for hiding valuable resources from networks and attackers. Access is given only to those who gain trust by undergoing a series of validation and authentication processes. SDP framework is implemented and showcased how only an authorized host can access the webserver. This in turn proves that SDP is resilient to the attacks that were discussed before. The main aim of this research is to highlight the importance of SDP and brief the security advantages of its implementation. In Sect. 3, the VPN implementation and the proposed attack model were discussed. In Sect. 4, an overview of cryptojacking and the results of CPU usage was discussed. In Sects. 5 and 6, the concept of zero-trust security [8] and SDP along with its architecture was explained. In the later section, the implementation and results were mentioned.

2 Related Works

In the previous research, cryptojacking was performed which is a browser-based hijacking technique on anonymous proxies and tor network exit nodes without the user's knowledge [6]. Also, the effect of cryptojacking on a large-scale environment was analyzed. For intercepting HTTP traffic, the Mitmproxy tool was used and for performing the cryptojacking, Coinhive was used which is now obsolete. In the other research, an overview of cyberthreats and the traditional approach to resolving the security concerns was provided [3]. And also to overcome the security drawbacks, an advanced security model known as software-defined perimeter was introduced. Major differences between the traditional security approach and the software-defined perimeter were also highlighted [4]. The generation of simple and dynamic perimeter along with the authentication process involved in dynamic perimeter was also discussed. Several previous types of research have focused on the concept, architecture, and implementation of SDP as well as the performance analysis of SDP during DoS attack, port scanning attack [1] and how it is resilient to such attacks. During the performance evaluation, the connection setup time, as well as the network throughput, was tested in different scenarios. The main layers of security protocols involved in SDP like single-packet authentication, mutual transport layer security, device validation, dynamic firewalls, and application binding were also mentioned. The main responsibilities of the client (which is also called as initiating host since it initiates the HTTP request), controller, and gateway (which is also called as accepting host since it accepts the HTTP request and sends HTTP response), which plays a key role in the functionality of software-defined perimeter's architecture, were briefed [2]. Open challenges and potential applications of SDP in core networks, mobile networks, and internal enterprise networks were discussed. Some recent articles related to deep learning on domain generation algorithm [16] and anomaly detection techniques using pyshark [17] were referred for machine learning approach strategies to pave a path for future work.

3 Cryptojacking

Cryptomining is a process that requires a group or network of interconnected miners who are responsible for solving the complex mathematical problem through which the blocks are linked and integrity is maintained during the transactions [7]. In exchange, the miners get rewarded with mined cryptocurrency. To solve complex mathematical problems, there is a need for huge CPU cores which are responsible for doing the calculations. Since cryptocurrencies have higher values, some malicious miners tend to hijack resources for mining. This kind of cryptomining is a threat, since hijacking the resources is considered a illegal act and it has grown drastically over the recent years. Usually, cryptomining script is written in Javascript and it is embedded in HTTP [10] responses. In this case, coinIMP is used which is an effective Javascript miner that can be easily embedded in HTTP responses. The impact of cryptojacking [18] may lead to performance-related issues such as device slowdown, overheating batteries, devices becoming unstable which in turn leads corporate networks to shut down for facilitating a cleanup. This will also increase the costs due to heavy electricity usage [6]. If this is performed on cloud instances, then additional financial implications may occur if the organizations opted billing based on CPU usage.

4 Experimental Setup

The idea behind implementing a VPN from scratch using python involving three different virtual machines is to perform cryptojacking by intercepting the traffic coming from the webserver using mitmproxy and injecting the coinIMP [15] API in the response so that the VPN client's CPU cores are used to the fullest. For implementing the VPN, three virtual machines with Ubuntu (preferably) as operating systems are required. The initial virtual machine setup requires the VPN server to have two adapters in which the first adapter is attached to the NAT network and the second adapter is attached to the internal network. The web server will have a single adapter attached to the internal network, and the IP address will be manually given since the webserver should be from an external network. Now by default, the webserver is unreachable to the VPN client because they both are not from the same network. A new virtual interface which in this case is TUN/TAP interface needs to be created using the python script. This is achieved by constructing a structure that contains the name and flags which are 16 bytes and 2 bytes long, respectively. This structure along with a hexadecimal command is sent to a device which in this case is /dev/net/tun to create the virtual interface. This virtual interface is responsible for sending the packets to the application (Fig. 1).

To redirect the packets to the virtual interface routing table can be used. This whole VPN setup is taken reference from the seed labs [11, 12]. Once the VPN tunnel is implemented, the VPN client will now be able to send the HTTP requests to

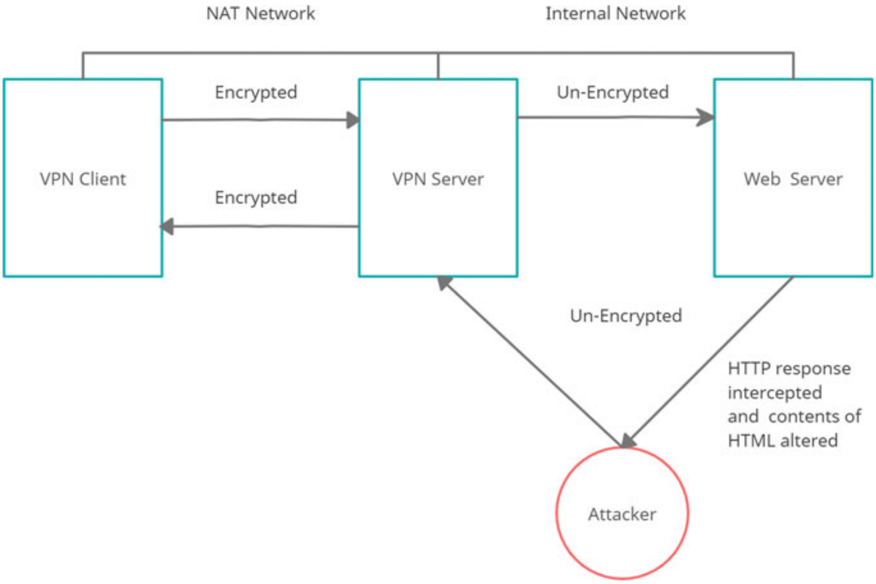
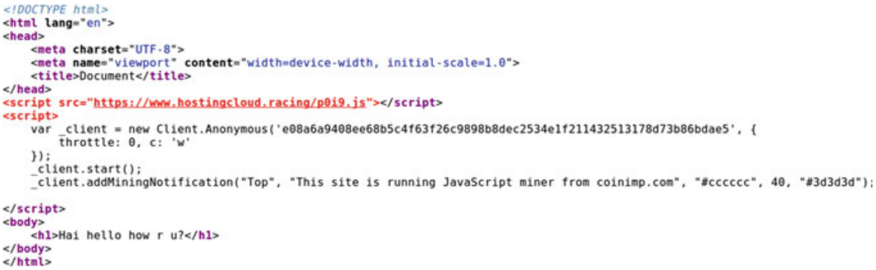
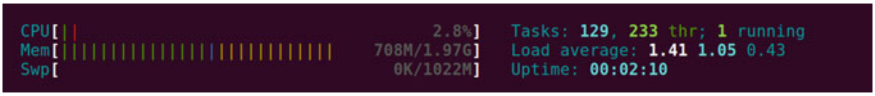


Fig. 1 Experimental setup



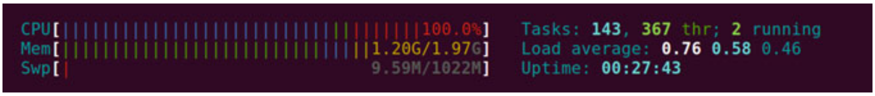
```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Document</title>
</head>
<script src="https://www.hostingcloud.racing/p0i9.js"></script>
<script>
    var _client = new Client.Anonymous('e08a6a9408ee68b5c4f63f26c9898b8dec2534e1f211432513178d73b86bdae5', {
        throttle: 0, c: 'w'
    });
    _client.start();
    _client.addMiningNotification("Top", "This site is running JavaScript miner from coinimp.com", "#cccccc", 40, "#3d3d3d");

</script>
<body>
    <h1>Hai hello how r u?</h1>
</body>
</html>
```

Fig. 2 CoinIMP script injected in the HTTP response



```
CPU[||                              2.8%]   Tasks: 129, 233 thr; 1 running
Mem[|||||||||||||||||||||||||||   708M/1.97G]   Load average: 1.41 1.05 0.43
Swp[                              0K/1022M]   Uptime: 00:02:10
```

Fig. 3 CPU usage before HTTP response execution in client browser



```
CPU[||||||||||||||||||||||||||||||||||||100.0%]   Tasks: 143, 367 thr; 2 running
Mem[||||||||||||||||||||||||||||||||1.20G/1.97G]   Load average: 0.76 0.58 0.46
Swp[|                               9.59M/1022M]   Uptime: 00:27:43
```

Fig. 4 CPU usage after HTTP response execution in client browser

the web server and based on the request, the webserver will send back the response to the VPN client through the VPN server. Once the VPN client sends the HTTP request to the web server, the response is sent back in plain text to the VPN server. Now, in the case of VPN setup, the attacker intercepts the HTTP response through the mitmproxy tool and injects the coinIMP API as shown in Fig. 2. CoinIMP is the new effective JavaScript miner that can be embedded into the Website. After successfully creating an account in coinIMP, in the dashboard, a site-key can be found which is appended in the script which is a 64-bit hexadecimal value. The throttle value can be set to zero for 100% CPU usage. The VPN server is configured with mitmproxy which is an open-source tool.

This tool is capable of performing all types of man-in-the-middle attacks. Mitmproxy can eavesdrop on all the incoming and outgoing HTTP traffic. This tool is also capable of modifying HTTP traffic, and in this case, the interception is done manually. In the script that is injected, there is a feasibility of throttling the maximum CPU usage limit, and in this case, it is set to 100% and the CPU usage history is shown in Fig. 5. If the mining is done anonymously, then it is an undetectable cryptojacking attack. To achieve this, the processing speed of the CPU during the mining process can be throttled. If this is done, then the possibility of detection rate is highly decreased.

As already mentioned, the throttle speed can be controlled inside the script, it is very convenient for the attacker to change the value dynamically during the interception of the HTTP response. Once the script is injected, the attacker now

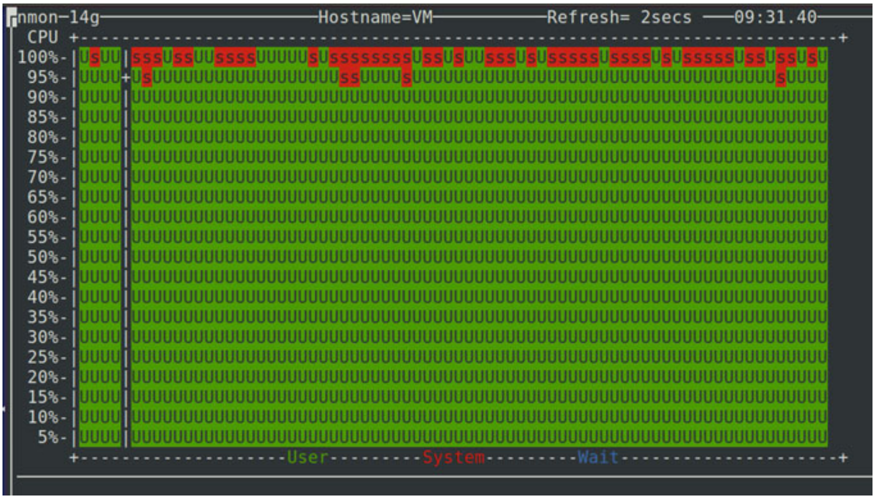


Fig. 5 CPU usage history after the attack

sends the altered HTTP response back to the VPN server. The VPN server now encrypts the response and sends it to the VPN client. Now the client decrypts the response, and the browser executes it which will ultimately lead to 100% CPU usage as shown in Fig. 4 as compared to the initial CPU usage which was just 2.8% as shown in Fig. 3 (Fig. 5).

5 Zero-Trust Security

To avoid cryptojacking and MITM attacks, there is a need for a strict security model which is zero-trust security [8]. Business migration to the cloud and other advanced technologies pose a variety of security threats. Zero-trust security is a concept that enforces organizations to strictly verify the devices, users, or anything that tries to connect the network. It is an optimal solution for defense mechanisms against many security threats. In this security model, every single connection attempt is verified and resources are completely hidden from the network until trust is established. In a nutshell, zero-trust security means trust no one, verify everything. There are two main elements to verify for gaining trust which is users and devices [1]. Users are verified through authentication methods like PKI, OpenID, SAML, and devices are verified through user behavior analytics and monitoring tools like Splunk and Microsoft defender. The user behavior analytics can be detected in zero-trust security by always verifying the users, packets, and connections first before gaining the trust and giving the access.

6 Software-Defined Perimeter

Software-defined perimeter [5] is a security concept that is based on the zero-trust security model which has some concepts related to SDN [9]. To avoid the threats that have been highlighted before in the form of cryptojacking can be mitigated with the help of software-defined perimeter. SDP was proposed by the cloud security alliance which is having the ability to protect the networks dynamically. The SDP architecture [13, 14] provides access to the client only if it can verify and authenticate the client's identity. The SDP architecture is dependent on single packet authentication which is a security layer and is responsible for authenticating the device. It is dependent on mutual transport layer security (mTLS) which is responsible for maintaining confidential communication over the networks. SDP is also dependent on dynamic firewalls which are responsible for hiding the resources behind the gateway and only allow them once authenticated. The SDP architecture consists of three main components which are the SDP controller, initiating hosts, and accepting hosts (Fig. 6).

The main responsibilities of the core components will be discussed. The controller is responsible for acting as a trusted broker between the client and the gateway. This includes storing and validating the details of the client and sending it to the gateway and determining the services that each client is authorized to access. It is also responsible for creating the mTLS tunnel. The client's main responsibility is to request the controller for accessing the service hidden behind the gateway. It also sends the details about the device which is one of the processes of authentication. Gateway's main duty is to implement a firewall rule which will by default reject all the incoming packets from all the hosts except the controller. Once the controller is successful in authenticating the client with positive feedback, then it proceeds in sending the details about the client to the gateway. Now the gateway will verify the details received and updates the firewall rule to give access to the service that the

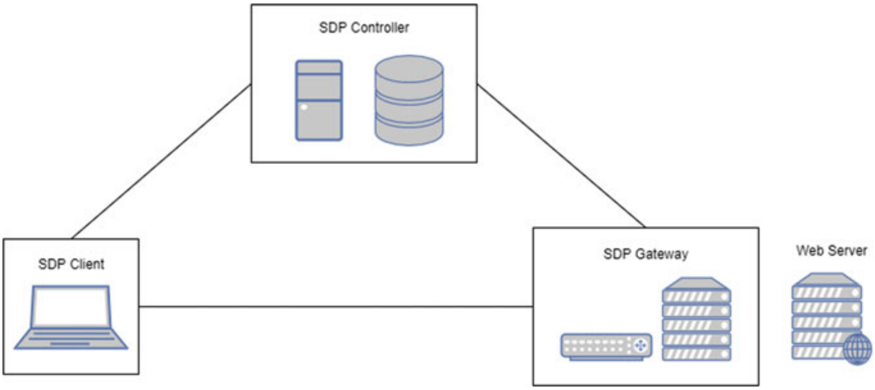
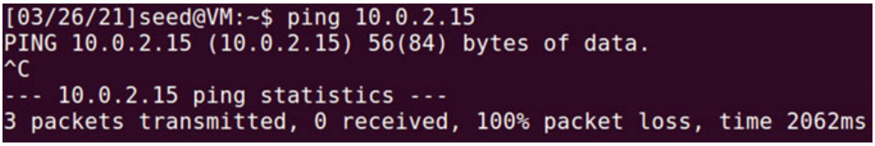


Fig. 6 SDP architecture with Web server hidden behind the gateway



```
[03/26/21]seed@VM:~$ ping 10.0.2.15
PING 10.0.2.15 (10.0.2.15) 56(84) bytes of data.
^C
--- 10.0.2.15 ping statistics ---
3 packets transmitted, 0 received, 100% packet loss, time 2062ms
```

Fig. 7 Unauthorized hosts before validation

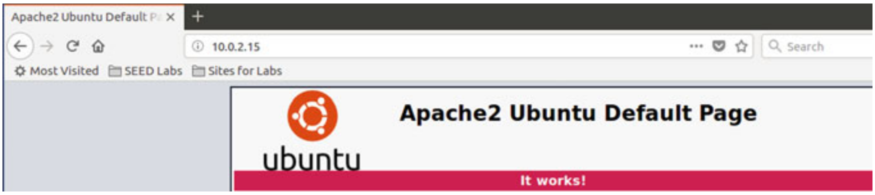
client had requested. Iptables can be used to accomplish this. The steps involved in the process of SDP are briefed below [1]. The gateway will begin a TLS connection to the controller and sends a SPA packet. Then the controller decrypts it and cross-checks the payload with the data present inside the database. If it is successful, then an mTLS tunnel is established between them. The gateway then receives the list of authorized clients and services from the controller. Similarly, the client begins a TLS connection to the controller and sends the SPA packet. The controller then decrypts it and cross-checks the payload with the data present inside the database. If successful then an mTLS tunnel is established. There is no much operational difference between client TLS connection and gateway TLS connection. The client repeats the same process with the gateway, and here the client can send the SPA packet to the gateway because it is listed in the authorized client's table sent by the controller. Once the gateway validates the request, it then proceeds in giving access to the client's request. Now the client will be able to view the web page that has been requested from the webserver.

7 Implementation and Results

In the implementation of SDP, three virtual machines are used in which the first VM is for the SDP client, the second VM is for the SDP controller, and the last VM is for both the SDP gateway and the webserver. By default, the gateway rejects all the connections except the controller. The client sends the details about itself as well as the service it needs access for, to the controller. Now the controller stores the details in the database and sends the same to the gateway. Now the client requests the gateway with the service and gets access to it as the firewall rules get updated in the gateway. In this experimental setup, IP address is given manually, but fetching IP address dynamically can also be done. In Fig. 7, it can be seen that the client initially couldn't even ping the gateway since it is not yet validated by the controller.

Now it is understandable that the client could not access the web page as the gateway rejects all the connections if not authenticated. So, the client now needs to undergo a validation process from the controller to gain access. Figure 8 shows the client getting access to the webserver after the authentication is successful.

There are many advantages of implementing SDP over VPN as the resources can be completely hidden from the network thus protecting it from attackers. The access is only given to the hosts who successfully undergo the verification process which



Apache2 Ubuntu Default Page

It works!

Fig. 8 Authorized hosts after validation

involves the devices and users to authenticate. This way the security of the hidden resources is enforced, and the unavailability of access to resources for unauthorized clients is confirmed.

8 Conclusion and Future Work

Although VPNs provide great security option to protect the networks from attacks, it also allows some attacks in certain scenarios which in this case was cryptojacking and the attack were performed and the impact of the attack was also discussed. To overcome these situations, a zero-trust security model concept was introduced which is software-defined perimeter. SDP's architecture along with some core components involved in the functionality of SDP was highlighted. Implementation of SDP to showcase how this security framework is capable of mitigating such scenarios where VPNs failed to handle it had always been the primary focus of this paper. With the help of SDP's security layers such as dynamic firewalls and mTLS tunnel, attacks such as MITM and cryptojacking can be mitigated. The future work of this research extends to implementing SDP in a more complex manner. It also involves implementing this model in an enterprise network for getting the more real-world applications of SDP into reality.

References

1. A. Moubayed, A. Refaey, A. Shami, Software-defined perimeter (sdp): state of the art secure solution for modern networks. IEEE Netw. **33**(5), 226–233 (2019)
2. P. Kumar et al., Performance analysis of sdp for secure internal enterprises in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE (2019)
3. D. Puthal, et al., Building security perimeters to protect network systems against cyber threats (future directions) IEEE Consumer Electronics Mag. *6*(4), 24–27 (2017)
4. A. Sallam, A. Refaey, A. Shami, Securing smart home networks with software-defined perimeter, in *2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE (2019)
5. J. Koilpillai, Software defined perimeter (SDP) a primer for cios. Waverley Labs LLC (2017)

6. R. Harish, et al., Facilitating cryptojacking through internet middle boxes, in *Advances in Electrical and Computer Technologies: Select Proceedings of ICAECT 2020*. Springer, Singapore (2021)
7. S. Eskandari, et al., A first look at browser-based cryptojacking, in *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE (2018)
8. S. Rose, et al., *Zero Trust Architecture*. No. NIST Special Publication (SP) 800-207 (Draft). National Institute of Standards and Technology (2019)
9. P. Wallker, R. Santhya, M. Sethumadhavan, Anonymous network based on software defined networking, in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)* (48184), IEEE (2020)
10. M. Suresh, et al., An investigation on HTTP/2 security. J. Cyber Security Mobility, 161–180 (2018)
11. VPN Tunneling. https://seedsecuritylabs.org/Labs_16.04/Networking/VPN_Tunnel/. Last accessed 4 Apr 2021
12. Virtual Private Network. https://seedsecuritylabs.org/Labs_16.04/Networking/VPN/. Last accessed 5 Apr 2021
13. SDP Controller. https://github.com/WaverleyLabs/SDPcontroller. Last accessed 14 Apr 2021
14. Software Defined Perimeter Client and Gateway Components. https://github.com/WaverleyLabs/fwknop. Last accessed 4 Mar 2021
15. CoinIMP. https://www.coinimp.com/documentation. Last accessed 24 Mar 2021
16. P. Karunakaran, Deep learning approach to DGA classification for effective cyber security. J. Ubiquitous Comput. Commun. Technol. (UCCT) **2**(04), 203–213 (2020)
17. G. Ranganathan, Real time anomaly detection techniques using pyspark frame work. J. Artificial Intelligence **2**(01), 20–30 (2020)
18. M. Suresh, et al., Exploitation of HTTP/2 proxies for cryptojacking, in *International Symposium on Security in Computing and Communication*. Springer, Singapore (2019)

Privacy Preserving Authentication and Access Control Scheme in Data Markets

Bhaskaruni Sai Praneetha and B. Sekhar Babu

Abstract Online company for various fields now has high margins due to its more transparent and complex nature. Several data holders' data can be shared with third-party vendors when conducting online business, and those third-party vendors can then share it with end users through those data vendors. If the data owner wishes to sell any digital content over online sources when processing these things, it is presently very difficult to keep track of data protection, privacy, and honesty while doing it. In this paper, privacy preserving and authentication and access control scheme in data markets are implemented. In this, online database performs the revocation and registration process. First, user should register with raw data. The raw data will be authenticated using authenticated data encryption technique. Next, this encrypted data will be processed using data processing. Pseudo-identity generation will provide the password for driver authentication. After data is encrypted, it is verified. This is verified based on two batches. First batch will be used for verification, and second batch will collect the encrypted data. At last, online data will be secured very effectively.

Keywords Data markets · Data truthfulness · Privacy preservation · Security · TPM · Homomorphic encryption

1 Introduction

As a result of the massive growth in the business paradigm, multiple online data levels have exploded in order to meet or fulfill society's minimum criteria for single user-centric data. The data contributors can build the data and send it to data server, which will be accessible to service provider. Data contributor's specific data information is stored by the service provider, who is responsible for data security [1].

B. S. Praneetha (✉) · B. S. Babu
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

B. S. Babu
e-mail: sekharbabu@kluniversity.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_23

Customers pay for data from service providers, and it is therefore important to provide them with the accurate and verified data. The key issue that data users face is that they are unable to obtain data because data can be changed by any third-party application. Gnip, Twitter API tool for business, as an example, gathers social media data from Twitter users, mines keen insight into personalized audiences, and offers data processing solutions to more than 94% of Fortune 500 companies. Even so, these market-based networks have serious security problems; namely, it is hard to verify the accuracy of data collection and processing, particularly when the data contributors' privacy must be protected.

Again, since the data contributor's information is confidential and sensitive, it should not be easily revealed [2]. The proposed framework includes Technical Programme Discussion Meeting (TPM), which proficiently coordinates privacy safeguarding and truthfulness in data centers that are accessible through large servers. TPM (Technical Programme Discussion Meeting) is set up in an Encrypt-then-Sign pattern within, with incompletely homomorphic encryption and character based markup.

Partially homomorphic encryption plans were used to allow useful calculations on scrambled data in order to achieve a trade-off between effectiveness and execution. Partially homomorphic encryption plans are different from restrictively moderate fully homomorphic encryption plans that aid discretionary tasks in that they focus on explicit function(s) and achieve better execution [3]. The cryptosystem is a great example of this, as it saves the expansion homomorphism and allows consistent development.

Individually, these plans enable the specialist organization and the information buyer to conduct successful information handling and result checks over scrambled data [4]. Furthermore, framework recognizes that result confirmation in the information markets differs from undeniable estimation in re-appropriating circumstances, since prior to information preparation, the information buyer, as a consumer, may refer to framework specialized report for increasingly relevant work.

To begin with, existing applications in information markets have not provided the security guarantees considered in the TPM scheme, as per the current trend in information security [5]. Second, when serving upward of 1 million data contributors and service providers, the measurement overhead at the professional company is 0.930 s per coordinating with 10 evaluating characteristics in each profile. Moreover, the measurement overhead at the specialist co-op for the information circulation gain is 144.944 s when helping 10,000 information givers and 8 arbitrary variables.

The enrollment emphasis primary responsibility is to set the structure parameters for the character-based mark plot and the cryptosystem. In addition, absolute decodings are needed in the profile coordination and information distribution administrations, respectively. When the percentage of invalid marks is up to 16, clump check is preferable to single mark confirmation. When invalid marks are bunched together, the most negative scenario of group search occurs; cluster confirmation execution should be improved.

Furthermore, as stated in the introduction, the expert co-op will set a practical following depth and allow those unidentified information supporters to resubmit. It also plots the correspondence profile coordinating overhead, where the personality based mark plot is actualized, the limit is set at 12 and the number of characteristics is fixed at 10.

The communication overheads are simply checking in the amount of content sent. Furthermore, the system only considers the correctness search. The common misconception is that the specialist organization's and information customer's correspondence overheads grow in lockstep with the number of legitimate information providers, while the correspondence overhead of each data pattern stays unaltered [6].

The cause for this is that each data provider only has to fill out one profile accommodation, so the cost is independent of m. In any case, the specialist organization must submit m encoded similitudes for decoding and files and cypher texts of organized knowledge patrons for verification [7]. With respect to information, the communication overhead is primarily caused by one information accommodation and the process of encoded likenesses for decoding.

According to this, Technical Programme Discussion Meeting (TPM) is the primary safe instrument for information markets that achieves both data integrity and data safety conservation.

- TPM is set up to encrypt-then-sign, with homomorphic encryption and identity-based signatures as part of the process. It allows the service provider to collect and process accurate data in an honest manner. Moreover, Technical Programme Discussion Meeting (TPM) combines a two-layer cluster validation scheme with an efficient result check scheme, lowering calculation overhead significantly.
- Technical Programme Discussion Meeting (TPM) is educationally instantiated with two types of useful information services, namely profile coordination and information distribution.

2 Literature Survey

Lackner et al. [1] creates and evaluates private stats, a system for processing complete insights over area data that achieves two properties at once: first, provable certifications on region security despite the server's knowledge of any side data regarding clients, and second, privacy securing responsibilities (i.e., a guarantee against causing harm to consumers by disseminating a lot of false information).

Private stats addresses two significant problems not addressed by previous work: It ensures that no extra data is released, even though there are self-assertive side data flaws, and it gives customers accountability without a restriction in data collection. Framework performed private stats on things like mobiles and servers and exhibited its reasonableness.

Long et al. [5] have worked on the technique, which entails connecting neural systems to CryptoNets. This neural system is linked to the user's scrambled info, which is encoded. As a result, data contributors can gain some faith and send personal encrypted data over the cloud with trust. Using General Processing Units (GPUs) and field-programmable gate arrays (FPGAs) to speed up calculations will significantly improve throughput and idle time. Another path to progress will be to find increasingly successful encoding strategies which take into the account littler parameters and subsequently quicker homomorphic calculation.

Sh et al. [2] suggest diagram encryption plots to help with inelegant most restricted segregation questions on large-scale jumbled charts. Shortest segregation inquests are a standout, one of the most important diagram activities, and they can be used in a variety of ways. For large-scale diagrams, advancements are level headed.

Alsharif et al. [4] intend to protect personal data from organization's while preserving the framework's functionality. To create proposals, the framework proposes encoding confidential data and ability to handle it under encryption. The whole job paves the way for confidential recommendations to be generated in a secure manner.

Lozano-Garzon and Donoso [6] implemented Venus; Venus is a major benefit-driven data acquisition method for the team-identified information markets. Venus is made up of two systems that are very similar: One is Venuspro, and another one is Venuspay. Venuspro is used for the sales enhancement and Venuspay for the installment minimization. In terms of installments, Venuspay circumvents the standard second-value sale. The present system is only handled in terms of transferring information without adding some form of protection to data, so information theft is common. There is no fallback recovery option available when it is discovered that a client was not using authorized information; this limitation is eliminated in the suggested approach.

Sultan et al. [7] suggests Account Trade, the set of responsible agreements for large-scale data exchange between dishonest buyers. Framework agreements achieve accounting power and accountability toward unreliable consumers who can create mischief during dataset transfers to stabilize the massive information sharing situation. Similarly, only some conscientious sharing agreements to encourage information representatives to warn the dishonest customer whenever the fraudulent behavior is detected.

Framework provides a structured validation tool to properly characterize, explain, and measure the obligation of framework conventions, much like it does in real-world datasets. Some of the difficulties make it unimportant to configuration account trade. The distinction between legal and illegal transactions is difficult to describe straight away. This is primarily due to the fact that fraudulent vendors may introduce additional irritation into others' datasets before achieving to exchange them, and determining how much information should be irritated in order to be free of the first one is not even in the software engineering space. This will allow efficient security models, for example, k-secrecy and assorted variety, to be used while reducing the loss of data caused by the existing procedure. The primary class is based on a rugged nearest neighbor relationship.

Data contributors have a number of privacy issues. Nonetheless, the service-based trade model, which conceals the confidential raw information, allays their fears. Semantically rich and informative data services will boost income for service providers.

3 Authentication and Access Control Scheme

Figure 1 depicts the architecture of proposed system. In this, online database performs the revocation and registration process. First, user should register with raw data. The raw data will be authenticated using authenticated data encryption technique [8]. Next, this encrypted data will be processed using data processing. Pseudo-identity generation will provide the password for driver authentication. After data is encrypted, it is verified. This is verified based on two batches. First batch will be used for verification, and second batch will collect the encrypted data. At last, online data will be secured very effectively.

Stage I: Data initialization and preprocessing.

Stage II: Generation of signing keys

Stage III: Submission of data

Stage IV: Processing of data and verification

Stage V: Revocation and tracing

Online data sets are facilitated on sites made accessible as programming as an assistance items open by means of an internet browser [9]. These days, increasingly more information markets give information benefits instead of straightforwardly offering raw information [10]. Information handling, which makes confirming the honesty of information assortment much harder.

An online dataset is an information base open from a neighborhood organization or the Internet, rather than one that is put away locally on an individual personal computer (PC) or its group [11]. Verified encryption and validated encryption with related information are types of encryption that ensure the classification and legitimacy of information [12].

Validation-based marking empowers you to sign at least one computerized archive electronically by first consenting to the substance of the document(s) and

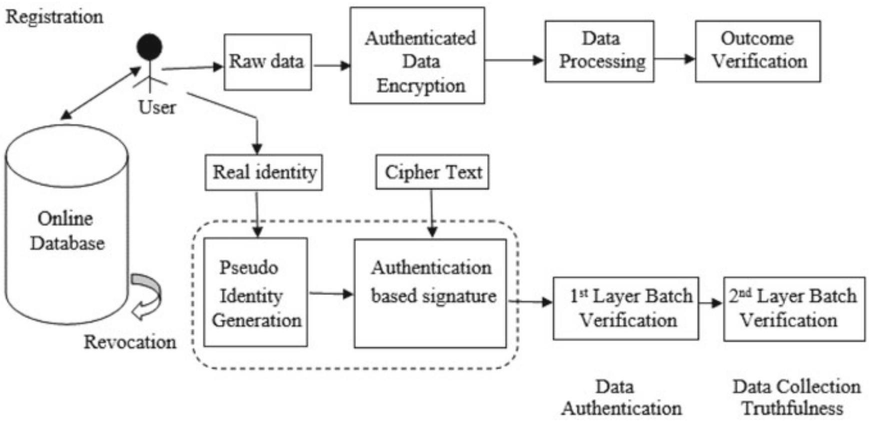

Registration
User
Raw data
Authenticated Data Encryption
Data Processing
Outcome Verification
Online Database
Revocation
Real identity
Cipher Text
Pseudo Identity Generation
Authentication based signature
1st Layer Batch Verification
2nd Layer Batch Verification
Data Authentication
Data Collection Truthfulness

Fig. 1 Authentication and access control scheme

hence confirming yourself utilizing one of various upheld strategies for verification, including ID suppliers.

At this stage, this considers two-layer bunch confirmations, i.e., checks directed by both the specialist cooperation and the information buyer. Between the two-layer bunch confirmations this present information handling and marks accumulation done by the specialist cooperation. Finally, this present result check directed by the information customer.

Rather than straightforwardly exchanging raw information for income, increasingly more specialist co-ops will in general exchange esteem added information administrations. Normal instances of information administrations incorporate interpersonal organization examinations, customized suggestions, area-based administrations, and likelihood dissemination fittings.

Result verification will have homomorphic properties which additionally empower the information shopper to check the honesty of information preparing. The two-layer group checks possibly hold when every one of the marks is available in any event, when there is a solitary invalid mark. Practically speaking, a mark cluster may contain invalid one(s) brought about by unplanned information process or potentially exercises dispatched by an outside scheme [13].

Traditional batch verifier would dismiss the whole bunch, regardless of whether there is a solitary invalid mark, and hence squander the other legitimate information things. In this manner, following as well as recalling invalid information things and their relating marks are significant. In the unlikely case that the second-layer group check the system, the information purchaser can require the specialist cooperation to discover the invalid signature(s). Additionally, if the principal layer cluster check fizzles, the specialist cooperation needs to discover the invalid one(s) without help from anyone else.

Subsequently, it is infeasible for the specialist organization to fashion substantial marks for the benefit of any enrolled information benefactor. A particularly engaging property keeps the specialist organization from infusing fake information imperceptibly and implements this to honestly gather genuine information [14]. Like information validation and information respectability, the information customer can confirm the honesty of information assortment by playing out the second-layer clump confirmation.

In the first place, the specialist co-op cannot purposely discard an information benefactor's genuine information. The explanation is that if the information patron has presented in this its scrambled raw information, without tracking down the pseudo-personality on the certificated notice board, he would get no compensation for information commitment [15]. In this way, this has motivating forces to report information missing to the enlistment community, which thusly guarantees the rightness of n.

Second, this considers that the specialist organization bargains the quantity of substantial information benefactors are only two: One is to place a legitimate information donor's pseudo-personality into the boycott; the other is to place an invalid pseudo-character into the white rundown. This examines these two cases independently: (1) During the first instance, substantial information benefactor would get no award, and (2) however in secondary case, it may likewise be renounced from the online enrollment dataset.

4 Results

Table 1 shows the result analysis of number of files processed using proposed system. In this, time to securing data in milliseconds and time required to regenerate original data in seconds are given in a detailed manner. From this, it can observe that the proposed system will secure the data in an effective way.

Figure 2 shows the analysis of time to protecting the data in milliseconds for 6 files.

Figure 3 shows the analysis of time it takes to re-create the actual information in seconds for 6 files.

Table. 1 Result analysis

S. no.	Total number of files processed	Time to securing data in milliseconds	Time required to regenerate original data in seconds
1	10	310	19
2	20	270	21
3	30	320	26
4	40	410	28
5	50	460	29
6	60	320	30

Fig. 2 Time to securing data in milliseconds

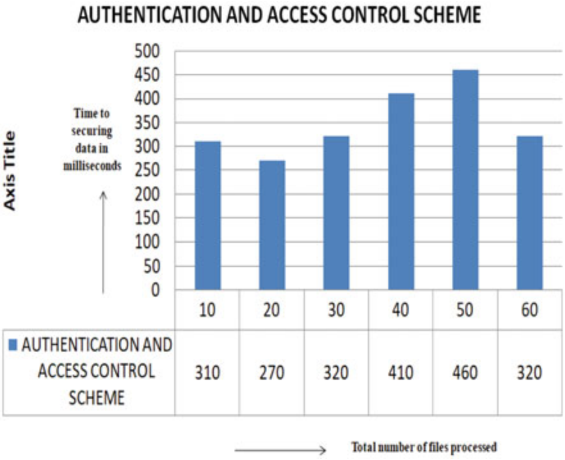
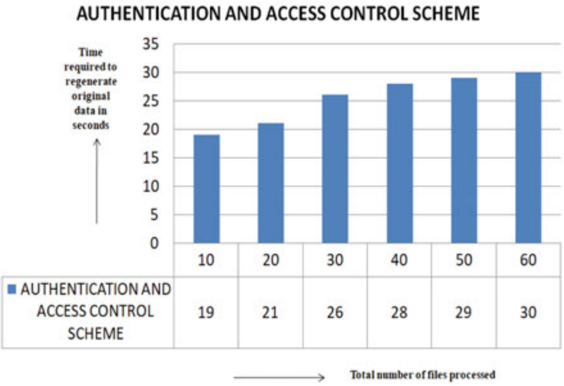


Fig. 3 Time required to regenerate original data in seconds



5 Conclusion

Hence in this paper, privacy preserving and authentication and access control scheme in data markets were implemented. In this, online database performs the revocation and registration process. First, user should register with raw data. The raw data will be authenticated using authenticated data encryption technique. Next, this encrypted data will be processed using data processing. Pseudo-identity generation will provide the password for driver authentication. After data is encrypted, it is verified. This is verified based on two batches. First batch will be used for verification, and second batch will collect the encrypted data. At last, online data will be secured very effectively. From results, it can observe that data is secured effectively. Information awareness is needed for authentication of digital signature systems, and it could actually reveal the true identity of a data contributor. Internally, TPM is organized in an encrypt-then-sign process that uses a combination of resemblance cryptography and identity-based signature. Its purpose is to check whether the authenticity of privacy protection in the information market is protected. In my next paper, I will work on comparison methods.

References

1. C. Lackner, D. Osipo, H. Cui, J.H. Chow, A privacy-preserving distributed wide-area automatic generation control scheme. **8**, (2020)
2. Y. Sh, E. Alsusa, M.W. Baidas, Downlink-Uplink decoupled access in heterogeneous cellular networks with UAVs. 978-1-7281-490-0/20/$31.00 ©2020 IEEE
3. Z. Erkin, T. Veugen, T. Toft, R.L. Lagendijk, Generating private recommendations effeciently using homomorphic encryption and data packing. IEEE Trans. Inf. Forensics Secur. **7**(3), 1053–1066 (2012)
4. A. Alsharif, M. Nabil, M.M.E.A. Mahmoud, M. Abdallah, EPDA: efficient and privacy-preserving data collection and access control scheme for multi-recipient AMI networks. **7**, (2019)

5. Y. Long, Y. Chen, W. Ren, H. Dou, N.N. Xiong, DePET: a decentralized privacy-preserving energy trading scheme for vehicular energy network via blockchain and K—Anonymity. **8**, (2020)
6. C. Lozano-Garzon, Y. Donoso, A multi attribute planning algorithm for the selection of joint processing schemes in a smart cities environment. 978–1–5386–1934–6/18/$31.00 ©2018 IEEE
7. N.H. Sultan, F.A. Barbhuiya, N. Sarma, SCAuth: selective cloud user authorization for ciphertext-policy attribute-based access control. 978–1–5386–2450–0/17 $31.00 © 2017 IEEE
8. G. Ghinita, P. Kalnis, Y. Tao, Anonymous publication of sensitive transactional data. IEEE Trans. Knowl. Data Eng. **23**(2), 161–174 (2011)
9. S. Sen, C. Joe-Wong, S. Ha, M. Chiang, Princeton University, "Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access", IEEE Commun. Mag. • 2012
10. D.H. Van, N.D. Thuc, A privacy preserving message authentication code, in *IEEE*, 2015
11. D. Liu, H. Li, Y. Yang, H. Yang, Achieving multi-authority access control with efficient attribute revocation in smart grid. 978-1-4799-2003-7/14/$31.00 ©2014 IEEE.
12. M. Ocenas, I. Homoliak, P. Hanacek, K. Malinka, Security and encryption at modern databases, in *International Conference on Cryptography, Security and Privacy*, pp.19–23 (2020)
13. J. Wu, M. Dong, K. Ota, M. Tariq, L. Guo, Cross-Domain fine-grained data usage control service for industrial wireless sensor networks. **3**, (2015)
14. Z. Niu, F. Kubotai, A Bursty transmission scheme for wireless ATM and its analysis, 1996, 0-7803-2916-3
15. C. Li, D.Y. Li, G. Miklau, D. Suciu, A Theory of pricing private data, in *Communications of ACM*, vol. 60, no. 12, pp. 79–86 (2017)
16. A. Sungheetha, R. Sharma, Novel shared key transfer protocol for secure data transmission in distributed wireless networks. J. Trends Comput. Sci. Smart Technol. (TCSST) **2**(02), 98–108 (2020)
17. D. Sivaganesan, Smart contract based industrial data preservation on block chain. J. Ubiquitous Comput. Commun. Technol. (UCCT) **2**(01), 39–47 (2020)

Communication Framework for Real-Time Monitoring of a Smart Grid Emulator

T. V. Sarath, P. Sivraj, and K. Kottayil Sasi

Abstract Wired and wireless communication technologies are the enablers of the future smart grid communication. Communication network and choice of communication technology are subject to technical, economic, and legal considerations. In this work, a simulation-based analysis of different communication technologies for real-time monitoring of smart microgrid is carried out. A hardware implementation of the communication network with the communication technologies selected from the analysis is successfully carried out on a smart microgrid emulator. It is observed that a heterogeneous network of wired LAN forming the high-speed backbone network supported by a WLAN for last mile connectivity is the most proper choice for smart grid applications.

Keywords Smart grid · Smart grid communication · Communication networks · IoT · WSN · NetSim · Communication performance metric · SMGE · MQTT

1 Introduction

The power system network has been growing vastly over the last two decades to satisfy the ever-growing power demand. Though the consumer demand for reliable and secure power is also increasing, the existing power grid is not capable of handling such demands in the current state. The importance of stable functioning of the power grid has gained paramount interest as critical services like transport, medical, communications, and finance depend on reliable, economic, and secure power [1]. Being a large interconnected network, a small disturbance at one location can propagate through the grid and may lead to cascaded grid failure. Even though the initial phase of grid automation improved the operating environment and efficiency of the system, it was no way near the expected objectives of automated power grid. This disparity

T. V. Sarath (✉) · P. Sivraj · K. K. Sasi
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: tv_sarath@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_24

has led to shift of focus on to smart grid (SG) which emphasizes on a distributed power generation paradigm in place of the existing centralized one, renewable source integration, integration of information and communications technology (ICT) on legacy infrastructure and direct involvement of customers. The fundamental triggers attributed to SG are aging power system infrastructure, rising energy demand, deteriorating reliability, and monopoly in the utility sector [2]. SG is a co-network of communication and power, which adopts a decentralized system topology with bidirectional communication based on real-time data. SG has the ability to self-heal, prevent rapid deterioration, and minimize the effect of any disturbance in the system. Extensive consumer participation, option of becoming a prosumer, fully automated real-time metering/control, outage prevention, minimal recovery time, and quality power are the other features provided by SG. SG will revolutionize the energy sector by making the grid more efficient, giving access to real-time grid parameters, facilitating increased penetration of renewable energy sources and offering provision for end users to involve in energy market [3].

The required level of SG control can be achieved only when the detailed data of operational parameters is accessible at any point of time. The network of intelligent electronic devices is deployed in various strata of the grid monitors and reports the parameters of the grid in real time. However, the number of phasor measurement unit (PMU) deployed is relatively less, and it covers only the transmission sector. This necessitates a distributed paradigm with many units of a PMU-like device deployed in the distribution network to access operational parameters. In such a large distributed network of sensing devices, the potential for utilizing concepts of wireless sensor networks and the Internet of things (IoT) is enormous. The proliferation of smart meters (SM) opens up a possibility to improve the reliability and quality of power as it provides fine-grained information about the distributed energy utilization [4]. Pervasive communication technologies can facilitate integration of devices like SM into the SG. The requirements of communication system in each SG application will vary in terms of reliability, coverage, responsiveness, and security. The availability of a myriad of wired and wireless communication technologies presents a unique and exciting challenge in terms of the right selection of technology for a specific SG application [5, 6].

This work presents a quantitative analysis of various communication technologies for a real-time monitoring application in a smart microgrid followed by implementation of the communication network using intelligent devices on a smart microgrid emulator (SMGE). The paper presents the background study in Sect. 2. The description of the system under study is in Sect. 3. Section 3.1 describes the simulation scenario considered for the analysis. The simulation results and hardware implementation results are presented in Sect. 4. The conclusions drawn on the work are in Sect. 5.

2 Related Work

The growing number of components in the SG subsystems and their interrelation to each other necessitates a simulation-based approach to understand the issues in a highly interconnected network prior to field implementation. There is a high degree of complexity in the SG due to interactions between the traditional power network and the communication system. The reliable operation of such a cyber-physical system is reliant on ICT, control algorithms, and data analytics. Before a large-scale deployment, the interaction among various components of the system must come under investigation. There is need for a co-simulation and real-time hardware-in-the-loop simulation approach for testing and validating the concepts of SG at different stages of its development [7]. The pilot projects to assess feasibility of SG concepts have reported promising results in terms of reliability. Due to limitations and complications, the utilities have implemented these projects in a limited scope of SG applications with minimal impact on existing grid operations [8]. A large-scale deployment needs to consider more actors/members and more services. A few of the challenges in the fore are high capital investment, involvement of regional governments and people, engagement of stakeholders, incentives for collaborations, security, and fear of obsolescence [7, 9]. But all these become non-issues when real-world simulators/emulators are used to validate SG concepts. A simulator that works on real-time data can speed up the development and implementation of SG. Apart from avoiding the costs and the risks, these electrical network simulators replicate various scenarios which would be hard to recreate in conventional power systems. The SG simulators/testbeds can essentially bolster the understanding of the three scenarios, namely normal system operation, expansion, and operation under disruption/disturbance [10]. The simulators provide a feasible platform for testing new algorithms for optimization of operations and minimization of losses. Even critical scenarios like disturbance or disaster/malicious attacks (physical and cyber) can be simulated in an emulator [11]. In short, a microgrid testbed that can simulate real-world microgrid projects can help accelerate the microgrid research and its applications. Such simulators provide a platform for testing and verification of SG research and projects. There are SG simulators like Tianjin University microgrid Testbed, SmartGridLab, GridSim, SCORE, and GridLAB-D which are currently used to evaluate the designed/developed protocols, simulate the real-time dynamic behavior of grid, and develop energy management tools based on the real-time data. In the simulators that were reviewed, the focus of the work was either limited to electrical network with flexibility to incorporate multiple distributed energy sources and complete software simulation models for electrical network or simulation model for communication network for specific SG applications. The choice of communication network for all the application scenarios in the hardware in loop simulators that were reviewed were mainly Ethernet-based communication to reduce the complexity of the system [12–14]. Power system optimization involves keeping track of power flow, collecting the data and analyzing the opportunity to optimize the performance. The accuracy of the result of the analysis is directly dependent on the amount of data collected; more data implies more accurate assessment of power system behavior. This requirement of data would demand a device that is capable of automatically monitoring and recording the power flow at regular intervals and pass this information on to higher levels of the system for analysis. The grid server will have enough data to study the behavior of the loads by placing the data collection units in the distribution sector of the power grid. The integration of communication technology provides additional advantages in sensing and control of the system. However, it adds further constraints such as latency and packet delivery loss. The degree of automation in the power system is impacted by the effectiveness of the communication infrastructure to deliver real-time data from sensing nodes. The communication system for the future SG should maintain the required levels of reliability, delay, and packet loss as well as handle the massive amount of data generated at all strata of power systems and must be robust against communication decadency [15]. The communication network chosen for a particular SG application must meet the critical communication performance metrics [16–18]. The transformation of the power grid into an SG paradigm craves for intelligent devices in the system which is capable of real-time monitoring and synchronized data acquisition [3]. For the change from a centralized model to a distributed architecture to happen, the need for distribution side PMUs arises whose functionality may not be entirely the same as the PMU deployed on the transmission side. Realization of concepts like demand response, dynamic energy management, and complete automation of the distribution sector requires a vast set of data from different sections in the distribution sector which can be acquired by increased use of real-time data sensing units [19]. The technological advancements in the field of wireless sensor networks and IoT yield a solution to the challenges in collecting the vast amount of SG data. Developments in SMs and sensory devices have seen promising outcomes in applications like power quality, reliability analysis, energy saving, dynamic pricing, power outage management, and appliance control. Geographical distance between various networks in SG poses a severe challenge in terms of communication latency and quality of service (QoS) in data transfer [4]. The emergence of the Internet of energy (IoE), a subset of IoT focusing on implementing concepts of IoT in smart power systems, will help faster integration of power system components [20, 21]. The communication model used in a network of sensory devices plays an essential role in ensuring the QoS of communication. Highly distributed nodes in a network generate considerable amount of data for analysis. Mechanisms like data aggregation in an intermediate node can bring in relaxation in network overhead [22]. The request–response model—one of the widely implemented models in Web services—provides a problem as the host server needs to be active all the time even when no data transfer occurs. Such a model deployed in the last mile connectivity network will lead to a burden in the communication network due to the presence of a vast number of communicating devices [23]. An alternative to this model is the publish–subscribe model, where data gets transferred only when needed. Message queuing telemetry transport (MQTT) is one of the best publish–subscribe protocols in terms of lightweight implementation and operation in constrained network bandwidth. It runs on top of TCP/IP and provides lightweight, asynchronous publish–subscribe protocol [22]. Constrained application protocol (COAP) uses a subset of hypertext transfer protocol (HTTP) methods to provide a synchronous request–response model over user datagram protocol (UDP) [23, 24]. A proper communication model and IoT protocol can be selected based on the application under consideration. In the existing SG simulators reviewed, the major focus was on the different operational aspects of the electrical network with minimal importance given to communication network selection. In most of the cases, Ethernet-based systems are preferred to reduce the cost and complexity of the system. In this work, a simulation-based selection of the best-fit communication system is done and implemented over a SMGE.

3 Proposed Work

In the typical wide-area monitoring architecture, PMU provides phase and magnitude of the voltage and current along with frequency, rate of change of frequency, MW, MVAR, etc. In the hierarchical bottom-up structure, the next device is phasor data concentrator (PDC) that correlates the data transmitted by the PMUs and forwards it to control centers via higher-level PDCs. It aggregates the synchrophasor data from multiple PMUs and feed these to applications. The communication between PMUs and PDC is standardized by the IEEE C37.118.2 standard. The standard specifies four types of messages—data, configuration, header, and command—that get communicated between PMUs and PDCs with each message type specified with its message format [25].

For this work, we have considered a five-bus system powering an 25 km^2 area having a peak demand of 15 MW at 11 KV. The microgrid under consideration has both renewable as well and conventional power generation integrated to it. For the collection of parameters from the microgrid real-time data collection units (RTDCU) are deployed on the distribution network [26]. The data generated by RTDCUs will be forwarded to server though a data concentrator (DC). Figure 1 shows the single line diagram of the microgrid with the placement of RTDCU, DC, and server. This work deals with a simulation-based communication technology selection for communication network for wide-area monitoring of a smart microgrid and subsequent hardware implementation of a communication framework for real-time monitoring in a smart microgrid emulator (SMGE) using RTDCU.

The sensor network integrated with the smart microgrid emulator follows the IEEE C37.118.2 standard for data communication with the data concentrator (DC). With the testing scenario fixed as eight RTDCUs communicating with one DC in the IEEE C37.118.2 format, the next challenge is the selection of communication technology as there is no general rule for the selection process. Performance metrics like delay, throughput, packet delivery ratio, overhead, etc., help to decide the best among the available communication technologies suited for the application [19]. Before the implementation of a particular communication protocol in the real world, it is essential to check its performance for the network under consideration through simulation using a suitable communication network simulator. NetSim is a prominent

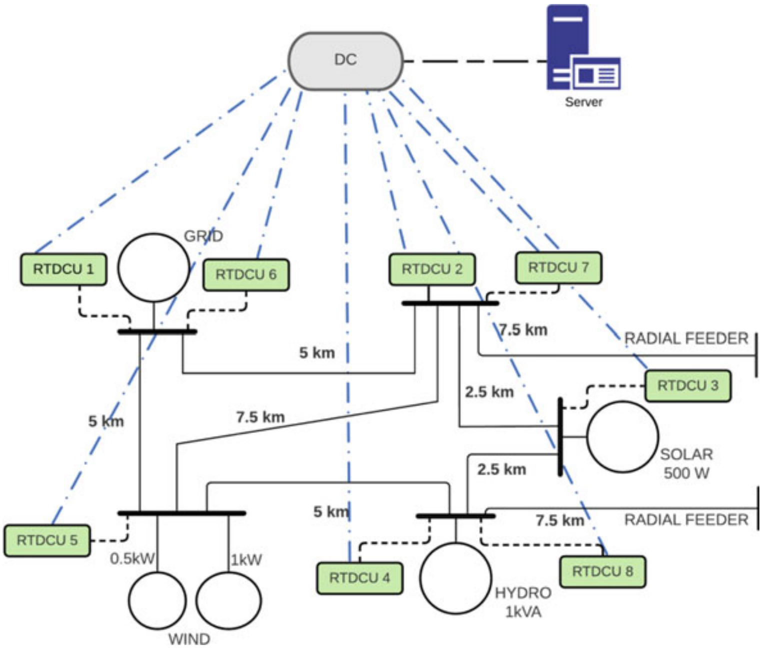


Fig. 1 Five-bus microgrid under consideration

discrete-event network simulation tool used for protocol modeling and simulation of communication protocols for networking studies. It allows us to design networks through drag and drop interface, visualize the simulation and analyze the output performance metrics at multiple levels. NetSim is capable of extending existing algorithms, creating custom protocols, and infusing add-on capability in external software such as MATLAB and SUMO. Further, it allows creation and analysis of computer networks with unmatched depth, ease, power, and flexibility [27].

3.1 Simulation Scenario

This section deals with the comparison of different communication protocols in a standardized simulation environment with the five-bus microgrid simulator as a reference framework. The communication protocols considered for the simulation are WLAN (IEEE 802.11), ZigBee (IEEE 802.15.4), wired LAN (IEEE 802.3), and a heterogeneous network of WLAN and wired LAN. The formats of the types of messages, namely command, configuration, header, and data, defined in IEEE C37.118.2 standard are followed for message passing in NetSim. Table 1 shows the simulation parameters fixed based on the IEEE C37.118.2 standard for PMU communication.

Table 1 Simulation parameters

Message type	Packet size(bytes)	Frames per second
Data	52	50
Configuration	454	1
Command	20	1
Header	18	1

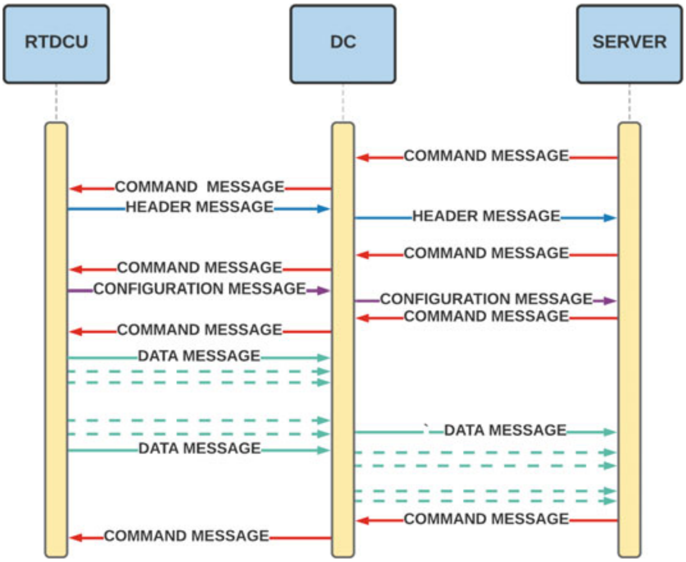


Fig. 2 Flow of messages in the network

Figure 2 illustrates the sequence of message flow in the network through RTDCU, DC, and server. The server initiates the communication by sending a command message to RTDCUs through DC to send the header message, for which the nodes respond with a header message to the server via DC. Once header message from all the nodes is received, the server follows up with the command for configuration message. The response from the RTDUCs contains the information regarding the device configurations. The server then sends a command message to start transmission of the data frame, which contains the actual sensed data. The nodes in the network respond with data frames until the server issues a stop transmission command message. Direction of header, configuration, and data message is from nodes to server via data concentrator, whereas command message flows from server to nodes through data concentrator.

Figure 3 shows the simulation scenario for determining the performance metrics for a homogeneous network of RTDCUs with wired LAN. Figure 4 shows the sim-

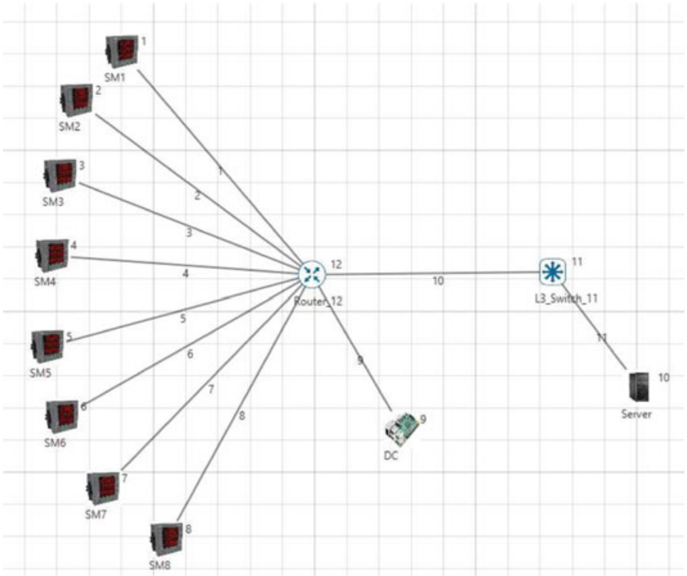


Fig. 3 Simulation scenario of wired LAN

ulation of the network of RTDCUs with WPAN. The data concentrator acts as the coordinator, which forwards the data sent by the RTDCUs to the server node.

Figure 5 depicts the simulation set up for WLAN. All the nodes here use homogeneous communication technology in different layers of the network. The eight RTDCUs use WLAN for communicating with the DC and also from the DC to the server. Figure 6 shows the simulation scenario for evaluating the performance metrics of a heterogeneous network. Here, eight RTDCUs are communicating with the data concentrator through WLAN technology, and communication from the data concentrator to the server is via Internet protocol. The heterogeneity of the system is due to the use of different communication technologies used in the different strata of the architecture.

4 Result Analysis

4.1 Simulation Results

The simulations are carried out for the four different types of messages with each one having different frame length. The results are analyzed with the performance metrics in the network simulation in order to choose the best-suited communication technology for real-time monitoring of the smart microgrid; variation in the performance

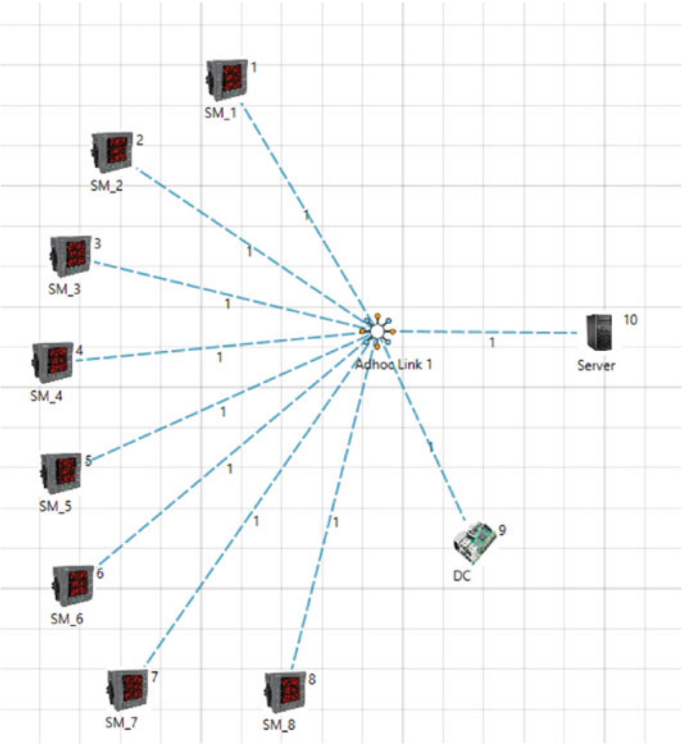


Fig. 4 Simulation scenario of WPAN

metrics like message overhead, average delay, packet delivery ratio, and throughput with packet size is analyzed.

Message overhead Figure 7 shows the variation of message overhead with packet size for four different communication networks. The total message overhead in a network is determined by the size of the data and the maximum permissible limit specified in the communication technology used.

The ZigBee network showed a higher message overhead in the simulation scenarios for packets with large payloads such as the configuration messages. The packets with large payloads are segmented and transferred over the network to the destination, which resulted in the significant overhead. The performances of the wired LAN and the heterogeneous network were slightly better compared to the Wi-Fi network for large payload messages. In the medium packet size messages, wired LAN showed the least overhead compared to the heterogeneous and the Wi-Fi network. The limitation of payload size in the ZigBee network started to add additional overheads in these data packets. However, for the messages with small packet sizes, message overhead was less in the ZigBee networks compared to the wired LAN, which in turn was better than the Wi-Fi and the heterogeneous networks.

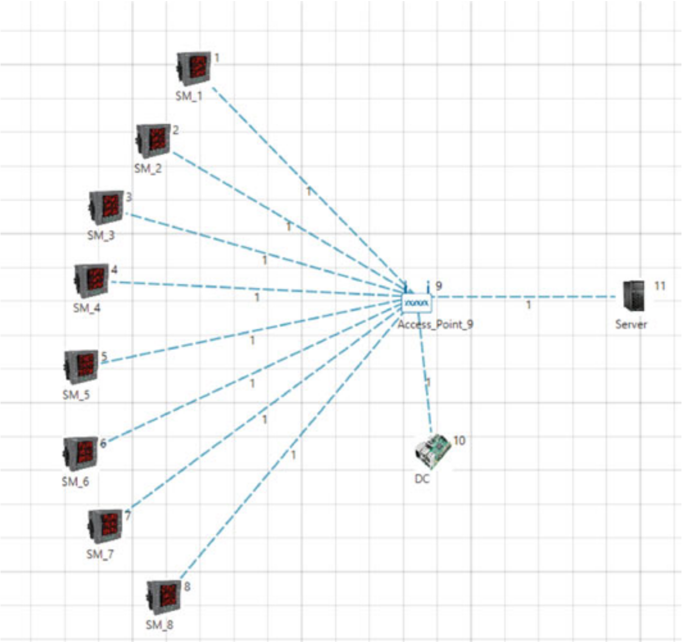


Fig. 5 Simulation scenario of WLAN

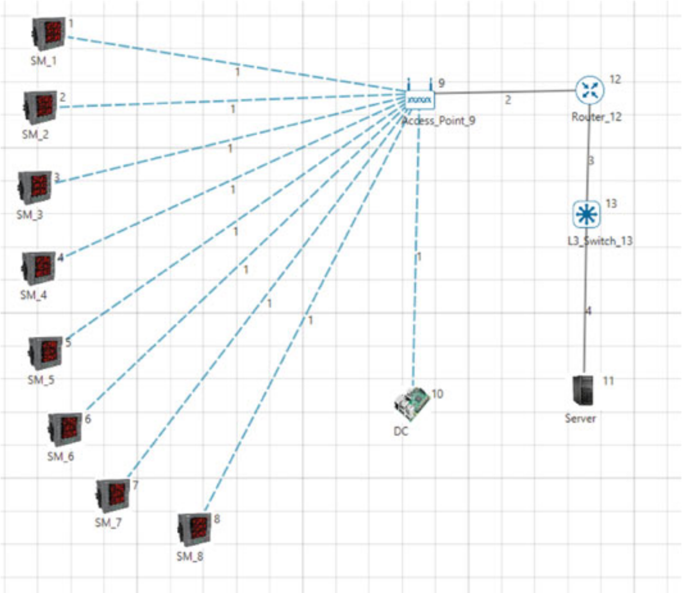


Fig. 6 Simulation scenario of heterogenous network of wired LAN and WLAN

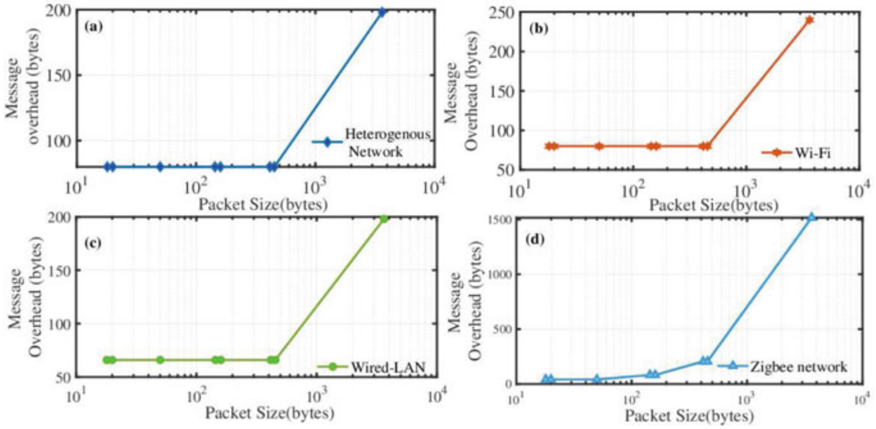


Fig. 7 Performance comparison of message overhead vs packet size: **a** heterogeneous network, **b** WLAN, **c** wired LAN, and **d** WPAN

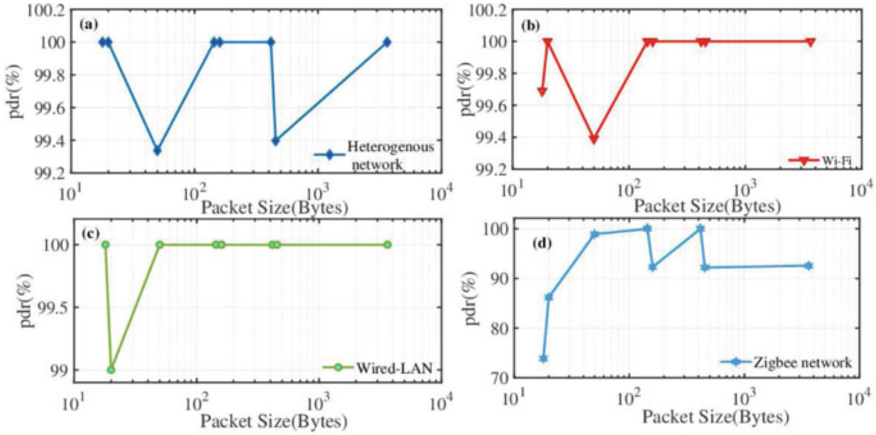


Fig. 8 Performance comparison of packet delivery ratio vs packet size: **a** heterogenous network, **b** WLAN, **c** wired LAN, and **d** WPAN

Packet delivery ratio Figure 8 shows the variation of packet delivery ratio with packet size for the four different communication networks. The heterogeneous, the Wi-Fi, and the wired LAN networks were highly efficient in maintaining high packet delivery ratio irrespective of an increase in the packet size transmitted across the networks. The ZigBee network showcased unsuccessful transmissions due to collision of control packets and expiry of the time to live metric. These collisions resulted in a relatively lesser packet delivery ratio for the ZigBee network compared to the other three networks.

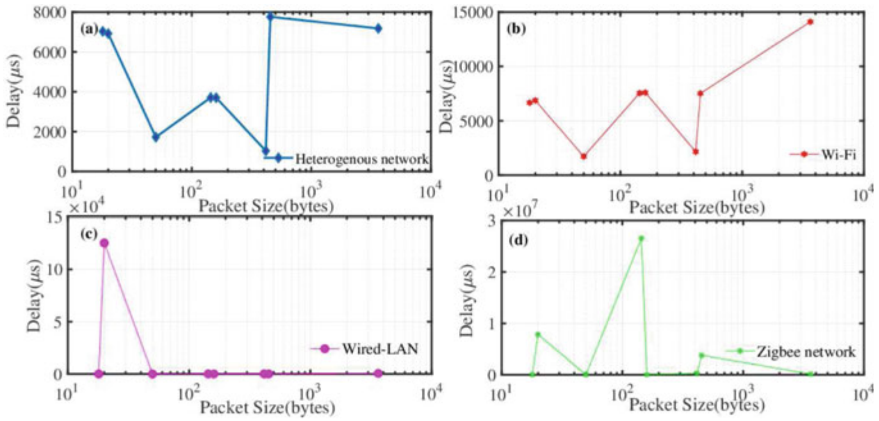


Fig. 9 Performance comparison of average delay versus packet size: **a** heterogeneous network, **b** WLAN, **c** wired LAN, and **d** WPAN

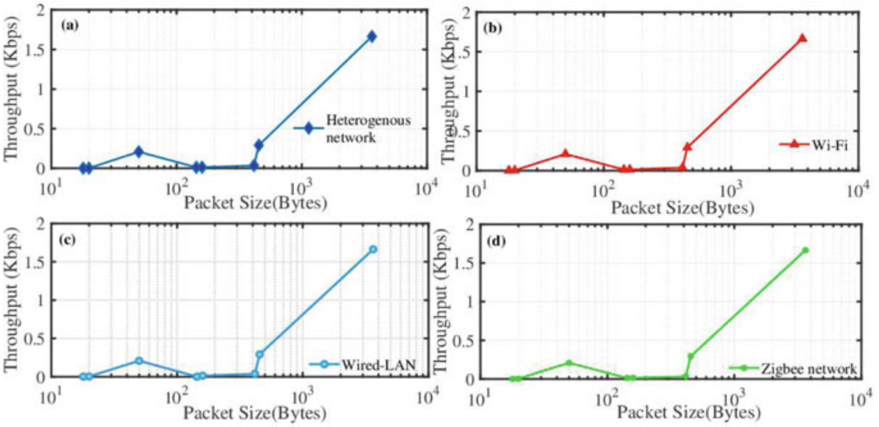


Fig. 10 Performance comparison of throughput vs packet size: **a** heterogenous network, **b** WLAN, **c** wired LAN, and **d** WPAN

Delay metric Figure 9 shows the variation of average delay with packet size for the four communication networks. The ZigBee network exhibited the delay in the range of seconds for specific packet sizes. The wired LAN network exhibited a high peak in the delay for smaller payload, which could be due to an increased number of control messages per packet. Wired LAN has the least delay for the remaining packet sizes. The delay for messages with small packet size in the Wi-Fi network was next to that of wired LAN. The network maintained a similar performance for medium packet size messages also.

Throughput Metric Figure 10 shows the variation of throughput with packet size for four different communication networks. The four networks showed similar performance in terms of throughput for different packet sizes. The networks showcased smaller throughput when the messages transmitted were of smaller packet sizes. The networks started to highlight higher throughputs for larger packet sizes since most of the packets were successfully delivered to the destination also.

From the analysis of the metrics, wired LAN would be the right choice for the communication framework based on quantitative analysis. Economic aspects of a homogeneous network of wired LAN is not attractive. The infrastructure required for this network would be extensive, especially for the last mile network where the chance of a more significant number of nodes joining is high. The wired LAN is underutilized in the last mile network as the bandwidth requirement for that network is relatively small. However, the section of the network from the DC to the server, which requires the capability to provide high bandwidth, can be effectively realized using wired LAN. As the traffic in the core network between the DC and the server will be substantial, choice of wired LAN would be appropriate considering the performance of the wired network in terms of delay, throughput, and packet delivery ratio. By considering two different technologies for the last mile network and the core network, we can compensate for the shortcomings of the homogeneous network. The choice for the last-mile connectivity comes between ZigBee and Wi-Fi network. Taking into account the performance in terms of delay and packet delivery ratio, Wi-Fi comes out as a better choice than ZigBee. The throughput and the message overhead for the packet size relevant in the last mile network came out to be similar for both the networks. The Wi-Fi technology addresses the scalability of nodes effortlessly. A high-speed backbone network of wired LAN and the last-mile connectivity using Wi-Fi makes the proper choice, given technical as well as economic merits.

4.2 Hardware Implementation

This section deals with the hardware implementation of a communication framework for real-time monitoring in a SMGE. A RTDCU mimics minimal functionalities of a PMU in the distribution sector and many such devices are interconnected to form a sensor network which performs data acquisition in a synchronized manner [7]. The network of RTDCUs is reminiscent of the synchrophasor data collection network in the transmission sector. Figure 11 shows the front panel view of the SMGE, the implementation platform.

The SMGE serves as the hardware implementation platform for the network based on the selected communication network. The RTDCU nodes in SMGE communicates with the data concentrator using Wi-Fi technology while the communication between the data concentrator and server adopts wired LAN. Figure 12 shows the scheme of the hardware implementation in which the data measuring nodes are RTDCUs placed on the SMGE. Each of this node is a part of the Wi-Fi network, communicating with the data concentrator, which in turn forwards the aggregated message to the server via Ethernet. Schneider EM6400 smart meter is used in SMGE; it is capable of providing real-time data in Modbus protocol. A Raspberry Pi translates the data

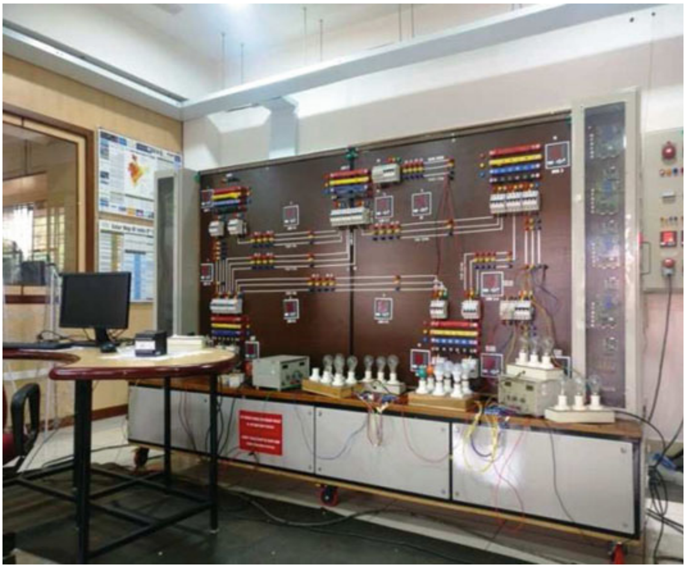


Fig. 11 Front panel of the SMGE

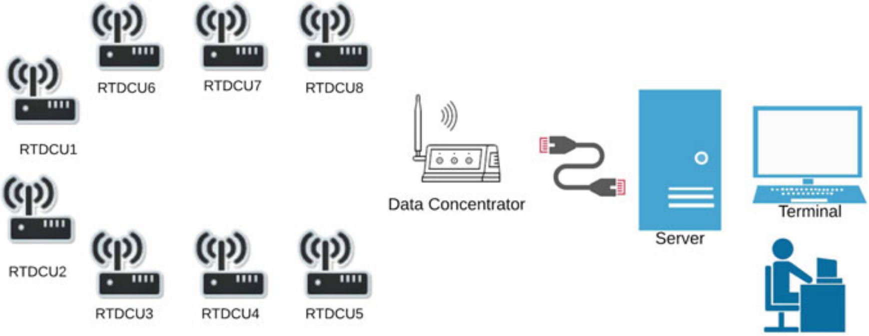
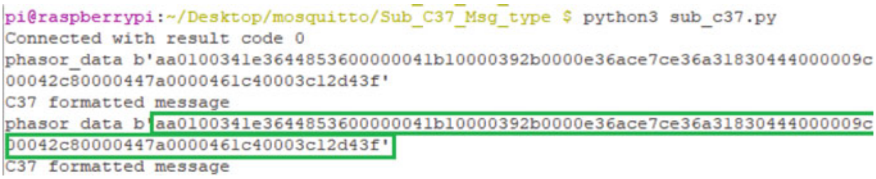


Fig. 12 System overview of HW implementation

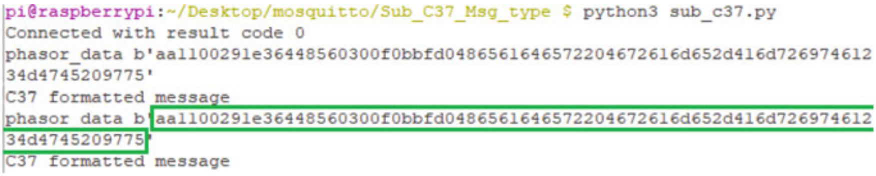
from the SM through the Modbus. Thus, a combination of SM and Raspberry Pi performs the function of RTDCU. Data retrieved from the SM is formatted to the IEEE C.37 standard using synchrophasor library, an open-source Python package, for synchrophasor data transfer [28].

All the RTDCU nodes communicate to the data concentrator node through Wi-Fi technology. In this hardware emulation, a Raspberry Pi functions as a data concentrator. The hybrid communication model considered for this work has a publish–subscribe model for the network of RTDCUs and DC and a request–response model for communication between the DC and the utility server. The C.37 formatted phasor data from the RTDCU gets published to a data concentrator through MQTT



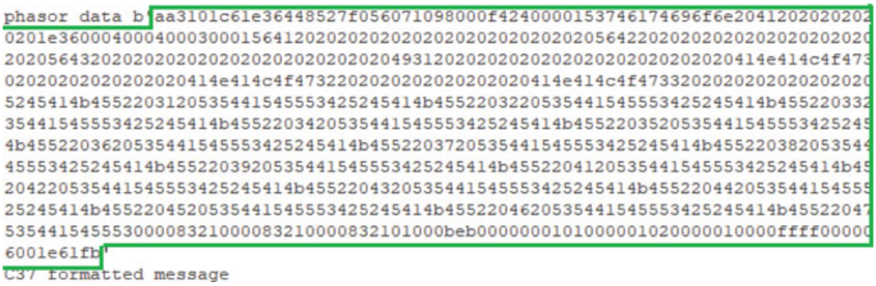
```
pi@raspberrypi:~/Desktop/mosquitto/Sub_C37_Msg_type $ python3 sub_c37.py
Connected with result code 0
phasor_data b'aa0100341e3644853600000041b10000392b0000e36ace7ce36a31830444000009c
00042c80000447a0000461c40003c12d43f'
C37 formatted message
phasor data b'aa0100341e3644853600000041b10000392b0000e36ace7ce36a31830444000009c
00042c80000447a0000461c40003c12d43f'
C37 formatted message
```

Fig. 13 C.37 formatted data message received in the DC



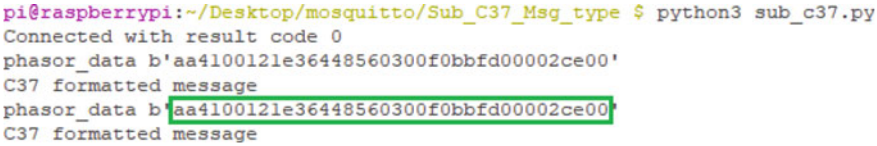
```
pi@raspberrypi:~/Desktop/mosquitto/Sub_C37_Msg_type $ python3 sub_c37.py
Connected with result code 0
phasor_data b'aa1100291e36448560300f0bbfd0486561646572204672616d652d416d726974612
34d4745209775'
C37 formatted message
phasor data b'aa1100291e36448560300f0bbfd0486561646572204672616d652d416d726974612
34d4745209775'
C37 formatted message
```

Fig. 14 C.37 formatted header message received in the data concentrator



```
phasor data b'aa3101c61e36448527f056071098000f4240000153746174696f6e2041202020202
0201e3600040004000300015641202020202020202020202020202020564220202020202020202020202
20205643202020202020202020202020202020204931202020202020202020202020202020414e414c4f473
020202020202020414e414c4f4732202020202020202020414e414c4f47332020202020202020202020
5245414b4552203120535441545553425245414b4552203220535441545553425245414b455220332
3544154555342525245414b4552203420535441545553425245414b45522035205354415455534252 45
4b45522036205354415455534252 45414b4552203720535441545553425245414b4552203820535 44
4555342525245414b4552203920535441545553425245414b4552204120535441545553425245414b45
2042205354415455534252 45414b4552204320535441545553425245414b455220442053544154555
25245414b4552204520535441545553425245414b4552204620535441545553425245414b45522047
53544154555300008321000083210000832101000beb00000001010000010200000100 00ffff00000
6001e61fb'
C37 formatted message
```

Fig. 15 C.37 formatted configuration message received in the data concentrator

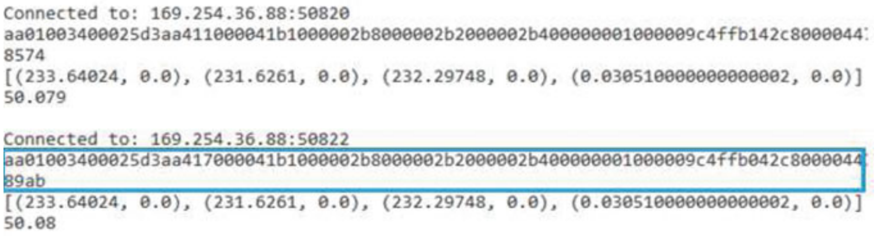


```
pi@raspberrypi:~/Desktop/mosquitto/Sub_C37_Msg_type $ python3 sub_c37.py
Connected with result code 0
phasor_data b'aa4100121e36448560300f0bbfd00002ce00'
C37 formatted message
phasor_data b'aa4100121e36448560300f0bbfd00002ce00'
C37 formatted message
```

Fig. 16 C.37 formatted command message received in the data concentrator

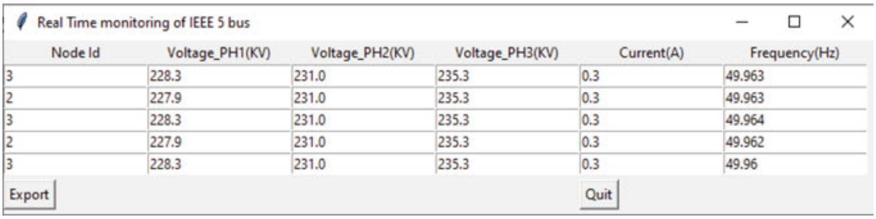
protocol. Figures 13, 14, 15, and 16 depict the data, header, configuration, and command messages received by the data concentrator from the RTDCU in IEEE C.37 format. The actual data are enveloped in boxes in the figures.

The DC subscribes to the phasor data published by RTDCUs and forwards to the server for real-time monitoring of the SMGE. A PC operates as a server that accepts a request from the data concentrator and services them. Figure 17 shows the data transmitted from the data concentrator to the server. A SQL-based database stores all the transmitted data in the server. Figure 18 illustrates a Python-based graphical



```
Connected to: 169.254.36.88:50820
aa01003400025d3aa411000041b1000002b8000002b2000002b400000001000009c4ffb142c8000044
8574
[(233.64024, 0.0), (231.6261, 0.0), (232.29748, 0.0), (0.030510000000000002, 0.0)]
50.079

Connected to: 169.254.36.88:50822
aa01003400025d3aa417000041b1000002b8000002b2000002b400000001000009c4ffb042c8000044
89ab
[(233.64024, 0.0), (231.6261, 0.0), (232.29748, 0.0), (0.030510000000000002, 0.0)]
50.08
```

Fig. 17 Data received by server



Real Time monitoring of IEEE 5 bus

Node Id	Voltage_PH1(KV)	Voltage_PH2(KV)	Voltage_PH3(KV)	Current(A)	Frequency(Hz)
3	228.3	231.0	235.3	0.3	49.963
2	227.9	231.0	235.3	0.3	49.963
3	228.3	231.0	235.3	0.3	49.964
2	227.9	231.0	235.3	0.3	49.962
3	228.3	231.0	235.3	0.3	49.96

Fig. 18 Python-based GUI

user interface (GUI) developed to view the data in tabular form with a provision for exporting all the data to a CSV file.

The simulation analysis will provide the communication technology suitable for the SG scenario but the hardware implementation of the communication network for real-time monitoring brings out the design aspects of the nodes that form the network from an embedded system perspective. The computation power and memory of the device forming sensing nodes are crucial. The hardware implementation helps to demonstrate the operational capability and facility automation of such SG applications.

5 Conclusion

A simulation-based quantitative analysis of performance matrices is used in real-time monitoring in a smart microgrid environment (SMGE), which aids in selecting communication technologies in this work. We observe that wired LAN performs well in terms of message overhead, whereas ZigBee networks have significant overhead with large packets due to data segmentation for transmission. Communication technologies such as heterogenous, Wi-Fi, and wired LAN exhibited excellent performance in terms of packet delivery ratio. The delay seen in ZigBee networks due to packet collision makes it less suitable for critical real-time application. The higher throughput and the minimized delay shown by wired LAN makes it a prime candidate for the network between DC and server. Our study reveals that a heterogenous network of wired and wireless LAN is necessary for real-time monitoring of economically viable SMGE. The wired LAN network can form the high-speed backbone network for a higher hierarchy of nodes, and WLAN can be the last mile connectivity network where more nodes may join in future. We have used a communication model which follows a mixed structure with publish–subscribe in the sensory node network and request–response for higher levels of the network.

Conflict of interest The authors declare no potential conflict of interests.

References

1. S.M. Amin, B.F. Wollenberg, Toward a smart grid: power delivery for the 21st century. IEEE Power and Energy Magaz. **3**(5), 34–41 (2005)
2. C.H. Lo, N. Ansari, The progressive smart grid system from both power and communications aspects. IEEE Commun. Surveys & Tutor. **14**(3), 799–821 (2011)
3. N. IqtiyaniIlham, M. Hasanuzzaman, M. Hosenuzzaman, European smart grid prospects, policies, and challenges. Renew. Sustain. Energy Rev. **67**, 776–790 (2017)
4. F. Al-Turjman, M. Abujubbeh, Iot-enabled smart grid via sm: an overview. Fut. Generat. Comput. Syst. **96**, 579–590 (2019)
5. D.W. Griffith, M.R. Souryal, N.T. Golmie, *Wireless Networks for Smart Grid Applications* (Tech, Rep, 2012)
6. T. Sauter, M. Lobashov, End-to-end communication architecture for smart grids. IEEE Trans. Ind. Electron. **58**(4), 1218–1228 (2010)
7. C. Steinbrink, S. Lehnhoff, S. Rohjans, T.I. Strasser, E. Widl, C. Moyo, G. Lauss, F. Lehfuss, M. Faschang, P. Palensky et al.: Simulation-based validation of smart grids–status quo and future research trends, in *International Conference on Industrial Applications of Holonic and Multi-Agent Systems* (Springer, 2017) pp. 171–185
8. M. de Reuver, T. van der Lei, Z. Lukszo, How should grid operators govern smart grid innovation projects? an embedded case study approach. Energy Policy **97**, 628–635 (2016)
9. R. Kappagantu, S.A. Daniel, Challenges and issues of smart grid implementation: a case of Indian scenario. J. Electr. Syst. Inf. Technol. **5**(3), 453–467 (2018)
10. R. Podmore, M.R. Robinson, The role of simulators for smart grid development. IEEE Trans. Smart Grid **1**(2), 205–212 (2010)
11. H.T. Zhang, L. Lai, An overview on smart grid simulator. in *2012 IEEE Power and Energy Society General Meeting*, pp. 1–6. IEEE (2012)
12. W. Li, X. Zhang, Simulation of the smart grid communications: challenges, techniques, and future trends. Comput. Electr. Eng. **40**(1), 270–288 (2014)
13. S. Nithin, K.K. Sasi, N. TNP, Development of a smart grid simulator, in *Proceedings of National Conference on Power Distribution* (CPRI, India, 2012)
14. C. Wang, X. Yang, Z. Wu, Y. Che, L. Guo, S. Zhang, Y. Liu, A highly integrated and reconfigurable microgrid testbed with hybrid distributed energy sources. IEEE Trans. Smart Grid **7**(1), 451–459 (2014)
15. S. Bukowski, S. Ranade, Communication network requirements for the smart grid and a path for an ip based protocol for customer driven microgrids, in *2012 IEEE Energytech* (IEEE, 2012), pp. 1–6
16. V.C. Prakash, P. Sivraj, K.K. Sasi, *Communication Network of Wide Area Measurement System for Real-Time Data Collection on Smart Micro Grid, in Artificial Intelligence and Evolutionary Computations in Engineering Systems* (New Delhi, Springer India, 2016), pp. 163–172

17. J.H. Teng, C.W. Chao, B.H. Liu, W.H. Huang, J.C. Chiu, Communication performance assessment for advanced metering infrastructure. Energies **12**(1), 88 (2019)
18. J. Zhang, A. Hasandka, J. Wei, S. Alam, T. Elgindy, A.R. Florita, B.M. Hodge, Hybrid communication architectures for distributed smart grid applications. Energies **11**(4), 871 (2018)
19. Nithin, S., Sivraj, P., Sasi, K., Lagerstöm, R.: Development of a real time data collection unit for distribution network in a smart grid environment, in *2014 Power and Energy Systems: Towards Sustainable Energy* (IEEE, 2014), pp. 1–5
20. Shahinzadeh, H., Moradi, J., Gharehpetian, G.B., Nafisi, H., Abedi, M.: Internet of energy (ioe) in smart power systems. in *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*. (IEEE, 2019), pp. 627–636
21. H. Shahinzadeh, J. Moradi, G.B. Gharehpetian, H. Nafisi, M. Abedi, Iot architecture for smart grids, in *2019 International Conference on Protection and Automation of Power System (IPAPS)* (IEEE, 2019), pp. 22–30
22. J.C. Hastings, D.M. Laverty, Modernizing wide-area grid communications for distributed energy resource applications using mqtt publish-subscribe protocol, in *2017 IEEE Power & Energy Society General Meeting*, pp. 1–5 (IEEE, 2017)
23. J.Y. Huang, P.H. Tsai, I.E. Liao, Implementing publish/subscribe pattern for coap in fog computing environment, in *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (IEEE, 2017), pp. 175–180
24. F. Bellido-Outeirino, J. Flores-Arias, E.J. Palacios-Garcia, V. Pallares-Lopez, D. Matabuena-Gomez-Limon: M2m home data interoperable management system based on mqtt, in 2017 *IEEE 7th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)* (IEEE, 2017), pp. 200–202
25. I.S. Association et al., Ieee standard for synchrophasor measurements for power systems. IEEE Std C **37**, 1–61 (2011)
26. S. Nithin, S.K. Kottayil, R. Lagerstöm, Direct load control on smart micro grid supported by wireless communication and real time computation. in *Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing* (2014), pp. 1–5
27. M. Lord, D. Memmi, Netsim: a simulation and visualization software for information network modeling, in *International MCETECH Conference on e-Technologies*. IEEE Computer Society, Los Alamitos, CA, USA (Jan 2008). https://doi.org/10.1109/MCETECH.2008.12. https://doi.ieeecomputersociety.org/10.1109/MCETECH.2008.12
28. S. Šandi, B. Krstajić, T. Popović, Pypmu—open source python package for synchrophasor data transfer, in *2016 24th Telecommunications Forum (TELFOR)* (IEEE, 2016), pp. 1–4

DOE to Identify the Most Important Factors to Enhance the Efficiency of Reactive Routing Protocols in MANETs

Pattisapu Manikanta Manohar and Dusi Venkata Divakara Rao

Abstract MANET is one of the prominent main areas beyond the 4G wireless communications. They are infrastructure-less autonomous collection of mobile nodes with self-organizing, self-configuring, and multi-hop capabilities forming a temporary network in which the network's topology varies dynamically due to node mobility. These striking features attracted many application areas, and a significant amount of research has been contributed over the last two decades. In this paper, a new initiative has been taken to study and find the influential factors concerning throughput on reactive routing protocols, influencing MANET's performance using Taguchi Design of Experiments (DOE) through OPNET 14.5 modeler. We considered seven factors to find significant importance: routing protocol, packet size, network size, mobility, pause time, data rate, and node placement. Using Taguchi DOE, there is an advantage of the significant reduction in the number of experiments run. The optimum factor levels from DOE are tested and validated through experimentation. Also, the results obtained from the simulation on the OPNET14.5 modeler are cross validated with the NS2 simulator, and it is observed that the results reveal similar behavior. From the results, it is observed that AODV routing protocol with large packet size, dense network size, low mobility, minimum pause time, low data rate, and random node deployment leads to improved performance related to throughput, delay, and network load when matched with DSR and other factor levels. Major effects with percentages are packet size, routing protocol, and network size are 34.84%, 33.37%, and 22.98%, respectively.

Keywords MANET · OPNET14.5 · NS2 · Taguchi DOE · AODV · DSR

1 Introduction

MANET refers to a set of wireless computing devices that are autonomous and exchange information without any centralized administration, such as mobile phones, laptops, personal digital assistant (PDAs) or homogeneous devices [1]. Such a

P. Manikanta Manohar · D. Venkata Divakara Rao (✉)
Department of CSE, Raghu Engineering College (A), Visakhapatnam, Andhra Pradesh, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_25

network is called as infrastructure-less or ad hoc network. Nodes in these networks are self-configuring, self-organizing, and multi-hop to act as a sender, receiver, and router to carry out information exchange and control information.

Wireless nodes often join and leave the network due to these networks' dynamic nature and switch from one location to another without any fixed topology due to node mobility [2]. These networks' application includes defense areas, emergency rescue and search operations, conferences and meetings, education in remote areas, telemedicine, etc. The list of main challenges associated with MANET includes dynamic topology of the network, bandwidth utilization, routing, frequent link breakages, high mobility of nodes, energy constrained, routing with security and privacy, quality of service, scalability, and robustness [3]. Routing and quality of service (QoS) in MANET are significant challenges [4]. Wired network routing protocols are ineffective, so, researchers developed various routing protocols that fall into three categories: proactive, reactive, and hybrid.

Proactive routing protocols are table-driven routing protocols that save each node's routing information even before they are required. Routing tables are used to store route information and are modified on a regular basis as the network topology changes. Proactive routing protocols are better suited to less dense networks. The remarkable benefit of such protocols is that the routes are always accessible when they are required. The notable disadvantage is that they are not suitable for large networks since each node has to maintain the colossal amount of routing information that consumes most bandwidth while exchanging information [5, 6]. This category's major routing protocols include destination sequenced distance vector (DSDV), optimized link state routing (OLSR).

Reactive routing protocols are on-demand routing protocols, since they do not maintain routing data at nodes if there is no communication. When an originator node wishes to transfer data packets to the target node, it first examines for the existence of a valid path. If this is present, it will send accordingly, or else, it conducts a route discovery procedure to determine a path to the target on-demand. The route discovery is accomplished by flooding the network with route request packets. Compared to proactive routing protocols, the primary benefit of these protocols is that they have less routing overhead. The notable disadvantage is that there was a delay in sending the first data packet during the route discovery process [5, 6]. Major routing protocols used in this category include ad hoc on-demand distance vector (AODV), dynamic source routing (DSR), dynamic MANET organization (DYMO) protocol [4].

Hybrid routing protocols are the synthesis of the best aspects of proactive and reactive routing protocols by reducing proactive routing protocol's overhead control traffic and reducing reactive routing protocol's route discovery time delays by preserving routing details. The primary routing protocol used in this category includes the zone routing protocol (ZRP).

This paper examines AODV and DSR routing protocol's performance using Taguchi Design of Experiments by considering seven factors of considerable importance: routing protocol, packet size, network size, mobility, pause time, data rate, and node placement on Average Throughput QoS metric. Finally, it selects the optimum levels of the factors to enhance the QoS in MANETs. With the Taguchi approach, the results can be obtained in fewer experimental runs [7], and the significant factor along with the optimum level is obtained. The performance metrics are investigated using the main effects plot, response table for signal-to-noise ratio. The experiments are also conducted with the optimum level of factors from DOE being again tested and validated. The results obtained from the simulations run on the OPNET14.5 modeler are cross validated with the NS2 simulator, and it is observed that the results reveal similar behavior.

The paper's remainder is with the following sections—Literature review of the related work, AODV and DSR routing protocols discussion, research methodology employed, simulation process used, result, and analysis discussion. To end, the conclusions and future scope of work are discussed.

2 Literature Review of Related Work

In the last two decades, several studies and recommendations have been made on the different routing protocols and various factors revealing the best suitable factors for the MANET environment and simulator setting, but they have used without prior knowledge of the influential factors. Some researchers studied and found significant factors and optimized values to enhance the MANET's performance through a very limited number of simulations. A large number of simulation runs with different combinations need to be performed to find these influential factors. This consumes a lot of time as well as partial usable of results. Some of the papers listed below in the survey are based on the study of different reactive routing protocols along with different factors involved, and some research papers on finding the important and optimized levels of the factors such as packet size, network size, mobility, routing protocol, etc., has been summarized.

Tolani et al. [8] analyzed the influence of different packet sizes on various routing protocols AODV, DSR, and OLSR with 50 nodes in MANET. They observed that OLSR has less delay, and AODV has high throughput for the same packet sizes using OPNET Modeler 14.5.

Panda et al. [9] analyzes the impact of mobility in different terrain areas and observed the performance through packets delivered using AODV, DSR routing protocols using GLOMOSIM simulator. From the result analysis, with low terrain area and high-density network, packet delivery fraction is high compared to medium, larger terrain size.

Ismail et al. [10] aims to test the AODV routing protocol with various packet sizes in homogeneous and heterogeneous MANET using OMNeT++ network simulator. The results show that the increase in throughput and PDR efficiency is analogous to the increase in packet size with some restrictions, as the packet size exceeds the transport layer to a large extent.

Hakak et al. [11] analyzes the effect of three important factors, namely routing protocol, packet size, and node mobility with Pause time by using performance metrics average end-to-end delay and Average Jitter using QualNet 5.1 simulator.

The results show that the most important factor for optimum Average Jitter is packet size, followed by routing protocol. For Average Delay, routing protocol selection is an important criterion. The research work's contribution concludes that each factor influencing the network performance has to be evaluated rather than comparing the protocols.

Singh et al. [12] analyzes the network's performance with different packet sizes and varying pause times on AODV and DSR routing protocols with performance metrics as throughput, average end-to- end delay, Average Jitter with QualNet 6.1 simulator. From the results, throughput with different packet sizes is better for DSR, and average end-to-end delay and Average Jitter is better for AODV. Research work concludes that the throughput increases, and the average end-to-end delay decreases as the packet size increases.

Mohamed et al. [13] aimed to study and identify the important factors that affect DSDV routing protocol's performance using Taguchi's Design of Experiments technique with performance metric—Packet delivery ratio. According to the results, it was found that the traffic load, followed by the pause time, has a greater impact on the packet delivery ratio and also indicates optimal factors that deliver increased packet delivery.

Mohamed et al. [14] assess the performance of DSR protocol in MANETs for single-performance metric using Analysis of Means (ANOM) and multi-performance metrics using Analysis of Variance (ANOVA). The key impacts of six important factors on two performance metrics were analyzed using Taguchi's Design of Experiment. Optimum factors for single-performance metric and multiple-performance metrics are recommended for best results.

Lee et al. [15] presents the Taguchi approach in investigating the performance of MANETs concerning packet drop rates on AODV and DSR protocols through NS2 Simulator. The main effects of the quantitative factors influencing the routing protocols are studied using Taguchi approach. Most desirable drop rates have been estimated using various combinations of optimal levels of significant factors.

3 Routing Protocols

AODV

AODV [16] is an on-demand reactive routing protocol that enhances the proactive DSDV protocol. It does not keep track of network routes until they are needed, reducing control overhead [17]. A routing table is used by each node in the network to keep track of the routing details of its neighbors. When a node needs to send data packets to a destination, it first looks for the destination in its route table, and if it is not found, it begins the route discovery process by broadcasting a route request packet (RREQ). If the intermediate node is the destined node or next-hop to the destination, it generates a route reply (RREP) packet when it receives the RREQ packet. If a node finds a link break, it will update the information to neighboring nodes by using route error packet (RERR) for route maintenance [18]. To avoid stale or broken routes, as well as routing loops, sequence numbers are used [19, 20].

DSR

DSR [21] protocol is a popular reactive on-demand routing protocol in MANETs. In this, the routes are only preserved between the nodes that want to communicate, reducing control overhead, and bandwidth use. Route caching is introduced in this protocol to minimize the further route discovery overhead. In the route cache, each node stores information about all routes it is aware of. It starts the route discovery process if the route to the destination is not known. A single route discovery process generates multiple paths to the destination and is cached. It does not generate much routing traffic, as well as it avoids routing loops routing. Since the source identifies the entire route sequence to the destination, DSR is a source routing protocol. The sequence of hops that must be traversed is included in each packet header [19].

4 Research Methodology

Qualitative Methodology

To evaluate the designs proposed in this study and choose the most effective evaluation methodology, three assessment methodologies were described.

1. Simulation,
2. Experimental and
3. Mathematical

The mathematical methodology is highly restrictive, which contains assumptions and hypothesis that cannot suit to realistic environments. Researchers would most need accuracy to further enhance their work and perform their experiments on real devices through testbeds. Due to the high cost and lack of versatility, the experimental approach is not feasible and setup these networks; simulation is the most popular method of experimentation [22].

Simulator Selection

The simulations are useful to see that the simulated outputs match closely to reality [23]. They must experience a certain degree of imprecision as a result of simulation. There are many reasons for imprecision including the impact of granularity, mobility models, radio propagation models and simulation size [22]. A simulator is said to be best if it is dependable and realistic [22]. Considering the dependability elements such as granularity and support for mobility, popularity, open-source, and available documentation, three simulators were ranked top in the list, namely NS2, OPNET, and GLOMOSIM [22].

In this paper, we have selected Taguchi's DOE using an orthogonal array since DOE with the full factorial method requires more simulation runs, which requires maximum time to run the scenarios. Also, we have used OPNET Modeler 14.5. We have taken random scenarios executed in OPNET and are validated with the results of the NS2 simulator. It is found that the results comparably match with each other.

OPNETModeler:

Optimized Network Engineering Tools (OPNET), a discrete-event network simulator which is a graduate project proposed at MIT in 1986 and publicized in 2000 [22]. It is one of the most commonly used simulators for commercial use and also provides free license for the academic purpose [24]. It supports parallelism and distribution [24]. OPNET Modeler provides an easy, interactive design, and development environment to study various networks, different models, devices, protocols, and behaviors. Its interface is developed in C [22].

Some of the main characteristics of the OPNET Modeler are described as follows [24]:

- It provides an environment to design new protocols and test in realistic scenarios.
- It increases the pace of research by designing and analyzing the protocols in different networks.
- It is one of the dynamic and powerful simulators in which it has the capabilities of modeling, designing, developing, simulating, and analyzing the different networks.
- It presents easy to use and user-friendly graphical interface.
- It supports analyzing the results with different graphs and numerical display of the values.
- It supports the animation of running the simulation scenario.
- It supports simultaneous (parallel) and distributed simulation

It has the following drawbacks [24]:

- It supports only random waypoint mobility model.
- It does not support the energy model.
- By default, it supports a limited number of protocols

Quantitative Methodology

This paper presents the basic study to find the MANET significant factors concerning throughput in a novel approach by validation. To perform this, it requires two stages; in the first stage, finding the influential factors by Design of Experiments using the Taguchi technique. From that, we get significant factors along with the optimum level of factors. Through the simulation run with optimum factors, it is validated that these values are ideal for increased throughput. In Stage 2, we increased different levels of these significant factors and carried out the simulation for further validation.

5 Simulation Process

Seven factors were considered: routing protocol, packet size, number of nodes, mobility, pause time, data rate, and node placement to find significant among them. Two levels were considered for each factor, i.e., low and high value, according to Table 1.

Using Taguchi's DOE technique, for the performance metric throughput, with criteria as larger the better, the following factors were observed as significant.

Taguchi Design

Taguchi Orthogonal Array L8 Design with 7 factors and 2 levels requires 8 simulation runs as shown in Table 2.

L8 Orthogonal Array

In this array, there are a totally 8 experiments to be conducted, with each experiment is based on the combination of different level values of the seven factors as shown in Table 2. For example, the fifth experiment is conducted by keeping factor 1 at level 2, factor 2 at level 1, factor 3 at level 2, factor 4 at level 1, factor 5 at level 2, factor 6 at level 1, and factor 7 at level 2. These eight experiments are replicated

Table 1 Factor-wise two level values

S.no	Factors	Level 1	Level 2
1	Routing Protocol	AODV	DSR
2	Packet size	256	2048
3	Nodes	20	50
4	Mobility	5	20
5	Pause time	0	900
6	Data rate	2	11
7	Node placement	Random	Grid

Table 2 Orthogonal array with three simulation runs and SNR

↓	Cl	C2	C3	C4	C5	C6	C7	C8	C9	CIO	Cll
	A	B	C	D	E	F	G				SNRA1
1	1	1	1	1	1	1	1	518.50	373.05	408.19	52.4906
2	1	1	1	2	2	2	2	237.32	335.87	22967	48.1836
3	1	2	2	1	1	2	2	8906.98	8698.13	8867.02	78.9120
4	1	2	2	2	2	1	1	8353.84	8423.84	7982.72	78.3254
5	2	1	2	1	2	1	2	543.51	187.07	166.86	46.4541
6	2	1	2	2	1	2	1	156.58	160.67	162.42	44.0733
7	2	2	1	1	2	2	1	385.51	353.30	356.66	51.2299
8	2	2	1	2	1	1	2	340.68	34665	36467	50.8870

three times concerning Average Throughput on C8, C9, C10, and SNR on the C11 column, which is indicated in Table 2.

Taguchi Analysis

C8, C9, and C10 versus A, B, C, D, E, F, and G are indicated in Table 2. From the results, the significant factor with rank is shown in Table 3.

Main Effects Plot for SN ratios

Most influencing significant factors on the performance of MANET with respect to throughput are observed as packet size, routing protocol, and network size (Fig. 1).

The bar graph denoting the percentage of the impact of the factors is shown in Fig. 2.

From the response table in Table 3, the optimum level of factors identified was shown in Table 4

Again, conducting the experiment with the optimum level of the factors yield better throughput 8184.121 with a confidence level of 85%.

Table 3 Response table for signal-to-noise ratio with criteria larger is better is shown below

Level	A	BCD	E	E	E	F	G
1	64.48	47.80	50.70	57.27	56.59	57.04	56.3
2	48.16	64.84	61.94	55.37	56.05	55.60	56.11
Delta	16.32	17.04	11.24	1.90	0.54	1.44	0.42
Rank	2	1	3	4	6	5	7

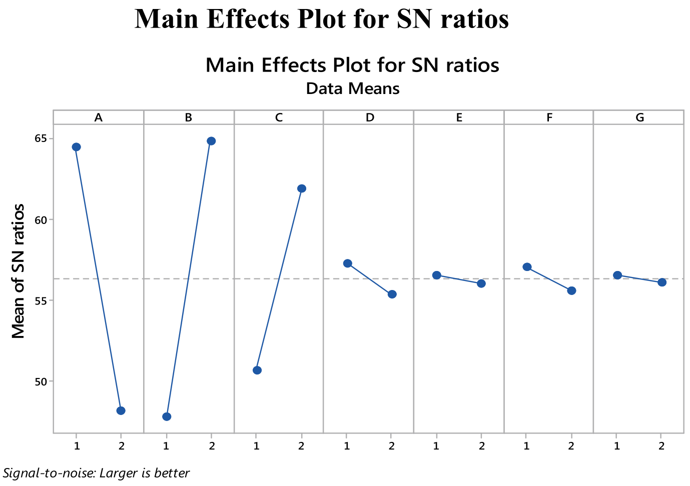


Fig. 1 Graph for main effects plot for signal-to-noise ratios

Fig. 2 Factor-wise percentage of effect

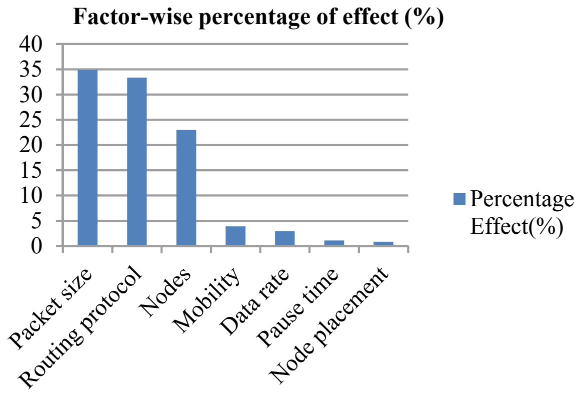


Table 4 Optimum level of factors obtained is shown below

Column no	Factors	Level Description	Level
1	Routing Protocol	AODV	1
2	Packet size	2048	2
3	Nodes	50	2
4	Mobility	5	1
5	Pause time	0	1
6	Data rate	2	1
7	Node placement	Random	1

Table 5 Different levels of significant factors

Factor	Different levels
Routing protocol	AODV, DSR
Packet size	128, 256, 512, 1024, 2048
Nodes	10, 20, 30, 40, 50
Mobility	5 m/s
Pause time	0 s
Data rate	2 Mbps
Node placement	Random
Traffic type	MANET Traffic generator - CBR
Simulation time	900 s
Terrain size	1000 × 1000 sq.mts
Mobility model	Random way point

In Stage 2, considering different levels of the significant factors keeping insignificant factors at the optimum level for further validness, as shown in Table 5.

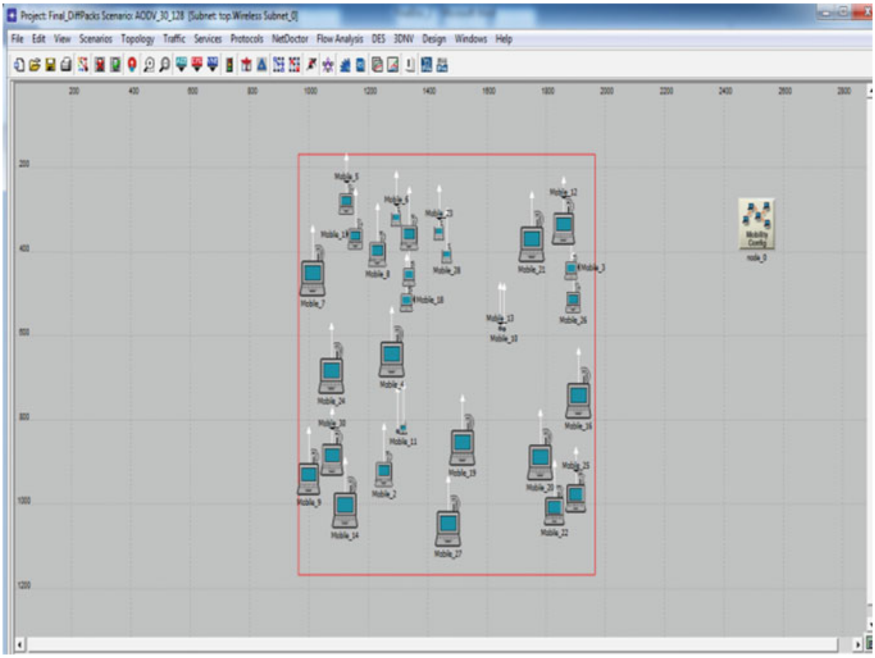


Fig. 3 Simulation scenario for 30 nodes

Simulation Scenarios:

Starting of the scenario for 30 nodes.

Simulation Scenario while running for 30 nodes (Figs. 3, 4 and 5)**.**

6 Results and Analysis

Performance metrics analyzed were Average Throughput, the Average Delay and Load.

Average Throughput: It is calculated by dividing the total number of successfully delivered packets to the destination by the total number of packets sent (Table 6).

The graph shows that the Average Throughput increases with network size and packet size for AODV but DSR, as the packet size and network size increases, Average Throughput remains minimal compared to AODV. Also, it is observed that Average Throughput is maximum at all (50) nodes with packet size 2048 for AODV. So, the significant factors observed are validated from the results. As the packet size and network size increases, the Average Throughput increases for the AODV routing protocol (Fig. 6).

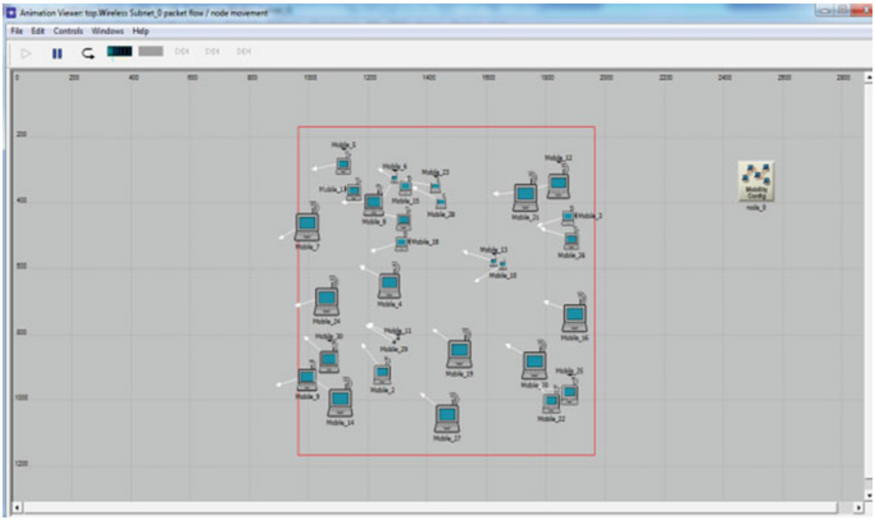
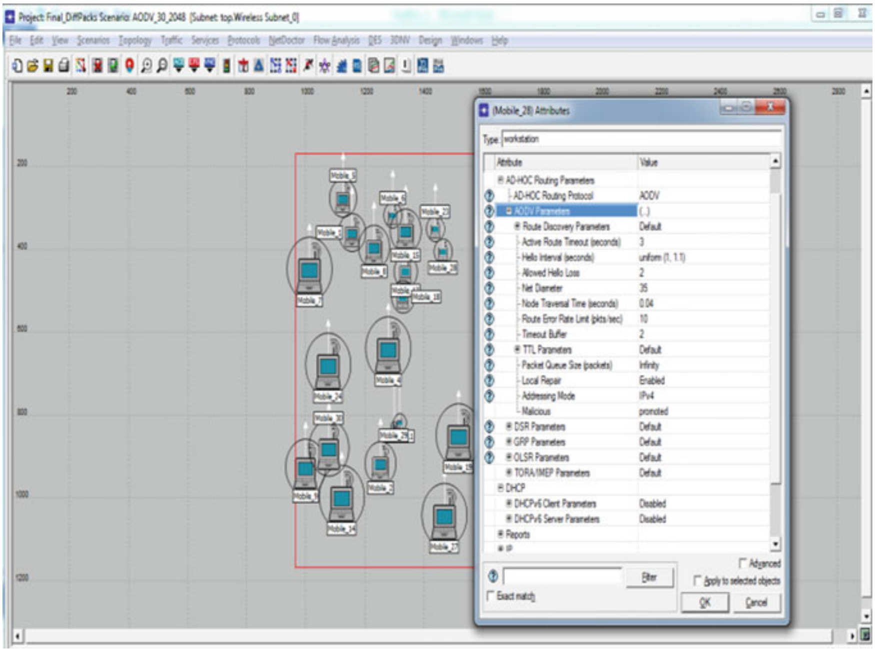

Animation Viewer: top.Wireless Subnet_0 packet flow / node movement
File Edit Controls Windows Help
Mobility Config

Fig. 4 Simulation scenario for 30 nodes while execution simulation scenario parameter setting—30 nodes


Project: Final_DiffPacks Scenario: AODV_30_2048 [Subnet: top.Wireless Subnet_0]
File Edit View Scenarios Topology Traffic Services Protocols NetDoctor Flow Analysis DES 3DNV Design Windows Help
(Mobile_28) Attributes
Filter
Exact match
Advanced
Apply to selected objects
OK
Cancel

Type: workstation

Attribute	Value
AD-HOC Routing Parameters	
AD-HOC Routing Protocol	AODV
AODV Parameters	(...)
Route Discovery Parameters	Default
Active Route Timeout (seconds)	3
Hello Interval (seconds)	uniform (1, 1.1)
Allowed Hello Loss	2
Net Diameter	35
Node Traversal Time (seconds)	0.04
Route Error Rate Limit (pkts/sec)	10
Timeout Buffer	2
TTL Parameters	Default
Packet Queue Size (packets)	Infinity
Local Repair	Enabled
Addressing Mode	IPv4
Malicious	promoted
DSR Parameters	Default
GRP Parameters	Default
OLSR Parameters	Default
TORA/IMEP Parameters	Default
DHCP	
DHCPv6 Client Parameters	Disabled
DHCPv6 Server Parameters	Disabled
Reports	
IP	

Fig. 5 Simulation scenario for 30 nodes—parameter setting

Table 6 Average Throughput for AODV, DSR for 10 and 50 nodes with different packet sizes

Average Throughput (kbps)				
	AODV-10	DSR-10	AODV-50	DSR-50
128	46.04247	17.9068	7991.465	108.5479
256	56.38985	26.21309	8192.285	552.8616
512	76.31225	47.05401	8494.746	282.8761
1024	118.692	89.2896	8292.746	639.2618
2048	198.8089	172.8687	8363.894	666.6343

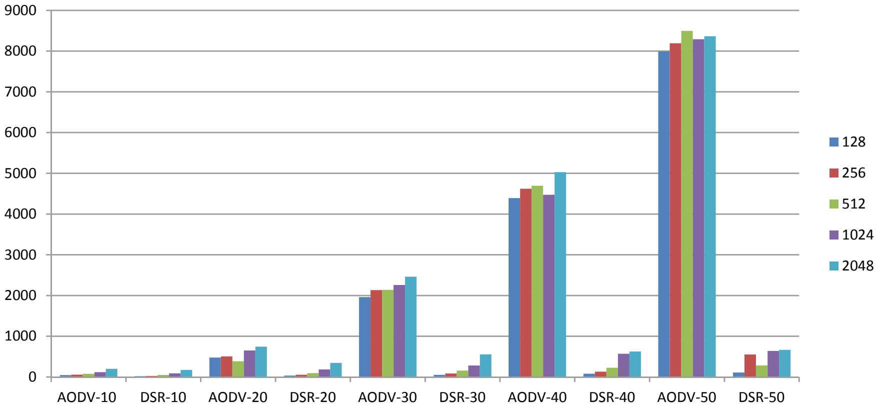


Fig. 6 Variation of Average Throughput for different packet sizes and nodes for AODV and DSR

Average Delay($\times 10^{-3}$): It is the delay in time it takes for packets to travel from source to destination, on average.

As seen the graph that the Average Delay increases with network size and packet size for AODV and DSR, but delay in DSR is higher when compared to AODV. Also, it is found that Average Delay is maximum at 50 nodes and with packet size 2048 for DSR. As the packet size and network size increases, the Average Delay increases. DSR is observed to have maximum delay when compared to AODV (Fig. 7).

Load: It is defined as the average control packets used for the data packets to reach the destination.

As seen the graph that the Average Load increases with network size and packet size for AODV and DSR, but for DSR, load is higher when compared to AODV. Also, it is observed that load is maximum at 50 nodes and with packet size 2048 for DSR. As the packet size and network size increases, the load increases. DSR is observed to have a maximum load when compared to AODV (Fig. 8).

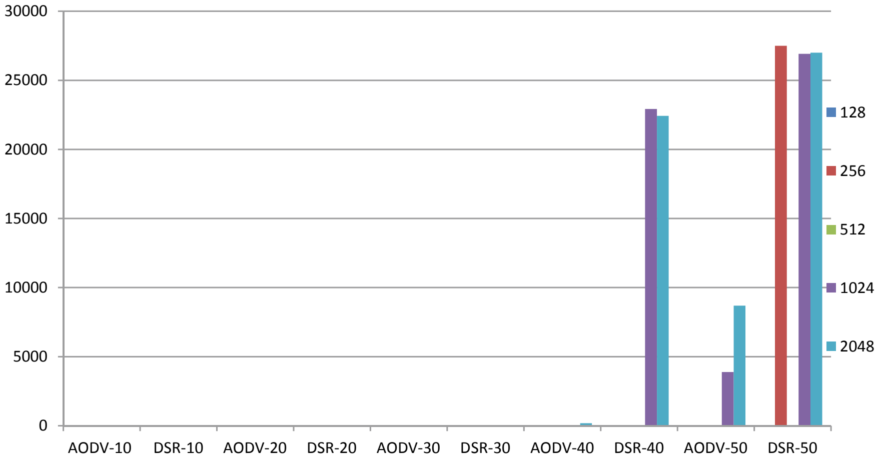


Fig. 7 Variation of Average Delay for different packet sizes and nodes for AODV and DSR

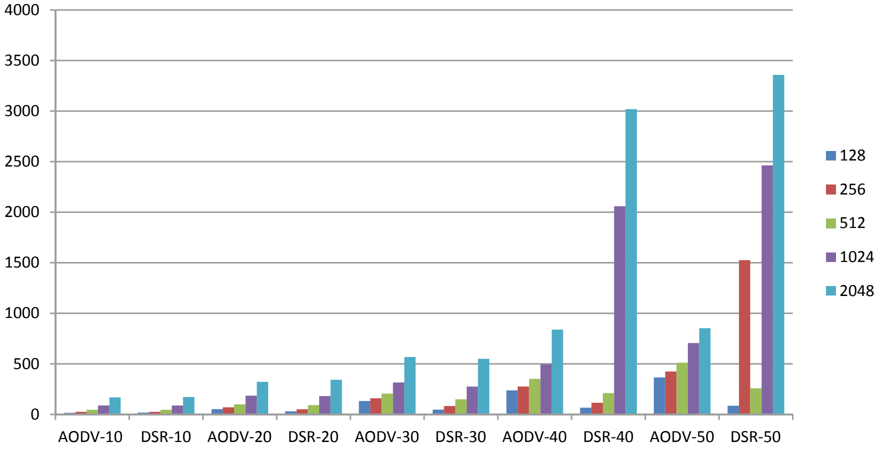


Fig. 8 Variation of Average Load for different packet sizes and nodes for AODV and DSR

7 Conclusions and Future Directions

This work presents the novel way to find the significant factors with respect to the performance metric Average Throughput from a set of factors, also validates the results with optimum factors, and it is found that packet size with 34.84% is most significant, next to that is routing protocol with 33.37% and network size with 22.98%. For further validation, we considered different levels of these significant factors and found that these factors are influencing the performance of a MANET. From the results, it is observed that for large packet size, high network size, and

AODV is the best protocol to be used with less delay and load when compared to DSR. We extend our work further to find the significant optimal level of the factors by using soft computing techniques and also obtain other influencing factors for the performance of a MANET and, finally, choose the ideal parameters to enhance the study for further research.

References

1. Z. Ismai, R. Hassan, Effects of packet size on AODV routing protocol implementation in homogeneous and heterogeneous MANET, in *2011 Third International Conference on Computational Intelligence, Modelling& Simulation* (2011), pp. 351–356
2. A. Khan, T. Suzuki, M. Kobayashi, M. Morita, Packet size based routing for route stability in mobile Ad-hoc networks, in *The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)* (2007)
3. D. O. Jorg. Performance Comparison Of MANET Routing Protocols In Different Network Sizes. *Institute of Computer Science and Applied Mathematics. Computer Networks and Distributed Systems (RVS).* University of Berne, Switzerland, 2003;.
4. V. Attada, S. Pallam Setty, Cross layer design approach to enhance the quality of service in mobile Ad Hoc networks. Wireless Personal Commun. **84**, 305–319 (2015)
5. R. Jhaveri, A. Patel, J. Parmar, B. Shah, MANET routing protocols and wormhole attack against AODV. IJCSNS Int. J. Comput. Sci. Netw. Security **10** (2012)
6. M.S. Chadha, R. Joon, Simulation and comparison of AODV, DSR and AOMDV routing protocols in MANETs. Int. J. Soft Comput. Eng. (IJSCE) **2** (2012)
7. P. Manikanta Manohar, S. Pallam Setty, Taguchi design of experiments approach to find the most significant factor of DYMO routing protocol in mobile Ad Hoc networks. i-Manager's J. Wireless Commun. Netw. 7, 1–11 (2018)
8. M. Tolani, R. Mishra, Effect of packet size on various MANET routing protocols. Int. J. Appl. Information Syst. (IJAIS). Foundation of Computer Science FCS **4**, 10–13 (2012)
9. B. Panda, Manoranjan Das, Impact of mobility and terrain size on performance of AODV and DSR in Mobile Ad hoc network 978-1-4673-1989-8/12, IEEE 2012
10. Z. Ismail, R. Hassan, Effects of packet size on AODV routing protocol implementation in homogeneous and heterogeneous MANET, in *2011 Third International Conference on Computational Intelligence, Modelling& Simulation* (2011), pp. 351–356
11. S. Hakak, F. Anwar, Impact of packet size and node mobility pause time on average end to end delay and jitter in MANETs, in *5th International Conference on Compuer& Communication Engineering*, pp. 56–59 (2014)
12. N. Singh, L. Shrivastava, Effect of change in packet size with different pause time in wireless Ad-hoc networks. Int. J. Comput. Commun. Eng. Res. **3**, 31–34 (2015)
13. H. Mohamed, M.H. Lee, M. Sarahintu, The use of Taguchi method to determine factors affecting the performance of destination sequence distance vector routing protocol, in Mobile Ad Hoc Networks. Journal of Mathematics and Statistics 4 (2008)
14. H. Mohamed, M.H. Lee, Taguchi approach for performance evaluation of routing protocols in mobile Ad Hoc networks, in *Proceedings of the Regional Conference on Statistical Sciences 2010 (RCSS'10)*, pp. 21–29 (2010)
15. M.H. Lee, M. Sarahintu, Performance evaluation of routing protocols for mobile Ad Hoc networks using statistical Taguchi's experimental design, in *Proceedings of the 13th WSEAS International Conference on Applied Mathematics (MATH'08)* (2008), pp. 220–226
16. C.E. Perkins, E.M. Royer, S.R. Das, *Ad Hoc On-demand Distance Vector Routing*. October 99. IETF Draft. 33 pages (1999). http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-04.txt

17. S. Taneja. A. Kush, A survey of Routing Protocols in Mobile Ad Hoc Networks. Int. J. Innov. Manage. Technol. **1** (2010)
18. A. Venkataramana, S. Pallam Shetty, Analyze the impact of transmission rate on the performance of AODV and DSR protocols in MANETs under responsive and non-responsive traffic. Int. J. Comput. Sci. Eng. Technol. (IJCSET) **5**, 516–523 (2014)
19. P. Manikanta Manohar, S. Pallam Setty, Performance analysis of reactive routing protocols AODV, DYMO,DSR,LAR in MANETs. Int. J. Future Revolution Comput. Sci. Commun. Eng. **4**, 1–7 (2018)
20. D.V. Divakara Rao, S. Pallam Setty, Performance comparison of geographic LAR1 with on-demand AODV and DSR routing protocols for MANETs. Int. J. Recent Innov. Trends Comput. Commun. **5**, 782–785 (2017)
21. D.B. Johnson, D.A. Maltz, *The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks*. October 1999. IETF Draft, 49 pages (2017). http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-03.txt
22. L. Hogie, P. Bouvry, An overview of MANETs simulation, in *Elsevier - Electronic Notes in Theoretical Computer Science*, vol. 150 (2006), pp. 81–101
23. D. Cavin, Y. Sasson, On the accuracy of MANET simulators, in *2002 ACM 1-58113-511-4/02/0010. Distributed Systems Laboratory*. EcolePolytechnique Fed´ er´ ale de Lausanne (EPFL).CH-1015 Lausanne. supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS). supported by the Swiss National Science Foundation
24. Manpreet, J. Malhotra, A Survey on MANET Simulation Tools, in *International Conference on Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity(CI PECH 14)* (2014)

IOT-Based Smart Street Light Control Application for Smart Cities

M. S. Muneshwara, M. S. Swetha, R. Anand, S. K. Pushpa, and T. N. Manjunath

Abstract The primary aim of smart streetlights system is to conserve electricity by reducing the wastage of electricity which in turn helps in reducing the manpower. Streetlights are often lit up all night even when unnecessary i.e. when no one is around which wastes huge amount of electricity. Electricity which can be used productively elsewhere. In a manual streetlight system, the streetlights are switched on with full intensity from sunset to sunrise. There is no variation in intensity even when it is not needed. Hence electricity is wasted. This can be avoided by installing a smart street lighting system which can detect when to increase the intensity or completely turn off the lights. This can be achieved with the help of motion detectors which can detect any moving objects like cars, people or animals. Smart streetlights can also serve many other purposes. This system will also be equipped with a depth sensor which can detect flooding in the streets and send data regarding this to a server which can in turn warn the vehicles intending to travel through that area. This would help in preventing accidents.

Keywords Internet of Things (Iot) · Temperature and humidity sensor · Light dependent resistor · Mq7 gas sensor · Ultrasonic sensor (Us) · Esp Wifi module

M. S. Muneshwara (✉) · R. Anand
Department of Computer Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka 560064, India
e-mail: muneshwarams@bmsit.in

R. Anand
e-mail: anandor@bmsit.in

M. S. Swetha · S. K. Pushpa · T. N. Manjunath
Department of Information Science and Engineering, BMS Institute of Technology and Management Bengaluru, Bengaluru, Karnataka 560064, India
e-mail: swethams_ise2014@bmsit.in

S. K. Pushpa
e-mail: pushpask@bmsit.in

T. N. Manjunath
e-mail: manju.tn@bmsit.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_26

1 Introduction

This Light has been the basic foundation for human civilization. The first significant creation that sparked the development of human beings was fire. The discovery of fire helped man fend off wild animals and lead to superiority from other beings and propelled humans to the top of the food chain [1].

In the modern world, streetlights play a very important role in our daily lives. It prevents darkness and gives security. It also extends the hours to which people can work and do productive things as well. It gives vision to drivers which prevents accidents and provides safety [2]. It also prevents unwanted bad activities like robbery, theft and murder. Driving outside daylight can be very dangerous. About 80% of the accidents take place during nighttime which are caused mostly due to visibility reasons. Thus, streetlights play a very important role in modern lifestyle.

Streetlights, although vital to society are yet be modernized. Streetlights require a lot of electricity to run. Streetlights typically consume 18–38% of the total power according to a recent report, which is a huge amount. This is mostly because the majority of the streetlights use traditional sodium vapour lamps and also the ones which are relatively modern like streetlights with LED bulbs also follow traditional methods of usage. Increasing efficiency in this department can lead to huge amount of power saving [3]. Power which can be utilized in other more demanding sectors like factories and industries. This is important especially in developing countries like Nepal and India. This efficiency can be achieved with the concept of smart or intelligent street lighting system [4]. The power usage of the streetlights can be optimized by the implementation of IOT along with other modern technologies. Simple things such as dimming the light when it is not needed and using more energy efficient technologies in terms of lighting can go a long way in conserving energy.

2 Literature Survey

Sritha et al. [1] in the paper has mainly focused on the reducing the wastage of electricity caused by the streetlights by creating a smart system which efficiently lights up the streets without wasting electricity. Here they have modelled the traffic by dividing it into three parts which are road traffic, streetlights and roads with multiple lanes. Different lighting systems are devised for these different conditions and intensity of light is controlled accordingly. In this paper they have also compared the different types of light technologies that are currently in use and the ones which are most efficient and suitable for streetlights. The conclusion of this paper is that with the combination of proper streetlight models like the combination of LOD and LDR can provide lighting to greater areas with significantly less power consumption with helps in saving electricity.

Aziera et al. [2], in the paper have focused mostly on the conservation of electricity by using LED (Light Emitting Diode) street lighting that helps in reducing the power consumption. The prototype design used in this system was built with components like Light Dependent Resistor, Infrared sensors, with a battery and LED. This system also uses motion detectors in order to sense if there are any people or vehicle around and the light intensity can be adjusted accordingly. In this system the main components are LED lamps and controlled sensors. The conclusion of this paper is that electricity can be conserved by using a smart street lighting system created with a combination of efficient lighting and different sensors which help in reducing the electricity consumed.

Omkar et al. [5] in this paper, propose a smart lighting system which controls the intensity of light by using a control circuit which controls the intensity of light based on TRIAC in combination with IOT hence reducing the electricity consumed. With the TRIAC and IOT combined, a smart system can be created which can be monitored and maintained using a simple GUI. NodeMCU, Wi-Fi Module, Intensity Control Circuit, Arduino Microcontroller and Ultrasonic Sensor are the main components used in this system. This system is proven to be economically viable however it may be difficult to implement.

Fathima et al. [3], have based the paper with the primary focus being conserving electricity with the help of an automated streetlight management system. This also helps in reducing the manpower. In a simple street lighting system, the lights are turned on at sunset and turned off at sunlight. This leads to high wastage of electricity. This system uses LED lights which do not consume high amounts of energy in contrast to traditional HID lamps which consume enormous amounts of electricity. Furthermore, LEDs emit directional light and can emit light in specific directions which further increases the efficiency. This system also consists of a DHT11 sensor which measures the temperature and humidity. This gives helpful information about the temperature and humidity in the current surroundings. This system consists of the following components: LDR (Light Dependent Resistors), Arduino Nano, Relay, ESP8266EX Wi-Fi Module and DHT11 sensor. This system mainly emphasizes on the usage of LED lights in contrast to sodium vapor lamps. It is very economical and even environment friendly as it helps in reducing the emission of carbon di oxide gas which is emitted by traditional light bulbs.

Bilam et al. [6] in this system, focuses primarily on not only the conservation of energy but also the reduction of environment pollution. This is achieved by optimizing the light intensity of streetlights with focus on reducing the carbon dioxide gas emission. The principle of this system is to control the intensity based on the amount of luminous energy present at a particular time. The ARDUINO model is responsible for controlling the intensity of the streetlights based on the intensity of ambient light. It also controls the intensity by detecting vehicles on the road. The main components used in this system are: LDR, IR proximity sensor, RTC, LCD module and Microcontroller. The system developed is a cost effective, practical and eco-friendly. It emits less CO_2 compared to traditional streetlights which helps in preserving the environment. However, the maintenance may be costly.

Abinaya et al. [7], have based this paper on smart lighting system in which the intensity of the streetlights is controlled by detecting the amount of sunlight. A camera is fitted in the streetlights which captures images of the environment and sends it to a server. The images are stored in a server. It also consists of a panic button which can be pressed by a person in case of danger or emergency. This notifies nearby police of danger. In the event of the panic button being pressed, footage of the events is sent directly to the cloud account. This system consists of a panic button, sensor, microcontroller, cloud account and a CCTV camera along with streetlight. This system provides additional security along with reduced energy consumption which can be very useful for societies which consist mainly of families. It also helps in crime detection and providing real-time footage which can be helpful in investigation.

Revathy et al. [8] in the paper focus on the reducing energy wastage done by streetlamps in the city. Depending on number of the passer-by's, the intensity of the light can be increased or decreased which can save a large amount of energy. In this system, high intensity lights are replaced with LEDs which can alter intensity based on need. The system also uses LDR which senses vehicles in the proximity and reduces the intensity when not needed. This system also consists of a GSM (Global System for Mobile communication) through which the system is notified if fault is detected in the system. This system when implemented on a small scale showed decent results. The intensity was reduced to 20% for slow moving people like pedestrians or bicycles and the intensity was increased to maximum for cars, bikes and other vehicles. This system is favorable for small scale use, but large scale implementation may prove difficult as it can be complex.

Ravikishore et al. [9], in the paper have proposed a system that consists of LDR sensors in combination with IOT which increases or decreases the intensity of the lighting system to conserve electricity. This system also consists of a low-cost Wi-Fi module ESP8266, light-dependent sensor and ultrasonic sensors. The ultrasonic sensor in this system measures the distance of the object from streetlight and adjust the intensity of the light. The proposed system is very robust however may be a little expensive to implement on a large scale.

Gangyong et al. [10] in the paper have focused on streetlights on safety and conservation of energy. In some cities the visibility is greatly hampered by weather conditions like fog and heavy rainfall. So, streetlights play an important role in providing visibility. SSL system is designed to be implemented in smart cities and it is based on the concept of fog computing. Its advantages are fine management, dynamic brightness adjustment and an autonomous alarm system to report abnormalities or malfunctions in the lamps. The features offered by it are reducing maintenance periods, satisfying fine grain control, decreasing energy consumption and an autonomous alarm system. The advantage of this system is its smart light intensity system however it may be difficult to apply in large scale.

Raju et al., in this paper discusses a smart streetlight control system which can control the intensity of light with the help of an IOT based system. It also includes a panic button to notify if someone is in danger or in an emergency. The system consists of a GSM which helps in sending message to the related authorities [11]. The footage is recorded by a Raspberry pi 8051 CCTV camera which can easily connect to the internet. The digital output of this is converted to analog by using a programming language. It also consists of a Msp430 microcontroller with a 16-bit

CPU which is easy to use and uses low power. The main purpose of this system is to reduce crime and to save electricity [12].

3 Proposed System

In the conventional streetlight system, the energy efficiency is very low due to its various disadvantages, the proposed system a smart street light system created by replacing traditional CFL lamps with LED (Light Emitting Diodes) [13]. In order to increase the efficiency of the streetlights, we use a Light Dependant Resistor which controls the illumination of LED lights according to the requirement. With the help of LDR the intensity of the light can be controlled based on the weather and ambient combinations of the environment [14]. In an LDR, the resistance may vary depending upon the weather condition having 5000 Ohms in daylight and about 20,000,000 Ohms during dark.

The proposed model also consists of Temperature and Humidity sensor which senses the temperature and displays in the LCD monitor [15]. The temperature and Humidity value can also be monitored remotely from BLYNK application.

The model also consists of MQ7 gas sensor which senses the presence of carbon monoxide (CO) gas. If the sensor detects the presence of CO gas it alerts the respective authority to take preventive measures through BLYNK application and the same warning is also displayed on the LCD screen.

The model also contains an ultrasonic sensor at the certain height of the light pole which continuously measures the distance of water level. If the water level exceeds certain level that may cause flooding the ultrasonic sensor sends an alert notification through the Wi-Fi module to the Blynk application to notify the control station from where the respective authority take action to prevent any accidents than can be caused by heavy rainwater (Fig. 1).

4 Components Used for Proposed Method

1. **Ultrasonic sensor:** An ultrasonic sensor is an electronic device capable of measuring an object’s distance by using sound waves by continuously transmitting sound waves at a specific frequency and assessing the distance depending on the time the sound takes to return [16]. It can be paired up with Arduino in order to measure the distance of moving objects which can be implemented in the proposed smart streetlight system. In our system, the ultrasonic sensor is used to determine the amount of rainwater on the roads, i.e. flooding on roads. If the level of water on the roads crosses a certain level it will alert the respective authorities by sending an SMS message through the ESP 8266–01 WIFI module (Fig. 2).

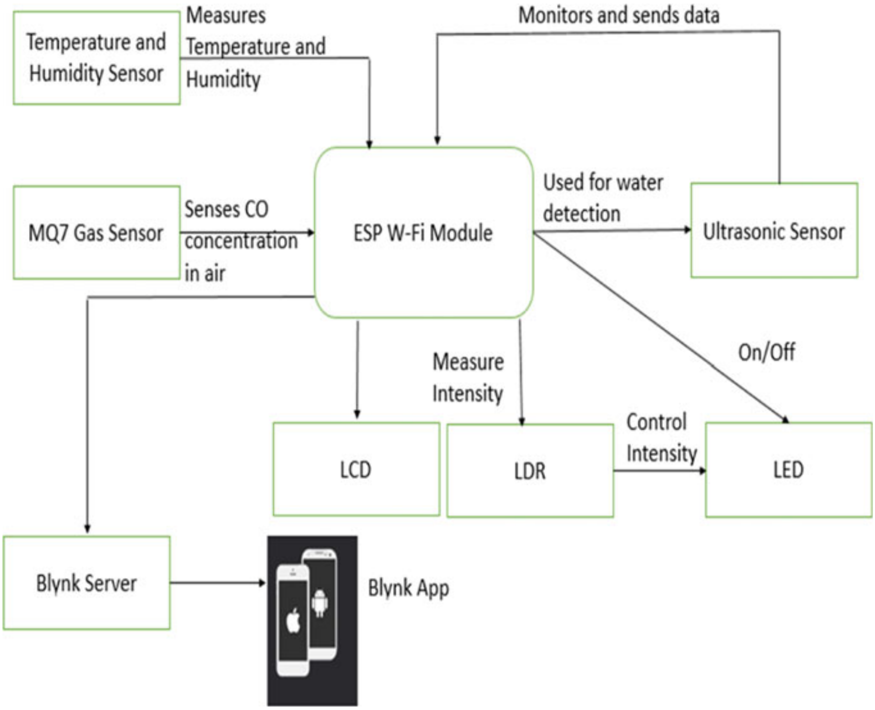
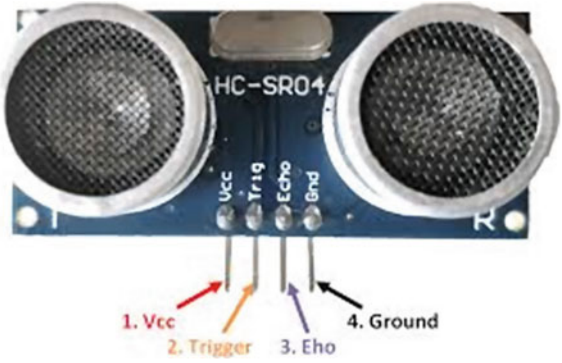


Fig. 1 System architecture

Fig. 2 Ultrasonic sensor



2. **Light Dependent Resistor:** LDR is a simple device in which the resistivity changes according to the electromagnetic radiation in the surrounding area. They change the resistivity based on the amount of ambient light in the surroundings [17]. It can be used to detect day or night in a smart streetlight system. This system helps in detecting the light around the environment and helps to control the illumination of the streetlight (Fig. 3).

Fig. 3 LDR sensor

3. **ESP8266-01 Wi-Fi Module:** The Wi-Fi module is one of the most popular and low cost Wi-Fi system on chip (SOC) low power 32 bit microcontroller [18]. With the help of ESP8266-01module we can send or receive data online or regularly upload data. This module can send data to the remote server. It is based on the TCP / IP protocol stack that can provide access to the Wi-Fi network to any microcontroller. This module is used in our system to send an alert message to the required authorities about the rise in water levels as detected by the ultrasonic sensor module (Fig. 4).
4. **LCD**: An LCD module has also been used in our system. It can be used to display different information such as the current temperature, humidity, the amount of carbon dioxide in the environment and any other user defined messages (Fig. 5).

Fig. 4 ESP8266 Wi-Fi module

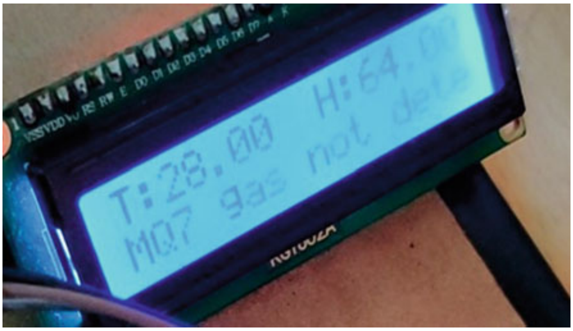


Fig. 5 LCD display

Fig. 6 MQ7 gas sensor

5. **MQ7 gas sensor:** An MQ7 sensor module is also connected with the system [19]. It detects the harmful carbon monoxide gas in the environment and the information about this is displayed in the LCD display (Fig. 6).
6. **DHT-22 Temperature and humidity senor:** The DHT-22 sensor module is also connected to the system which measures the current temperature [20] and humidity of the surroundings, and this information can be displayed on the LCD display (Fig. 7).

5 Results and Discussion

This part shows the entire hardware setup where all the hardware components are connected, and the screenshots of the message received in the Blynk application.

Figure 8 shows the setup of different hardware's. The LCD display, Temperature and humidity sensor, MQ7 gas sensor, LDR (Light Dependent sensor) is connected to the Wi-Fi module. The ultrasonic sensor is mounted at the top of the pole which is also connected to the Wi-Fi module (Fig. 9).

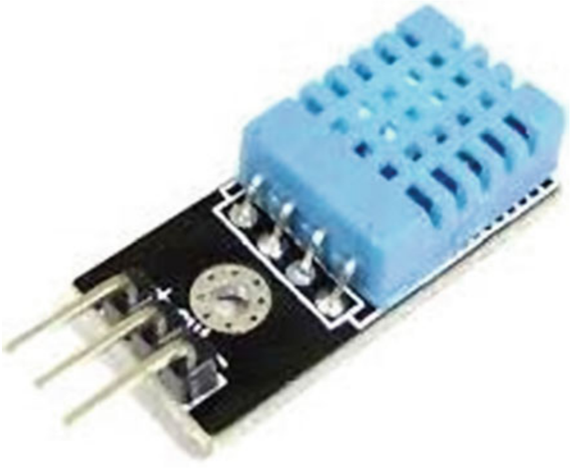


Fig. 7 Temperature and humidity sensor

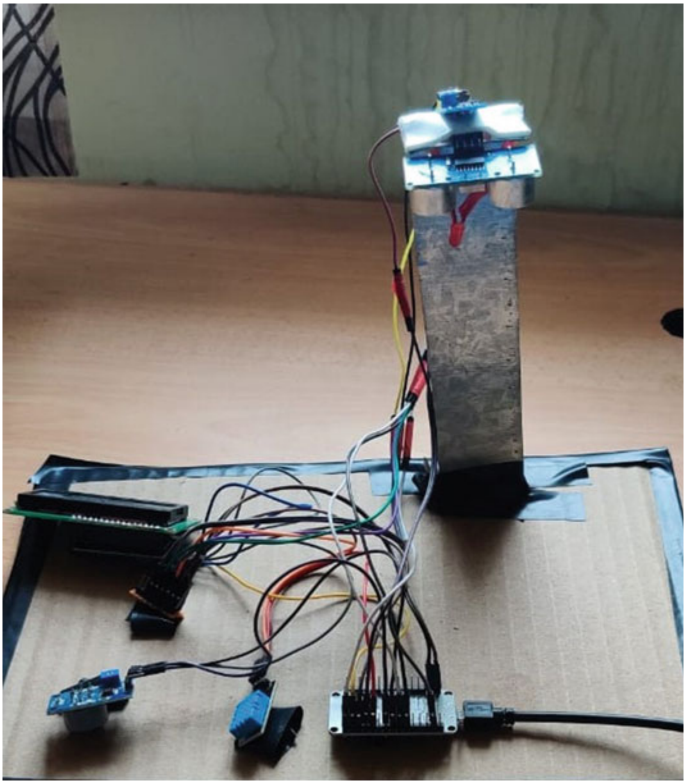
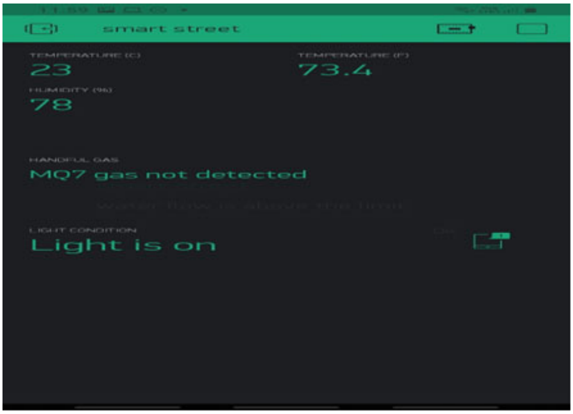


Fig. 8 Hardware setup

Fig. 9 Screenshot of the Blynk application



smart street

TEMPERATURE (C)
23

TEMPERATURE (F)
73.4

HUMIDITY (%)
78

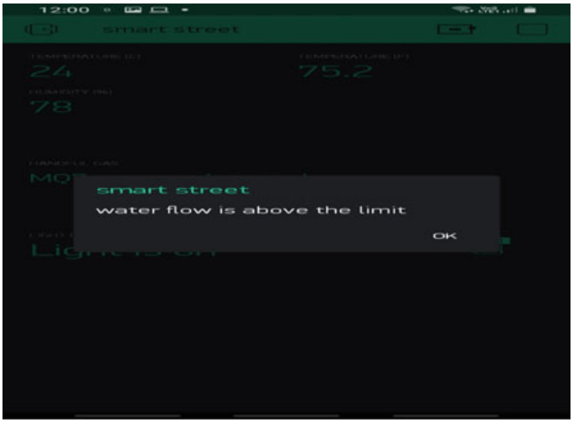
HANDFUL GAS
MQ7 gas not detected

LIGHT CONDITION
Light is on

In the above screenshot we can see different parameters, it consists of Temperature where the value 23 is in Celsius and 73.4 is in Fahrenheit format that can be monitored remotely using Blynk application. This application can also be used to monitor the presence of MQ7 gas, if the sensor detects the presence of CO-carbon monoxide gas it sends an alert to the mobile as "MQ7 gas detected". If there is no harmful gas in the air it shows MQ7 gas not detected. The light condition parameter shows whether the light is on or not. If it shows that the light is off during night-time one can conclude that the light is dis-functioning and not working properly. Thus, the working condition of lights can also be monitored remotely which reduces human dependency for frequently need of surveillant the lamps for frequent maintenance of lights (Fig. 10).

In the above screenshot we can see the alert that reads water flow is above the limit. When the water level increases and reaches a certain limit, the ultrasonic sensor sends an alert via Wi-Fi module to the Blynk application. The message alerts the respective authorities, the control station of the danger of increasing water level

Fig. 10 Screenshot of ultrasonic sensor sending an alert about increasing level of water via Wi-Fi module to Blynk application



smart street
water flow is above the limit

and the authorities can take various steps to warn the people or vehicle about the potential danger and can prevent accidents.

In this method it has been designed a smart streetlight system which has many other functionalities as well. The main goal of the work was to control the illumination of light in order to save energy and achieve higher efficiency and help in building towards a smart city. There are also many other components which provide various functionalities.

1. **Cost compared to LED and CFL:** LED bulbs are initially slightly more expensive than CFL bulbs. A ten-pack of CFL lights costs approximately Rs.1600–1800, but a pack of LED bulbs costs approximately Rs.2200–2600. LEDs have a higher initial cost, but they save money in the long term. A ten-pack of the identical LED bulbs will last approximately eight times as long as the same ten-pack of CFLs. Replacing CFL bulbs eight times costs Rs.14,000, compared to a single purchase of Rs.2600 for a ten-pack of LED bulbs.
2. **Energy Efficiency of LED:** CFLs consume 25–35 percent less energy than incandescent or conventional light bulbs. This is adequate but not exceptional. LEDs, on the other hand, utilise 75% less energy than incandescent bulbs. As a result, LED lamps are extremely energy efficient. Furthermore, CFL bulbs emit nearly 80% of their energy as heat, but LED bulbs emit very little to no energy as heat, increasing their efficiency even further.
3. **Efficiency of ultrasonic sensor:** It takes about 10 s for the ultrasonic sensor to alert the ESP8266 module. The time it takes to display it on the BLYNK app is determined by the internet speed. The proposed method controlled the illumination of the light using the LDR resistor and detected flooding on the roads caused by significant rain, particularly during monsoon seasons. When the water level surpasses a specific barrier level, the ultrasonic sensor transmits a notice message to the authorities. This message can be distributed to the general public to direct them to a safer route. Using a MQ7 module, this approach may also detect the presence of hazardous CO_2 gas.

6 Conclusion and Future Enhancement

Proposed Smart lighting system is achieving higher efficiency in terms of energy consumption. System is also more environment friendly because it uses LED lighting system that has significantly less carbon emission. It also helps in detecting the harmful gases in the environment as well used as alert system that detects the flooding of roads during the rainy season and sends a notification alert to the authorities with the help of an ultrasonic sensor and an ESP 8211 Wi-Fi module. The system also has an LCD display that displays the information such as temperature and humidity as well as the presence of harmful gases as mentioned. This can also be used to display any user defined messages such as advertisements in order to generate revenue as well.

In the future, this lighting system can be further optimized, and the components can be more real-time. When implementing on a large scale better components can be used to get better results.

References

1. S. Bandla, S. Basavaraju, N. Gangrade, Smart street lighting with reduced sensors for sustainable and efficient smart cities, in *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*
2. A. Abdullah, S.H. Yusoff, S.A. Zaini, N.S. Midi, S.Y. Mohamad, Smart street light using intensity controller, in *2018 Technologies for Smart-City Energy Security and Power (ICSESP)*
3. P.P. Fathima Dheena, G.S. Raj, G. Dutt, S. Vinila Jinny, IOT based smart street light management system, in *2017 IEEE International Conference on Circuits and Systems (ICCS)*
4. R. Lohote, T. Bhogle, V. Patel, V. Shelke, Smart street light lamps, in *International Conference on Smart City and Emerging Technology (ICSCET)* (2018)
5. O. Rudrawar, S. Daga, J. Raj Chadha, P.S. Kulkarni, Smart street lighting systemwith light intensity control using power electronics. in *2018 Technologies for Smart-City Energy Security and Power (ICSESP)*
6. B. Roy, A. Acharya, T.K. Roy, S. Kuila, J. Datta, A smart street- light intensity optimizer, in *2018 Emerging Trends in Electronic Devices and Computational Techniques (EDCT)* (2018)
7. B. Abinaya, S. Gurupriya, M. Pooja, IOT based smart and adaptive lighting in street lights, in *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*
8. M. Revathy, S. Ramya, R. Sathiyavathi, B. Bharathi, V. Maria Anu, Automation of street light for smart city, in *2017 International Conference on Communication and Signal Processing (ICCSP)*
9. R. Kodali, S. Yerroju, Energy efficient smart street light, in *2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*
10. G. Jia, G. Han, A. Li, J. Du, SSL: smart street lamp based on fog computing for smarter cities. in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11 (2018).
11. R. Anitha, M. Nishitha, K. Akhila, K. Sai Anusha, G. Srilekha, IOT based smart and flexible lighting in streets. Int. J. Eng. Technol. **7**(2,8), 291–294 (2018)
12. H.B. Khalil, N. Abas, S. Rauf, Intelligent street light system in context of smart grid, in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*
13. H. Ahmad, K. Naseer, M. Asif, M.F. Alam, Smart street light system powered by footsteps, in *2019 International Conference on Green and Human Information Technology (ICGHIT)*
14. C.B. Soh, J.J. Tan, K.J. Tseng, W.L. Woo, J.W. Ronnie Teo, Intelligent street lighting for smart cities, in *2018 IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia)*
15. E. Bingöl, M. Kuzlu, M. Pipattanasompom, A LoRa-based smart streetlighting system for smart cities, in *2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)*
16. M.S. Swetha, M. Thungamani, A novel approach to secure mysterious location based routing for manet, Int. J. Innov. Technol. Exploring Eng. (IJITEE) 8(7) 2019. ISSN: 2278-3075
17. G. Sarraf, M.S. Swetha, Intrusion prediction and detection with deep sequence modelling, in *Security in Computing and Communications. SSCC 2019. Communications in Computer and Information Science*, vol. 1208. Springer, Singapore 978-981-15-4825-3
18. M.S. Swetha, N. Ashwini, A novel approach for weather monitoring system using IoT. Int. J. Innov. Res. Sci. Technol. 7(3):15–23
19. V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, S. Madden, A measurement study of vehicular internet access using in situ Wi-Fi networks, in *Proceedings of the 12th annual international conference on Mobile computing and networking*, pp. 50–61 (2006)

20. G. Han, L. Liu, N. Bao, J. Jiang, W. Zhang, J. Rodrigues, AREP: an asymmetric link-based reverse routing protocol for underwater acoustic sensor networks. J. Netw. Comput. Appl. **92**, 51–58 (2017)
21. Y. Chen, Z. Liu, Distributed intelligent city street lamp monitoring and control system based on wireless communication chip nRF401. Proc. Int. Conf. Netw. Secur. **2**(2), 278–281 (2009)
22. M.S. Muneshwara, M.S. Swetha, M. Thungamani, G.N. Anil, Digital genomics to build a smart franchise in real time applications, in *2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT)*
23. C.N. GireeshBabu, M. Thungamani, S.K. Pushpa, M.S. Muneshwara, *2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*
24. M.S. Muneshwara, V. Haritha, M.S. Swetha, M. Thungamani, Comparison on hyper ledger fabric and hyper ledger composer of block chain technology, in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*
25. M.S. Muneshwara, A. Lokesh, M.S. Swetha, M. Thunagmani, Ultrasonic and image mapped path finder for the blind people in the real time system, in *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*
26. R. Anand, Priyanka, R.M. Patil, Health monitoring in aerospace system. Int. J. Inf. Futuristic Res. (IJIFR) (2017)
27. R. Anand, M. Pushpalatha, R.M. Patil, A social networking for sharing infrastucture resources in the social cloud computing. Int. J. Inf. Futuristic Res. (IJIFR) (2016)
28. M.V. Vijaykumar, P. Jagadish, K. Shryavani, R. Anand, Authorized deduplication in hybrid cloud. IJCSN Int. J. Comput. Sci. Netw. (2016)
29. M.S. Muneshwara, B.R. Rajendra, Intelligent robot positioning system (IRPS) for tracing the contemporary location. IAETSD J. Adv. Res. Appl. Sci. Sci. J. Impact Factor—5.2 Indexed by: Thomson Reuters' Research ID : H-2404–2017 Volume 4, Issue 1, Indexed Journals SCOPUS SUGGESTED JOURNAL ID:50E4CF07B9886F83 UGC APPROVED JOURNAL—JARAS
30. M.S. Swetha, S.K. Pushpa, MS Muneshwara, T.N. Manjunath, Blockchain enabled secure healthcare systems, in *2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*

Recognition of Facial Expression Using Haar Cascade Classifier and Deep Learning

J. Premaladha, M. Surendra Reddy, T. Hemanth Kumar Reddy, Y. Sri Sai Charan, and V. Nirmala

Abstract Every human has emotions in life at all phases of life's moment or instance. These emotions keep changing ever so often and person to person because of the biological reflexes of human brain produced by the nervous system due to the neurophysiological changes which are closely associated with behavior, mood swings; thoughts, etc. are tending to be reflected as an expression on the human face. The application of study of these emotions of humans using facial expressions became very prominent in the fields of biomedical engineering, neuroscience, product screening, marketing, and advertisement, etc. In early works of emotion recognition is for the most part useful for analyzing product reviews or some methodologies by monitoring videos and images in real time. Imparting artificial intelligence to the future of the world where every industry thrives well. In recent years, progressive growth in artificial intelligence is seen especially in the field of machine learning with the predominant subset called deep learning. Mainly, deep learning algorithms made an overriding burgeon in image classification thereby avoiding the complex process in facial recognition. The framework proposed in this paper describes conceptual and theoretical knowledge for human facial emotion recognition with a labeled model and applying a Haar cascade classifier using CNN classification—a class in deep neural networks used in its implementation.

Keywords Image Preprocessing · Facial Emotion Recognition · Haar Cascade Classifier · Data Augmentation · Deep Learning · CNN—Convolutional Neural Network · Adaboost Training · Prediction

M. Surendra Reddy · T. Hemanth Kumar Reddy · Y. Sri Sai Charan
Computer Science Engineering, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamilnadu, India

J. Premaladha · V. Nirmala (✉)
Information and Communication Technology, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamilnadu, India
e-mail: nirmalaveeramani@ict.sastra.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_27

1 Introduction

Human communication is a heterogeneous mixture of both verbal and nonverbal conversations. People react through facial expressions, body gestures, and nonverbal prompts. Facial expressions and emotions are dynamical entities with frequent observable variations. Also, it is more specifically true in the communications of emotions. In fact, studies have shown that a staggering 93% of emotional communication happens either non-verbally or Para-linguistically [1] by the way of facial expressions, gestures, or vocal inflections [2]. This modern era has a raise in massive number of products in each and every industry, which is primitive in the task of publicizing their products. On the other hand, it is quite difficult for the companies based on the rival or crown companies in their particular field of industry. So, there is a need for a support system which is a computerized intelligent assistant. There is also a need for a faster popularization of the product, which is a key requirement. As each time a new business pops up in the market with challenging methods the conventional procedure may lead to a fall in a product of the company results in market risks. To overcome this effect, the company should get familiarized not only in the perspective of product prices and discounts but also on innovation. This proposed system will improvise the product popularization by capturing the customer's emotion towards the product by detecting the six basic and universal facial expressions like sadness, anger, surprise, disgust, fear, and happiness [3] which can give us the desired results accordingly.

Despite all these, there comes a question "How facial expressions recognition reflects in one's company familiarity and some other aspects?" this may act to be irrelevant but it is absolutely relevant and profitable for innovative, developing, newly emerging companies. For Instance, Video Gaming Industry, which is the most exciting tech industry performing photorealistic graphics which in terms has a challenging plan of action and innovative releases consistently. In this gaming industry, we have versatile companies working on different platforms like PC, mobiles, tablets, gaming consoles by PlayStation, Microsoft, Nintendo, etc. [4]. So if a company wants to design a game for any of these platforms that should be user likely means the game should be creative, interesting, challenging, and not boring. In general, these companies should trail the game by recording the video of the player and know about the useful information like enthusiasm, likeness, or boredom from his facial expressions caused by emotions. And the other fields have its own custom of benefits relating to facial expression detection that are overlooked. Automobile industries are also in need of predicting the response of the new launch of a vehicle, security system for drowsiness detection by monitoring drivers face expressions and signs of sleep on duty [5], face recognized car ignition system [6] and these are other challenging factors. In medical and psychological fields this helps for improving brain disorders and behaviors [7]. Among the various fields that can benefit from this technology are automotive, healthcare, and gaming, it plays a vital role as a part in the field of security. It can be used for identifying the individual person in a crowd [8], monitoring citizens for unusual or suspicious behavior [9] by tracking current emotional state of that person. Extending its usage with strongly trained computations can also be used to pre-emptively stop criminals from law enforcement [10] and potential terrorist activity tracking in public places.

The main objective of this paper is to extract the facial features of the human face such as eyebrows, eyes, and lips and later recognizing the facial expression of the person using Haar Cascade Classifier. The classifier detects only the face part from the input image and that particular face part is given as an input to Haar for recognizing the facial expressions. Data Augmentation procedure is done in order to increase the dataset size so that the model can learn every detail from the image. And as a final step where all the low level and high-level features are extracted from the preprocessed input image for assigning appropriate weights to each feature to reach the results of classification as output.

2 Background and Literature Survey

Emotion is a strong feeling make out of one's state of affairs, or relationships with others, tremendously associated with the brain waves. The contemporaneous reaction that occurs during the neurological activity within the brain [11, 12] early research had substantial confirmation of that brain waves reflection is the most real human body reaction. Brain wave emotion analysis is then compared with other emotion analyses. The authenticity and non-immutable credibility of brain waves are notably high. The Federation of society's electroencephalography and clinical neurophysiology distinguishes the brain waves using different frequencies of sizes [13]. These distinguished brain waves have various emotional features, which are meant to be combined as human emotions. In order to find the best facial emotions for the detection, the background research of brain waves considerably plays a major role in our input for the implementation. The brain waves are parts into five major kinds of waves, such as follows in Table 1. This helps to understand the more correlation among the brain waves and its outcome the human emotions where the brain waves are more likely to be the case of emotion features which lies in biological messages when they are differentiated.

The human frame of mind is perplexed. It is the coordinated response of physiology, behavior, and neurological mechanisms. Pail Ekman, an American psychologist confirmed that basic emotions are human physiological responses [14]. The most common human emotions can be divided into six categories namely fear, anger, sadness, happiness, surprise, and disgust [15]. Psychologists have defined the following views on basic emotions.

- **Fear**: The instinctive behavior of a common person who faces some danger in life. Fear can cause instant changes in the human body specifically change in the heartbeat rate, which in turn results in distinct raise of eyebrows, and also the lips are well stretched.

Table 1 Emotion Classification based on the brain waves

Waves	State of consciousness	Frequency (Hz)	Psychological state
δ	Level of Unconscious	1–3	Refers to the deep sleep state which is used to determine the person whether has attained the deep sleep state or not. Delta waves are needed to regain sleep physically
β	Level of consciousness	Low β (13–17)	Beta waves are most commonly associated with concentration where it emits higher energy and positively found increased attention
		High β (18–30)	
γ	Level of consciousness	Low γ (31–40)	Refers to the higher sense of happiness which is mostly related to reducing stress when the energy is released in terms of pressure here
		High γ (41–50)	

- **Anger**: Emotional agitation, being violated, disrespected, or mistreated, can lead to instinctive self-preparedness for its combat response. This response action involves the wide opening of eyes and glaring. Most cases the eyebrows are pulled close down together notably with the lips that are open wide.
- **Sadness**: It is usually defined as the psychological frustration due to loss or failure; the mood is lower meaning. Emotions signs are the lips may either have drawn so tightly or otherwise pouted outwards. The eyebrows angled up to the inner corners.
- **Joy**: It is the psychological state of pleasure, with the meaning of happiness. The signs on the face feature are wrinkles around the eyes with raised cheeks and also the diagonally raised lip corners.
- **Surprise**: By unexpected stimulation in the living environment, resulting in temporary action to stop or freeze. The heavily raised eyebrows are curved and found high with wrinkles horizontally in the forehead.
- **Disgust**: Facing negative stimuli in the environment. Always defined as the feeling of aversion to something offensive which can be found by high proportionality of nose scrunches.

3 Proposed Methodology

This paper explains the multi-staged implementation of recognition of facial expressions using Haar cascade classifier and deep learning where the features are selected without the manual intervention since the neural nodes itself connected to each other and chooses its best feature for the recognition and delivers the desired output with labels.

The first stage is to develop a theoretical and conceptual model. The next step is analyzing facial expressions of humans to relate with its different facial expressions.

The proposed method takes image inputs and detects the face using Haar cascade classifier. Then data augmentation is done. Such duplication of data takes place in the process of image augmentation technique employed with basic methods, i.e., cropping, scaling, rotation, translation of image is done for increasing the datasets to work with which reflects in accuracy terms.

The final step is CNN where the image passes through five layers of CNN with the softmax activation function for human facial emotions is then predicted with the labeled rectangular window. Figure 1 depicts the details of all the phases with the components and its roles.

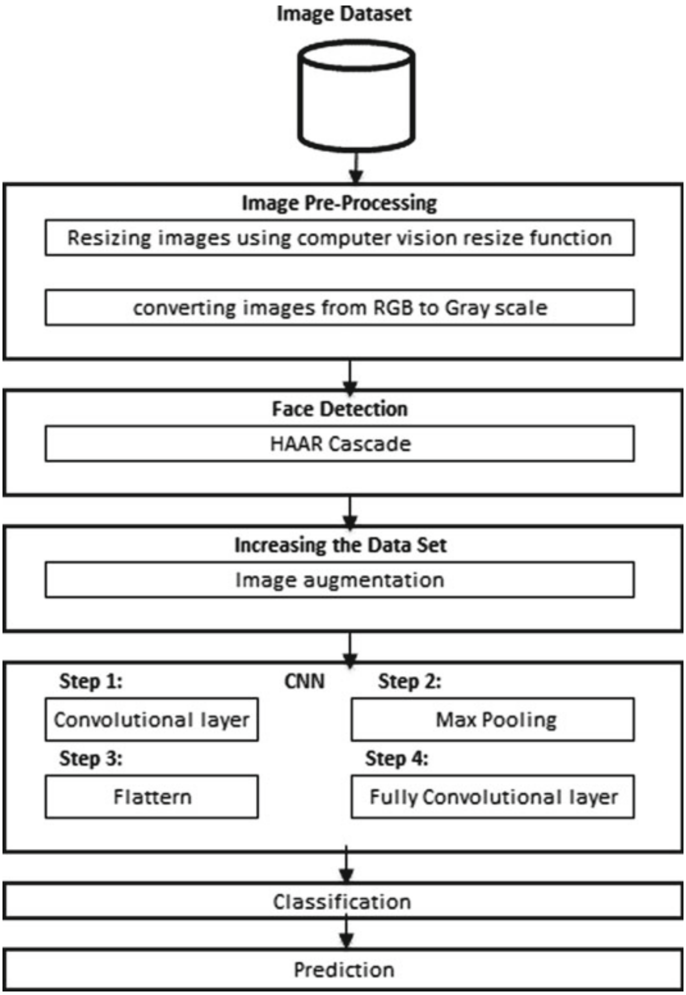


Fig. 1 Proposed model for recognition of facial expression detection using Haar cascade classifier and deep learning

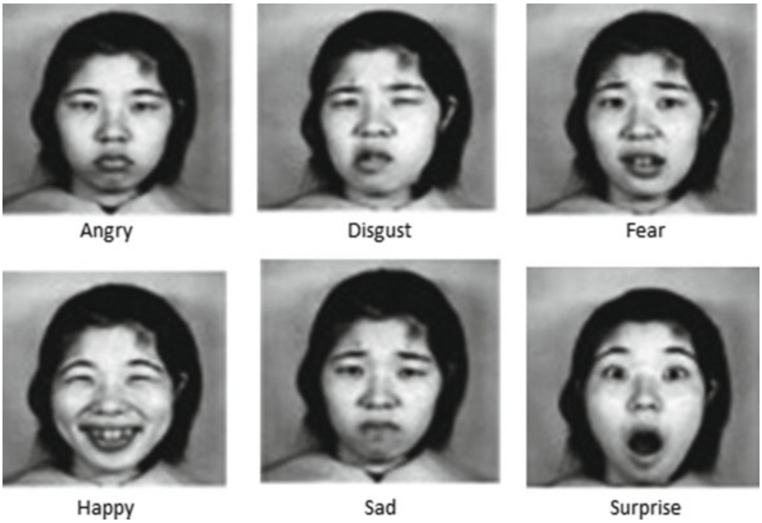


Fig. 2 Different expressions of human emotions

3.1 Image Dataset

In this work, Fig. 2 shows the image dataset of the basic human emotions with labels. In this implementation, the Labeled Faces in the Wild Home dataset of human facial expressions are obtained from the kaggle source [16] where 13,000 samples are taken and the ratio of training and testing is of 10,000 and 3000 samples respectively. It is a well-established world's largest data science community.

3.2 Image Preprocessing

The large volumes of data sets for the construction of a system increase the computational time and the hardware needed to train them in the system. This issue is solved by means of an efficient preprocessing technique. Here, the preprocessing consists of two basic steps involving image resize and conversion of color from RGB to Grayscale. First step which is image resizing is done because to set down the range of size of an image and its pixels or inch perspective which may cause inaccurate results due to its different sizes. This issue is tackled by using the inbuilt computer vision function in python. Once all the resizing is done the second step performs the changing the colors of the image. Here, conversion of RGB to Grayscale is done for providing less information to each pixel. This is also achieved by using the same computer vision library present in python.

3.3 Face Detection

A cascade classifier uses a machine learning algorithm, an idea presented in the research paper by Viola and Jones [17]—An image is extracted from a video or an image by cropping and then the classifier detects the object based on the cascade function which is essentially a pool of negative and positive images. This algorithm is implemented in four different stages:

Haar Feature Selection

Haar identifies and calculates the sum of pixel intensities in each region of a particular rectangular location chosen from a specific location in the detection window. As a result, the difference between the sums is calculated. Edges determine the boundaries between regions in an image by placing sharp discontinuities in pixel values, which helps with the identification of segmentation and entity. Line detection algorithm takes a set of edge points and identifies all the lines on which that edge point's lies. The convolved features are used in the implementation process. Among many line detector techniques, the most popular Hough transformation and convolution [18] are used, it performs transformation of Hough matrix where location of peak values is determined and superimpose a plot on the attribute representing the original image. The different features as shown in Fig. 3 are as follows.

Creation of Integral Images

An Integral image is formed by applying the above Haar features with the resized grayscale images. When we calculate these features, some will be irrelevant. For instance, take line feature on an image as one integral image and edge feature on the same image as for another integral image, same near the nose and eyes region. However, one feature utilizes a property that seems the eyes to be darker than the

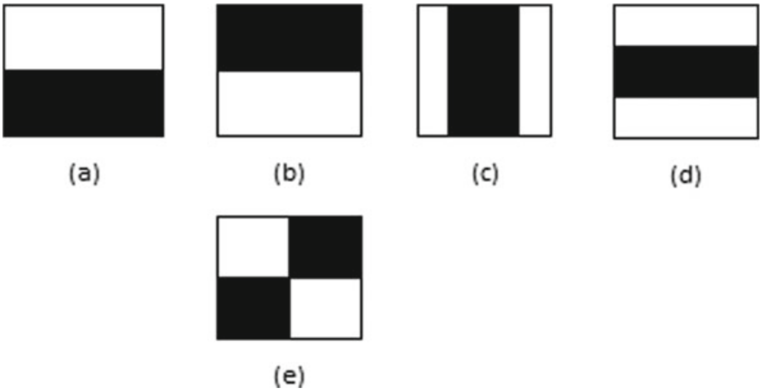


Fig. 3 Convolution and Hough Transformation features (**a**), (**b**) Edge Features (**c**), (**d**) Line features (**e**) Four Rectangle Features

other region, others recognize the region where eyes are to be the darker portion as compared to the other portions of the image. So, we can say that some are irrelevant.

Adaboost Training

Adaboost, a short for "Adaptive Boosting," is Freund and Schapire's first functional boosting algorithm proposed in 1996 [19]. It focuses on classification problems and tries to transform a collection of poor classifiers into a powerful one. Adaboost training can be used to overcome the shortcomings of Integral images. It programs the classifiers to focus on the best features of an image.

Input: Adaboost training algorithm takes a bunch of weak classifiers as input.

Output: Forms a single weighted strong classifier by the summand sequence of weights of weak classifiers.

Step 1: Let $w_t(i) = \frac{1}{N}$ where N denotes Number of Training samples.

Step 2: For t in T Where T denotes chosen Number of Iterations

(a). Pick $\mathrm{h_t}$ the weak classifier that minimizes $\in t$

$$\in_t = \sum_{i=0}^{m} w_t(i)[y_i \neq h(x_i)] \tag{1}$$

(b). Compute the weight of the classifier chosen

$$\propto_t = \frac{1}{2}\ln\frac{1- \in t}{\in t} \tag{2}$$

(c). Update the weights of the training examples

$$w_{t+1}^{i}$$

and go back to step (a)

Step 3: $H(x) = \text{Sign}\ (\alpha 1h1(x) + \alpha 2h2(x) + \ldots + \alpha T\ hT\ (x))$.

To understand the intuition behind w_i^{t+1} the formula is:

$$w_{t+1}(i) = \frac{w_t(i)}{z} e - \alpha^t h^t(x) y(x) \tag{3}$$

At the time of training the data there generates the '*N*', i.e., number of decision trees. Once the very first decision tree is made, more priority is given to the record which is classified incorrectly. And those particular records are the inputs for the second model of the training.

The consequence of Adaboost classifier is the strong classifiers which are divided into levels to form cascade classifiers. The term "cascade" means that the classifier thus produced consists of a set of simpler classifiers which are applied to the region of interest until the selected object is discarded or passed. The records are repeated, and they are allowed with all the boosting techniques.

Cascading Classifiers

Haar feature-based cascade classifiers are effectual machine learning based approach. Employing these ideas creates a "strong" classifier that is comparable to a line combination of weak classifiers. Since Haar is a feature used by a weak classifier, in strong classifiers we use only significant Haar features so that an accurate and precise description of the object is provided into the cascade classifier.

The cascade classifier split up the work of classification into two stages: training and detection. The work of gathering samples that can be classified as positive, and negative is carried out in training stage. Whereas the cascade classifier employs some supporting functions to generate a training dataset and also rates the eminent of classifiers. For training the cascade classifier a set of positive and negative samples are required. The work is incorporated with the utility called opencv_createsamples to create the positive samples for opencv_traincascade. The output file of this particular function provides input to opencv_traincascade to train the detected face. Arbitrary images, where the negative samples are meant to be collected which are not included in the objects to be detected.

Figure 4 shows the workflow of the cascade classifier. Primarily, the classifier was trained with a few positive and as well as negative samples, which are the arbitrary images of same size, in which both the samples were scaled equally in their size. So then the classifier generates "1" if the region likely identifies the face and generates "0" otherwise. The major desire of the cascade classifier is to detect the face objects of interest at various sizes, making the classifier efficient without modifying the size of the input images.

Each stage is edified using a technique called boosting which confers the ability of obtaining an accurate classifier that operates by taking an average of the decisions made by the decision stumps. If a region receives a negative label, the classifier moves the sliding window onto a different location since the classification of previously mentioned location is considered complete. If a region receives a positive label, the region is moved on to the next level of processing by the classifier. In this final stage, the detector announces the detection of an object in the present location of the slider. The region is tagged to be positive and where negative samples are rejected or fired out of the processing as soon as possible.

However, the fact that true positives are time consumed for the process, verification is well suited for meticulous identification. Cases which are resulted, true positive,

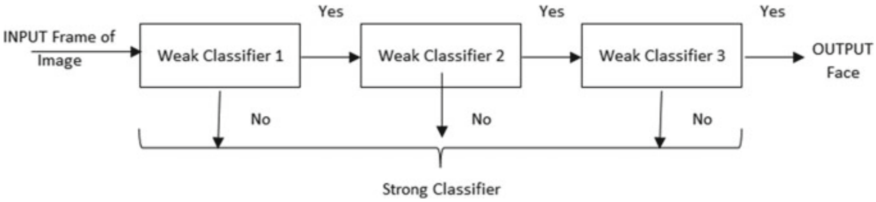


Fig. 4 Classification flow using a cascade classifier

i.e., correctly identified samples. False Positive where the negative samples deemed to be positive and the more important case called false negative, where the real positive sample deemed to be negative which may lead to the greater fall in the term of accuracy.

Efficient functioning of the cascade requires low percentage of false negatives. This can be tackled by increasing the number of stages. However, this also reduces the number of true positives obtained.

3.4 Data Augmentation

Enlarging the image data set is a technique that can be used by creating modified image versions within the data set in order to expand the size of a training dataset. Simply, to increase the diversity of the data available for the training models. The Keras, deep learning neural network library offers the ability to fit models to increase the image dataset using the image data generator class. Data augmentation is the major step for accurate results. Just taking the single perspective of an image will overfit. To overcome this issue duplication of data is performed. So, in image processing, duplication of a particular image is obtained by changing its angle which is considered as another saving of the image. This proof is called image augmentation. Figure 5 A process that artificially creates the images for training in different combinations or by performing random shifts, rotations, tilts, etc. This is done by using some inbuilt libraries of Keras.

Here the process is carried out after the face detection where the detected face is augmented rather than taking the original image for the process to reduce complexity.

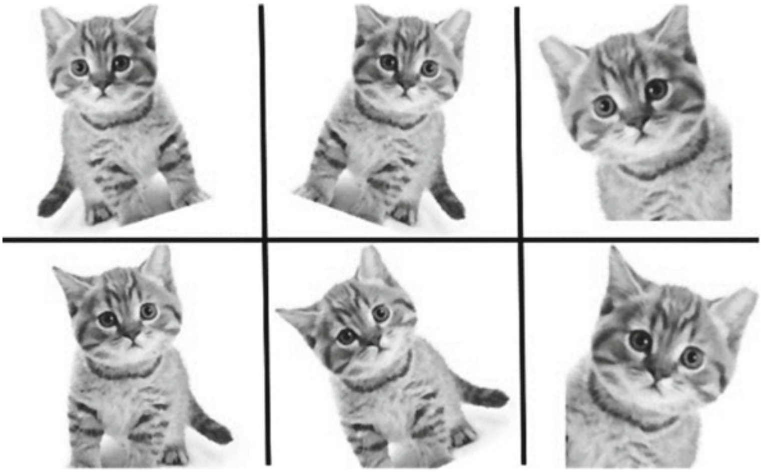


Fig. 5 Data augmentation

3.5 Convolutional Neural Network

A Convolution Neural Network (CNN/ConvNet) Fig. 6 implies the network uses a mathematical operation called convolution that is capable of taking an input image, assigning its significance with learnable weights and biases of various pixels of an image, and being able to differentiate from one another. Also capable of capturing the spatial and temporary dependencies in an image by applying appropriate filters. The main role of employing CNN is to obtain more features. The requirement of preprocessing in the ConvNet is far lower than other classification algorithms. The filters are hand designed in such a way that works in primitive methods.

This algorithm is widely used in many fields and companies. For instance; Tag people in Facebook and Instagram this works on CNN, used in self-driving cars, health care, security, etc. where the network is trained for better understanding of the image. As CNN uses relatively a smaller number of steps in preprocessing which results in comparatively increased efficiency. It indicates that the network is mastering the filters that have been hand crafted in conventional algorithms.

Multilayer perceptrons are regularized variants of CNNs. Multilayer perceptrons typically mean completely linked networks, which implies that each neuron in a single layer is related to all neurons in the next layer. In other words, the network can be trained to better understand the sophistication of the image.

Convolutional Layer

Convolutional layer is an important layer in CNN in which the filters are applied on the input image where both high and low level features are extracted. Convolution operation is performed on the image that changes the function. Some of the filters used in this layer are sharpened filter, edge detection filter, etc. Also, convolutionary layer contains a set of filters that need to learn the parameters. The filter height and weight are smaller than the volume of the input. To calculate an activation map made of neurons, each filter is converted with the input weights and bias.

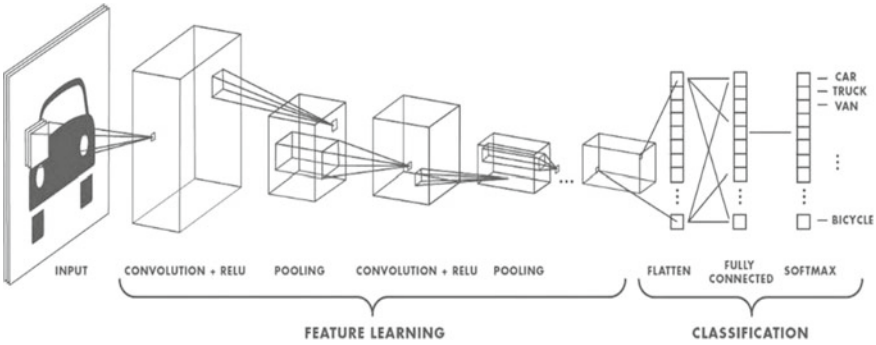


Fig. 6 Deep Convolutional Neural Network

The linear convolution involves the multiplication of a series of weights with the input, much like a conventional neural network. Since the technique was developed for two-dimensional input, the multiplication is carried out between an array of input data and a two-dimensional array of weights, called a filter or kernel. The type of multiplication applied between an input filter of size patched and the filter performs the element wise multiplication between the input and filter of size patched is then summed up, resulting always with a single value.

Padding

Padding is for adding additional pixels outside the image. So zero padding means the value of each pixel you add is zero. If zero padding is equal to 1, the original image will be one pixel thick with a pixel value of 0. Every time we use the filter kernel to search the image, the image size will go smaller and smaller. We don't want that, because we want to preserve the original image size in order to extract some features at low level. So we'll be adding some extra pixels outside the image.

Kernel Size

The kernel in Convolutional neural network is nothing more than a filter that is used to remove the features from the images. The kernel is a matrix that moves over the input data, performs the point product with the input data sub-region, and gets the output as the dot product matrix.

Activation function

Activation functions are termed as the mathematical expression that helps the network to learn the complex patterns in the data as it ends inferring what is to be fired to the next neuron and passes that throughout the network. The task is attached to each neuron, i.e., nodes in the network, and determines whether or not it should be activated or fired depending on whether the feedback of each neuron is relevant to the model prediction. The normalized the output of each neuron to range between '1' and '0' or '-1' and '1'.

ReLu Layer

A linear rectified activation network or a ReLU activation functions in short. The ReLU layer applies the $f(x) = \max\ (0, x)$ function to all values in the input range. Essentially, this layer effectively sets all negative activations to 0. This layer raises the nonlinear properties of the model and the overall network without impacting the receptive fields. The activation feature of the neural network is responsible for translating the total weighted input from the node to the control of the node or output for that information. The rectified linear activation function is a vector piece by piece function that explicitly outputs the input if it is positive, otherwise, it outputs zero.

Max Pooling Layer

A pooling surface of CNN is to gradually reduce the spatial size of the representation in order to minimize the number of parameters and calculations within the network. Max pooling is the most popular method used for pooling. It is necessary to check the

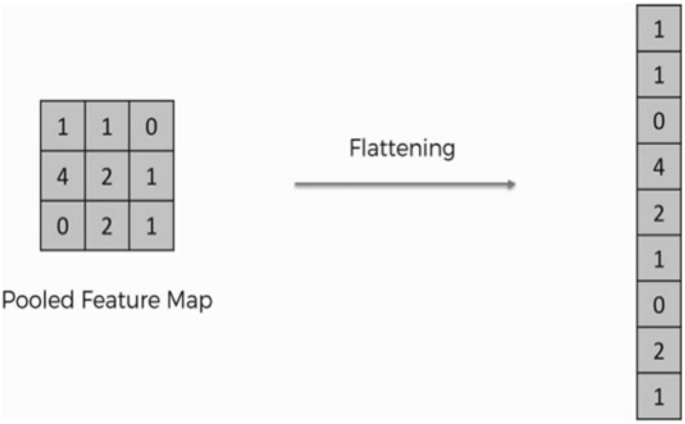


Fig. 7 Flattening

functionality of the map function to fix the vulnerabilities of output function which are sensitive to the position of the input features.

Average pooling and max pooling are two common pooling approaches that contribute to the average presence of the feature and the most active presence of the feature. Average pooling calculates the average value for every patch in the function diagram. Max pooling calculates the highest value for every area of the function chart involves the process of sliding a two-dimensional filter across each feature map channel and summarizing the features inside the filter covered area.

Flatten Layer

There is a 'Flatten' layer in between the convolutionary layer and the completely attached layer Flattening converts a two-dimensional matrix of features into a vector that can be fed into a fully connected neural network classifier. Flattening, Fig. 7. Resulting in a long input data vector which you then transfer through the artificial neural network to have it further processed.

One classifier is the last stage of a CNN. It is called dense layer, which is essentially a classifier of the artificial neural network (ANN). The output of the convolutionary part of CNN must therefore be converted into a 1D feature vector. This is called flattening operation It gets the output of the convolutionary layers, flattening all its structure to create a single long feature vector to be used for the final classification of the dense layer. Flatten layers allow you to adjust the shape of the data from a 2d matrix vector (or n-d matrices actually) into the appropriate format for interpreting a dense layer. This basically allows the data to be managed by a particular type of layer.

3.6 Fully Connected Layer

The contribution from the final pooling is the reference to the completely connected layer, which is flattened and then inserted into the fully connected layer. The flattened vector is then connected to a few completely connected layers that are similar to ANN and perform the same mathematical operations.

The final layer uses the softmax activation function instead of ReLU. The softmax function forces the outputs of each unit to be between 0 and 1, Just like à sigmoid function. But it also divides each resulting value such that the total sum of the outputs is equal to 1. The output which is equivalent to a categorical probability distribution, tells you the probability that any of the classes are true where in some cases of ReLU, the neurons essentially vanish or die for all inputs and remain idle no matter what input is supplied, here no gradient flows and a large number of dead neurons are directly proportional to the performance factor of the neural network (Fig. 8).

After passing through the fully connected layers, this is used to obtain classification. Completely linked layers are an essential component of CNN, which has proved to be very successful in the recognition and classification of computer vision images. The process ends with convolution and pooling; the picture is split down into features and evaluated separately. To determine the most accurate weights, the fully connected segment of the CNN network moves through its own back-spreading method. Rising neurons are given weights that offer preference to the mark that is most fitting.

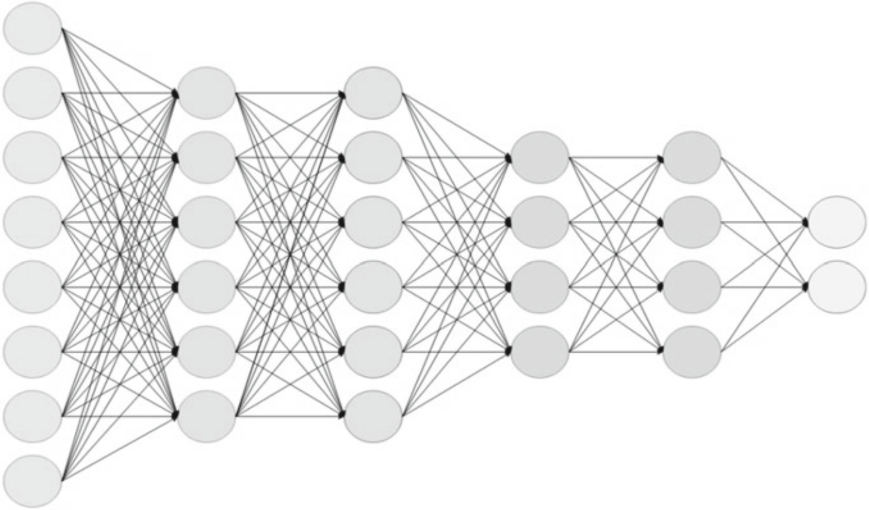


Fig. 8 Fully connected layer

3.7 Softmax Function

The softmax function is a function that transforms the vector of real K values into a vector of real K values which sum up to 1. The input values may be positive, negative, zero, or greater than one, but they are transformed by the softmax into values between 0 and 1, so they can be interpreted as probabilities. It is only applied until the output line, utilizing a neural network line. The layer shall have the same number of nodes as the output layer. Softmax assumes any example to be a member of exactly one class. However, such instances may be a part of several groups at the same time.

4 Results and Discussion

This work helps in classification of images with respect to facial expressions formed due to human emotions which in terms provide wide analysis and extending works based on this implementation may help in resulting successful outcomes for industry, drowsiness detection of drivers, and other popular problem of gesture detection. Work flow of CNN is done and detected the faces using Haar Cascade and achieved 98% as the better accuracy. By the results for true positive prediction rate obtained and for the true negative prediction rate of 87% and 93% respectively (Fig. 9.)

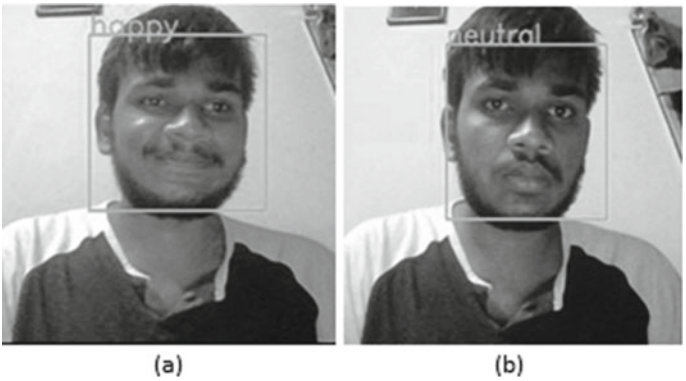


Fig. 9 Sample results for emotions recognition of facial expression using Haar cascade classifier and deep learning **(a)** Happy and **(b)** neutral

5 Conclusion and Future Enhancement

In the proposed model we used Haar cascade to predict five major human facial emotion recognitions. We obtained good accuracy rates with our dataset. Comparatively the prediction of true positive rates is improvised. The future scope is extended in the field of virtual sight with emotion capturing to visually challenged people sophisticated with understanding the emotions of their neighbors, friends, and family members. Further, we can deploy the model into the spectacles facilitating the visually challenged people where the camera is attached to the spectacles to predict the emotions of their peers.

References

1. A. Mehrabian, *The Book "Silent Messages"* (Wadsworth, Belmont, 1971)
2. R.W. Picard, Towards agents that recognize emotion, in *IMAGINA, Actes Proceedings*, Monaco (1998), pp. 153–155
3. H.-W. Ng, V.D. Nguyen, V. Vonikakis, S. Winkler, Deep learning for emotion recognition on small datasets using transfer learning, in *Conference: ACM International Conference on Multimodal Interaction*, Seattle, Nov 2015
4. E. Ghaleb, U. Demir, H.K. Ekenel, A face recognition based multiplayer mobile game application, in *IFIP Advances in Information and Communication Technology*, Turkey, Sept 2014
5. E.E. Galarza, F.D. Egas, F.M. Silva, P.M. Velasco, E.D. Galarza, *Real Time Driver Drowsiness Detection Based on Driver's Face Image Behavior Using a System of Human Computer Interaction Implemented in a Smartphone* (ICITS 2018), Ecuador, Jan 2018, pp. 563–572
6. N. Nagendran, A. Kolhe, Security and safety with facial recognition feature fornext generation automobiles. Int. J. Recent Technol. Eng. (IJRTE) ISSN: 2277-3878, vol. 7, issue-4S, Nov 2018
7. I. Mario, M. Chacon, P. Pablo Rivas, *Face Recognition Based on Human Visual Perception Theories and Unsupervised ANN*, Chihuahua Institute of Technology Mexico, Jan 2009
8. P. Mendis, J. Lai, E. Dawson, H. Abbass, Recent advances in security technology 2007, in *Proceedings of the 2007 RNSA Security Technology Conference*, Melbourne (2007)
9. M.B. Ayed, S. Elkosantini, S.A. Alshaya, M. Abid, Suspicious behavior recognition based on face features. IEEE Access (99), 1–1, Article in IEEE Access October 2019
10. J. Lynch, *The Book—Face Off: Law Enforcement Use of Face Recognition Technology*, February 12, 2018
11. S. O'Regan, S. Faul, W. Marnane, Automatic detection of EEG artefacts arising from head movements, in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. (IEEE, New York, 2010), pp. 6353–6356
12. Z. Lan, O. Sourina, L. Wang, Y. Liu, Real-time EEG-based emotion monitoring using stable features. Vis. Comput. **32**(3), 347–358 (2016)
13. Z. Mohammadi, J. Frounchi, M. Amiri, Wavelet-based emotion recognition system using EEG signal. Neural Comput. Appl. **28**(8), 1985–1990 (2017)
14. P. Ekman, Basic emotions. Handb. Cogn. Emot. **98**(45–60), 16 (1999)
15. P. Ekman, W.V. Friesen, P. Ellsworth, *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*, vol. 11 (Elsevier, Burlington, 2013)
16. T.U. Ahmed, S. Hossain, M.S. Hossain, R. Ul Islam, K. Andersson, *Facial Expression Recognition using Convolutional Neural Network with Data Augmentation*, Conference Paper April 2019

17. P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in *Accepted Conference on Computer Vision and Pattern Recognition* (2001)
18. V. Ra, Y.V. Kinia, K. Manikantana, S. Ramachandran, *Face Recognition using Hough Transform based Feature Extraction* (ICICT) (2014)
19. Y. Freund, R.E. Schapire, Experiments with a new boosting algorithm, in *ML: Proceeding Soft the Thirteenth International Conference* (1996)

Swastika-Shaped Uniplanar EBG Antenna for X-band Applications

Regidi Suneetha and P. V. Sridevi

Abstract In this paper, it is aimed to design an antenna for X-Band applications using a combination of defected ground structure and uniplanar EBG structure to attain a simple, robust, small size antenna at low cost for civil and commercial applications. A novel X-band (8–12 GHz) compact microstrip patch antenna with uniplanar swastika-shaped EBG structure is used to enhance bandwidth, no connecting vias are placed between the ground plane and patch, which makes the structure simple and easy to fabricate. The proposed antenna measured results obtained are −10 dB impedance bandwidth of 1000 MHz from 10 to 11 GHz with a minimum S_{11} value of −30.6 db, and gain of 9.14 db. This antenna finds a wide range of applications in radar, satellite communications and biomedical fields.

Keywords Bandwidth · Defected ground structure · Electromagnetic band-gap structures · Metallo-dielectric EBG · Gain · Impedance bandwidth · Uniplanar · X-band antenna

1 Introduction

In the present world of wireless communication, the design of low profile, flexible, planar antennas with high gain [1] at low cost for various applications is very much required. The congested spectrum in the sub 6 GHz band [2] and millimeter-wave frequency bands is becoming a big challenge, and it is required to design small antennas at a proximity that occupy less volume for civil, commercial and various other applications, at different microwave frequency range. When X-band is considered, it finds applications particularly in satellite communications, and it is also extended from motion capture to tomographic motion capture which can be used for commercial and biomedical applications [3, 4].

The existence of surface waves is the main disadvantage in microstrip patch antennas, it may decrease the gain and directivity of the antenna, the propagation of surface waves [5] can be limited with the help of electromagnetic band-gap (EBG)

R. Suneetha (✉) · P. V. Sridevi
AUCE(A), Andhra University, Visakhapatnam, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_28

structures, and it is an effective way for bandwidth improvement and to reduce the power loss through the substrate. EBG structure presents a smoother radiation profile with low-back radiation, high gain and high efficiency than a conventional ground antenna. It also ensures low interference in the presence of other microwave elements and provides shielding between the antenna and the other communication devices. The perforated dielectric EBG structure and the metallo-dielectric EBG structure are the two types of EBG structures commonly used. EBGs are periodic structures that can limit the propagation of the electromagnetic wave in a specific frequency band.

1.1 Literature Survey

Nagendra kushwaha presented an article on the study of different shapes of EBG structures [6], various methods for identification of the EBG properties and development of various EBG structures are discussed [7] by Md. Shahidul Alam. The quarter-wave EBG structured antenna is presented [8] by Norhana Mat Salleh, and different uniplanar EBG structures are discussed [9, 10]. V N Koteswara Rao Devana presented a compact triple band-notched tapered microstrip-fed ultra wideband antenna for wireless communication applications [11]. Artificial intelligence antenna for fifth-generation mobile networks, a monopole-antenna integrated with the switchable pattern equipped with four antennas with four diverse signals is presented by Abul Bashar [12], Xiaoxuan Gu proposed a novel F-shaped Tx/Rx microstrip antennas with EBG-mushroom structure that shows good radiation performance [13]. Yi Wang proposed a mushroom-shaped electromagnetic-band-gap (EBG) structure with a good wideband radiation performance from 2.64 to 12.9 with two separated notched bands 4.8–5.9 GHz and 7.1–7.8 GHz for modern UWB antennas [14]. Raghavaraju proposed an antenna in the combination of coplanar waveguide feeding and EBG structure for the improvement of bandwidth, obtained dual bands, 1.5–3.6 GHz and 4.8–15 GHz for GPS and X-Band applications [15].

In this paper, a simple swastika-shaped EBG structured antenna for X-band applications is designed and fabricated on Fr4 material with 10 × 10 mm^2 dimensions and the proposed antenna used a uniplanar metallo-dielectric EBG structure which is simple without any vias, unlike perforated structures.

2 Antenna Design

The proposed microstrip antenna geometry and configuration are illustrated in Fig. 1. Following the rules of standard design procedure, a rectangular microstrip patch antenna is designed with a length and width [16] of 4.9 × 4.26 mm^2 with 50 Ω transmission line, design simulations are performed on Ansys HFSS software and fabrication is done on Fr4 material with relative permittivity of 4.4 and thickness of 0.5 mm, that is, used as the substrate of 10 × 10 mm^2 dimensions, defected ground

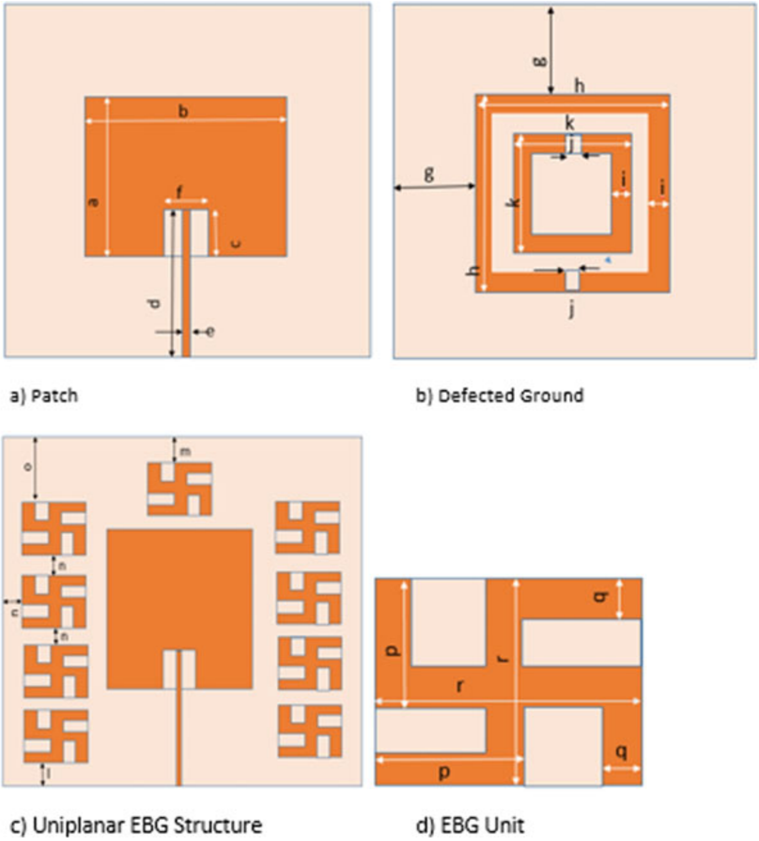


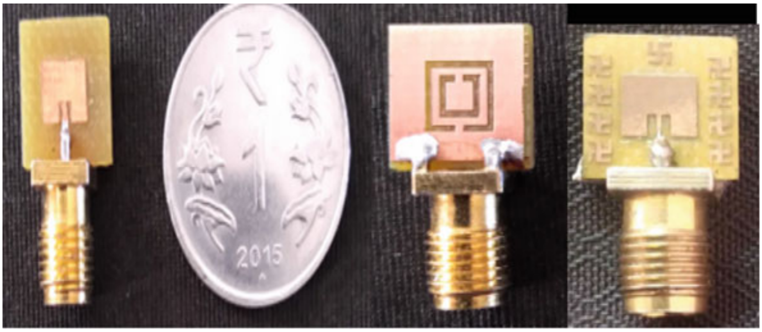
Fig. 1 Designed antenna with patch, defected ground, uniplanar EBG structure, EBG unit (swastika)

structure [17] is used along with the EBG in the design of X-band antenna. The designed patch antenna, with the defected ground and uniplanar EBG structure with the individual unit, is as shown in Fig. 1. The dimensions of the antenna are tabulated in Table 1.

Uniplanar EBG [10, 18] is used to reduce the surface waves and as a result, and gain enhancement can also be achieved. Swastika-shaped EBG structure which is inspired from a usual mushroom-like structure is used, but with the simple difference of no connecting via between patch and ground plane, to decrease the spurious radiations from the vias and also for the ease of fabrication. The individual unit of EBG is 1.5 × 1.5 mm^2 and a total of 9 EBG units with four on either side of the patch and one above the patch are used, and the simple structure of swastika shape provides the ease of fabrication. The fabricated antenna is as shown in Fig. 2.

Table 1 Antenna dimensions

S. No.	Alphabet	Value (mm)	S No	Alphabet	Value (mm)
1	A	4.26	10	J	0.4
2	B	4.9	11	K	3
3	C	1.55	12	L	1
4	D	4.42	13	M	0.5
5	E	0.2	14	N	0.4
6	F	1.62	15	O	1.5
7	G	2.5	16	P	0.75
8	H	5	17	Q	0.2
9	I	0.5	18	R	1.5



a) Patch b) Defected Ground c) Uniplanar EBG

Fig. 2 Fabricated antenna

3 Discussion of Results

The simulated and measured results show good agreement with each other. The simulated and measured graphs of S_{11} versus frequency of the proposed antenna are as shown in Fig. 3. With the inception of the EBG structure, there is a decrease in the S_{11} value along with enhancement in the bandwidth. The minimum value of S_{11} for normal patch antenna and patch antenna with EBG is −25.7 dB and −30.6 dB, respectively.

The simulated and measured graphs of VSWR versus frequency of the proposed antenna are as shown in Fig. 4. The minimum value of VSWR obtained is 1.1, and the value of VSWR lies below 2 in the operating frequency range. The simulated gain is obtained as 9.14 dB as shown in Fig. 5 (Table 2).

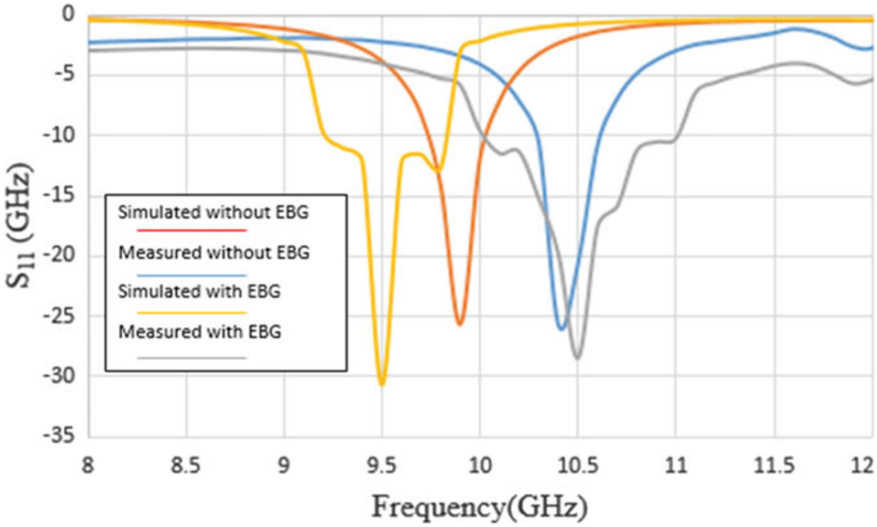


Fig. 3 The simulated and measured graphs of S_{11} versus frequency of the proposed antennas

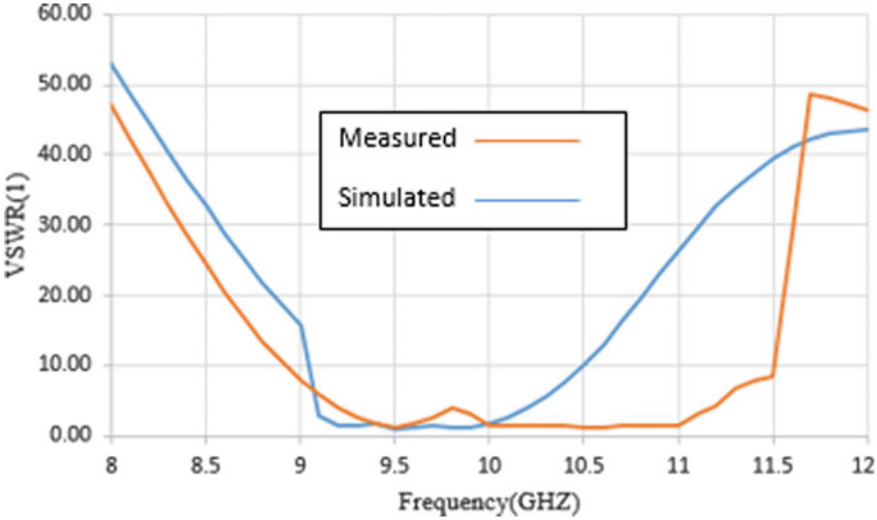


Fig. 4 The simulated and measured graphs of VSWR versus frequency of the proposed antenna

4 Conclusion and Future Scope

It can be concluded from the results obtained and the comparison of various EBG structures from Table 3 that the proposed antenna, a uniplanar swastika-shaped EBG structure is used to enhance bandwidth with no connecting vias placed between the

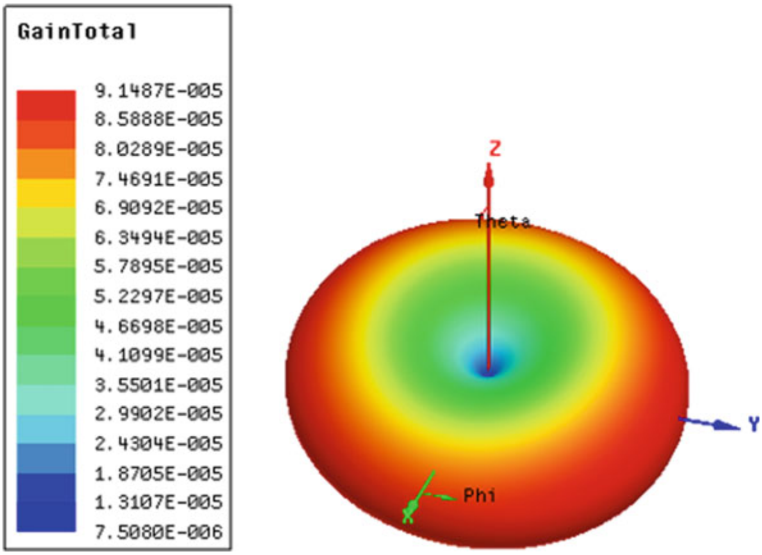


Fig. 5 The simulated gain

Table 2 Comparison table of proposed work

Antenna	S_{11} (dB)	Bandwidth (MHz)	Gain (dB)
Normal patch with DGS (without EBG)	−25.7	300	9.14
Proposed antenna (with EBG)	−30.6	1000	9.14

Table 3 Comparison table of various antennas

Model	Size (mm^2)	Gain (dB)	X-Band bandwidth (MHz)	EBG shape
R [11]	16 × 26	−0.95	(8–8.4) GHz, 400 MHz	U-Shaped slots
R [18]	63 × 63	3.38	(11.2–11.6) GHz, 400 MHz	New structure
R [19]	40 × 40	9.7	(9.875–10.125) GHz, 250 MHz	Truncated Edge (Hexagon shaped FSS)
R [20]	34 × 30	–	(9.6–11.13)GHz, 1.53 GHz	Hexagon shape
Proposed	10 × 10	9.14	(10–11)GHz, 1000 MHz	Swastika Shape

ground plane and patch, which makes the structure simple and easy to fabricate. The defected ground structure is used to achieve miniaturization. The proposed antenna achieved low profile, high gain, miniaturization with a larger bandwidth. An array of the proposed antenna can be used for different applications like MIMO.

Acknowledgements The 'author with Unique Awardee Number MEITY-PHD-2245 would acknowledge Visvesvaraya Ph.D. Scheme, Deity, New Delhi for providing the fellowship and Andhra University college of Engineering (A), Andhra University, Visakhapatnam for providing facilities to conduct research activity.

References

1. D. Qu, L. Shafai, A. Foroozesh, Improving microstrip patch antenna performance using EBG structures. IEE Proc-Microw. Antennas Propag. **153**(6), 558–563 (2006)
2. W. Hong, K.-H. Baek, Y. Lee, Y. Kim, S.-T. Ko, Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices. IEEE Commun. Mag. **52**(9), 63–69 (2014)
3. S.N. Muhammad, M.M. Isa, F. Jamlos, Review article of microwave imaging techniques and dielectric properties for lung tumor detection, in *The 2nd International Conference on Applied Photonics and Electronics 2019* (InCAPE 2019).
4. S.M. Hanna, Siemens medical systems-OCS, Concord, CA, "Applications of X-Band Technology in Medical Accelerators," in *Proceedings of the 1999 Particle Accelerator Conference*, New York (1999)
5. F. Yangand, Y. Rahmat-Samii, Microstrip antennas integrated with electromagnetic band-gap (EBG) structures: a low mutual coupling design for array applications. IEEE Trans. Antennas Propag. **51**(10), 2936–2946 (2003)
6. N. Kushwaha, R. Kumar, Study of different shape electromagnetic band-gap (EBG) structures for single and dual-band applications. J. Microwaves Optoelectronics Electromagnetic Appl. **13**(1), 16–30 (2014)
7. M.S. Alam, N. Misran, B. Yatim, M.T. Islam, Development of electromagnetic band-gap structures in the perspective of microstrip antenna design. Int. J. Antennas Propag. (2013)
8. N.M. Salleh, I.M.M. Yusoff, A.A. Azlan, M.T. Ali, The design of quarter-wave EBG antenna, in *RFM 2018—2018 IEEE International RF Microwave Conference Proceedings* (2018), pp. 266–269
9. N. Jaglan, S.D. Gupta, E. Thakur, D. Kumar, B.K. Kanaujia, S. Srivastava, Triple band-notched mushroom and uniplanar EBG structures based UWB MIMO/Diversity antenna with enhanced wide-band isolation. AEU—Int. J. Electronics Commun. **90**, 36–44 (2018)
10. G.S. Karthikeya, M.P. Abegaonkar, S.K. Koul, Low cost high gain triple-band mm-Wave Sierpinski antenna loaded with uniplanar EBG for 5G applications, in *2017 IEEE International Conference Antenna Innovative Mod. Technol. Ground, Aircraft and Satellite Applications iAIM 2017* (2018), pp. 1–5
11. V.N. Koteswara Rao Devana, A. Maheswara Rao, A compact 3.1- 18.8 GHz triple band notched UWB antenna for mobile UWB applications. IRO J. Sustain. Wireless Syst. **02**(1), 1–12 (2020)
12. A. Bashar, Artificial Intelligence based LTE MIMO antenna for 5th generation mobile networks. J. Artif. Intell. Capsule Netw. **02**(03), 155–162 (2020)
13. X. Gu, S. Liu, X. Kong, Q. Yu, R. Yang, Y. Xu, A novel F-shaped hybrid isolator employed by Tx/Rx antennas to improve antenna performance, in *International Symposium on Antennas and Propagation (ISAP)* (2019)
14. Y. Wang, T. Huang, D. Ma, P. Shen, J. Hu, W. Wu, Ultra-wideband (UWB) monopole antenna with dual notched bands by combining Electromagnetic-Bandgap (EBG) and Slot Structures, in *IEEE MTT-S International Microwave Biomedical Conference (IMBioC)* (2019)
15. A. Raghavaraju, T.V. Ramakrishna, T. Bhavani, B.T.P. Madhav, Investigation on EBG structured CPW fed CM antenna for WiMAX, WLAN applications, in *International Conference on Vision Towards Emerging Trends in Comm. and Networking (ViTECoN)* (2019)
16. Z. Wang et al., An accurate edge extension formula for calculating resonant frequency of electrically thin and thick rectangular patch antennas. IEEE Access **4**, 2388–2397 (2016)

17. M. Elhabchi, M.N. Srifi, R. Touahni, Bandwidth enhancement of trapezoid antenna using an open F-rotated and I-shaped slots Defected Ground Structures (DGS), in *2019 International Symposium on Advanced Electronic Communication Technology ISAECT 2019* (2019), pp. 9–12
18. M.I. Zaman, F.T. Hamedani, H. Amjadi, A new EBG structure and its application on microstrip patch antenna, in *15th International Symposium on Antenna Technology Applications Electromagnetics ANTEM 2012* (2012) pp. 1–3
19. M. AlyAboul-Dahab, H.H.M. Ghouz, A.Z.A. Zaki, High gain compact microstrip patch antenna for X-band applications. Int. J. Antennas (JANT) **2**(1) (2016)
20. A.S. Nagulpelli, D. Varun, Hexagonal shaped magnetically coupled EBG for X-band antenna bandwidth enhancement, in *3rd International Conference for Convergence in Technology (I2CT).* Apr 06–08, 2018

Hurdle Dodging Robot Using Lidar with a Built-On Gas Detection System

B. M. Ragavendra, R. Gogul Sriman, H. Saiganesh, K. S. Saileswar, R. Santhanakrishnan, and C. B. Rajesh

Abstract Motor robots are the machines which helps humanity to move forward where it can do some works that we humans can't do, for example it can reach places regardless of dangerous or poisonous environment. There are various sensors available in the market to scan and detect objects but in recent times, LIDAR is proven to be the best in terms of its characteristics in a closed environment, but other usual standard sensors have their restraint in phase of distance recognition and clarification of complexity. In this research paper we are making an autonomous motor bot which is equipped with a 2D 360° LIDAR sensor with Raspberry Pi 4 to avoid obstacles. The main purpose to avoid the obstacle here is to obtain data from hazardous surroundings. So the idea was, what if a robot in a closed environment can reach a final point by avoiding the obstacles without any pre marked line to reach the final marked point, where the real time mapping of the bot would be displayed in the system.

Keywords Obstacle avoidance · 2D LiDAR · Mapping · Hazardous gas detection · Webot · Robotic operating system · Robot

1 Introduction

Worldwide there are a lot of industrial accidents happening and many people lose their lives, the main cause of the deaths is due to the inability to assess the situation inside properly and quickly. This problem mainly happens in the chemical industries and the amount of gas released in the accident area will not be known.

When there is an explosion in the industry or any situation which needs a further inspection to take necessary steps where humans can't enter, we use LiDAR based robots to monitor the situation and give the information of the surroundings and addition to it we also use gas sensor to find whether toxic or harmful gas is present in the

B. M. Ragavendra · R. Gogul Sriman · H. Saiganesh · K. S. Saileswar · R. Santhanakrishnan · C. B. Rajesh (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb_rajesh@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_29

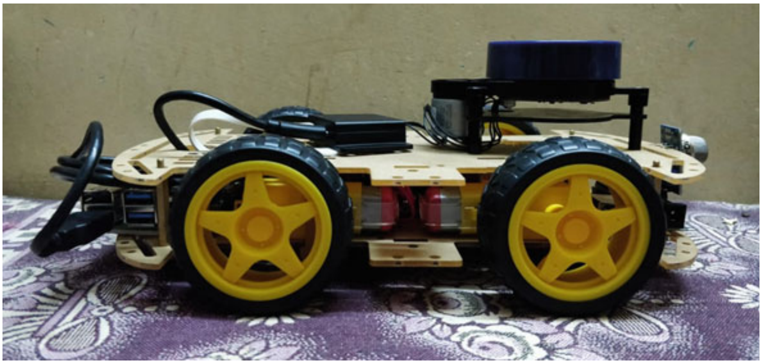


Fig. 1 The robot

surroundings. LiDAR is the main component for the autonomous robot. According to the usage we have the necessary equipment equipped along with lidar and the gas sensor.

LiDAR technology can be used to capture the structure of the place. This digital information can be used for mapping, which can be used to create models of the structure. The technology uses eye-safe laser beams to create a representation of the surveyed environment. So, this can be used in most parts of the surroundings. The obstacle will also be shown in the map. The path of the robot is programmed in such a way that it avoids the obstacle and takes every point in the room including obstacles while mapping. The chemical gas will differ from the main area of the accident and the surroundings of that area so the gas sensor will record the readings of the chemical gas for each and every second continuously. The map and the amount of chemical gas present can be known outside from time to time (Fig. 1).

2 Related Works

Dony Hutabarat, Muhammad Rivai, Djoko Purwanto, Harjuno Hutomo proposed to avoid obstacles, an automated mobile robot with a lidar sensor was created in this research. The robot's movements are navigated using the Braitenberg vehicle technique. On a single Raspberry Pi 3 model computer board, the sensor data collection and control algorithm were applied. The findings of the experiments show that lidar can calculate ranges between 0.12 and 10.5 m with a 0.9% error rate. The colour of the sample and the amount of ambient light have no impact on the lidar measurements. lidar is an intelligent monitoring device that detects distance and other information from a point by measuring the properties of radiated light. One technique for measuring the distance of an object is to shoot beams of laser beam onto the surface. In vacuum, light flies at 300,000 km/s [1].

T R Madhavan, M Adharsh studied and employed a low-cost Slam tech 2-D lidar device for obstacle collision avoidance. This paper presents an efficient algorithm for implementing 2-D lidar-based obstacle detection and avoidance. A method for obtaining data from the lidar sensor is proposed, as well as an algorithm for obtaining obstacle information through filtering and clustering. The paper makes use of a 2D RP Lidar method. The algorithm was created in such a way that the hurdle recognition and avoidance are accomplished using a 2D lidar device with a single measurement plane. The lidar data is screened, processed in advance, then clustered before being used [2].

Huixu Dong, Ching-Yen Weng, Chuangqiang Guo, Haoyong Yu and Ming Chen proposed that in order for a robotic arm to accomplish functions such as transporting barriers in home and commercial environments, it must have the potential to prevent them. When a robot and an obstacle both pass, the current avoidance strategy based on TEB local predictor and cost-map supported by Robotics Os (ROS) cannot achieve excellent results. Furthermore, these tests demonstrated that the suggested. In a cluttered world, the system is resilient and attentive. In the meantime, it will integrate the 3D sensing approach with other technologies. Recognize the challenges that are specifically depicted by complete 3D models. To boost the navigational capabilities of robots [3].

Yan Peng, Dong Qu, Yuxuan Zhong, Shaorong Xie and Jun Luo Jason Gu proposed that Here the filtering, preprocessing, and clustering of the beam cloud, the finding of obstacle system according to SICK LMS511 open-air lidar, could effectively filter noise from raw laser information, fragment, and group the laser-point cloud. This paper suggests a useful stumbling block. Centered on 2-D lidar, a detection and object tracking algorithm has been developed. A system for obtaining knowledge about obstacles is suggested. The data first from beam cloud is filtered and clustered [4].

Zoltan Rozsa proposed a methodology for intelligent transportation systems and autonomous vehicles, lidar devices allow object and free-space tracking. The whole article shows a lidar identification system based on a small number of detection planes. To define a scan plane, we used the Fourier descriptor, and to classify it, he used the (CNN) Convolutional Neural Network. If multiple scan planes are accessible, our system uses both time varying shape details and contours from multiple scan planes. In the case of near field, the program handles at least and even some state-of-the-art techniques, but this also extends the detection spectrum. They tested these on hundreds of thousands of observations across vast public databases, and we even tested it separately against far distance objects. We will use a LIDAR sensor with some of those detecting planes in the rest of this paper. The suggested data preprocessing measures are presented first, followed by the classification system in this paper. The point cloud obtained during each complete 3D scan cycle of the LIDAR sensor would be referred to as one frame throughout the preceding [5]. Similar works are available in [6, 7].

Andrei Fenesan, Daniel Szöcs, Teodor Pana, Wen-Hua Chen proposed an article, an electric concept vehicle is shown for its own data storage and computing system. Data, on which we'll put the collision avoidance to the test. Indoor and outdoor algorithms for GPS localization. The Potential Field approach was used in this research. Then opted for a lidar sensor capable of acquiring distance data from a two-dimensional world. As a result, a more detailed investigation is likely to be fruitful in future research. The read data from the lidar sensor is precise, and it will be used to construct a map of the surrounding atmosphere and objects. The lidar sensor's validation and analysis helped us to understand its characteristics as well as the replay mode of sensed distances [8, 9].

Kai-Tai et al., Song proposed an onboard robot, the lidar SLAM optimization model is carried out in a ROS device design that uses Cartographer SLAM and adaptive Monte Carlo optimization. This paper proposes an optimized guidance system that combines obstacle avoidance and SLAM to allow the robot to reach the desired location without crashing with any unnecessary obstacles. The robot's localization is built on the Cartographer Algorithm, which creates a map of the surrounding area. Goal navigation is combined with an obstacle avoidance controller that operates in real time to give the robot autonomous navigation motion control [10].

In this paper Steven T. Padgett and Aidan F. Browne discuss how to bring all systems together on a research platform and how to use them in a vector-based intrusion detection system. A controller is designed to stick to a wall that has doors, some identifying and protruding characteristics from columns. When the robot comes across an object in its direction that is larger than a certain height, it takes steps to avoid clashing with it. The robot may be designed to follow walls through rooms or corridors with up to half-foot differences in the wall and gaps close to 40 inches to handle open or close doorways and narrow columns. When the robot comes across an obstacle that protrudes or more than half a step from the wall, it determines the required vector to stop it and strafes over accordingly [11]. Similar works available in [12, 13].

Joshna, V., Kashyap, M., Ananya, V., and Manitha, P proposed a paper to measure the harmful gas concentration in the present environment after the flood disaster and degasify the harmful gas from the environment. A Mq sensor which is used to measure the gas concentration then gives analog output via arduino UNO then a 12 V immersion pump is used to spray the oxidizing agent to reduce the concentration of hazardous gas. There are different types of MQ sensors which can measure various harmful gases, Now for the autonomous part IR (infrared) sensors are used to avoid obstacles [14].

Subramanian, M. A., Selvam, N., S, R., Mahalakshmi, R., and Ramprabhakar presented about the gas alerting system which involves IOT. A MQ sensor for detecting the gas concentration in the environment is connected with arduino UNO which then transmits the gas concentration to IOT cloud platform (Thinkspeak) via an wifi module, now the user can see if there is any abnormal gas activity in his/her house via thing speak cloud platform [15]. Similar analysis can be found in [16–18].

Varma, N. P., Aivek, V., and Pandi presented simulated results by comparing the results produced by two types of fuzzy controllers in matlab using a Pioneer 3AT 3 wheeled robot. Fuzzy controllers are linear rule based systems which make decisions based on the present data or available data the comparison of two types type 1 and type 2 is presented in an elaborate manner with result why the later fuzzy controller is efficient than the previous one [19]. Similar works can be found in [20, 21].

R. Ramkumar, S. Adarsh and K. I. Ramachandran proposed an Obstacle avoidance system by a mobile robot which uses a 2d lidar as well as 2 ultrasonic sensors the fusion of these two sensors is tested using ANFIS (Adaptive Neuro Fuzzy Inference System) and FIS models to get an accurate distance of the obstacle. With this fusion precise distance of the objects in an unknown environment can also be known. FIS based control algorithm is designed and implemented to avoid the walls provided in the maze path [22].

3 Proposed Method

3.1 Initial Setup

Figure 2 represents the overall hardware component setup except the robot.

(1) Raspberry Pi: The Raspberry Pi 4 2gb variant is used with Ubuntu 18.04 LTS desktop version. This simple computer is also installed with the Robot operating system ROS package and other necessary ROS packages for ydlidar and arduino.

(2) LiDAR: Ydlidar X2 is used which has a maximum range of 8 m and minimum of 0.1 m with a scan angle of 360°.

Error checking has to be done in every lidar because there will always be some real time error. In our lidar the error checking is done by taking values from a sample run from the lidar where it was covered by a box from a known distance. So that range expected, and range measured can be used to find the error. So, the formula used in Eq. 1.

$$\text{Error} = (\text{Expected value} - \text{measured value})/\text{EXPECTED value} \qquad (1)$$

Fig. 2 Block diagram of assembly unit

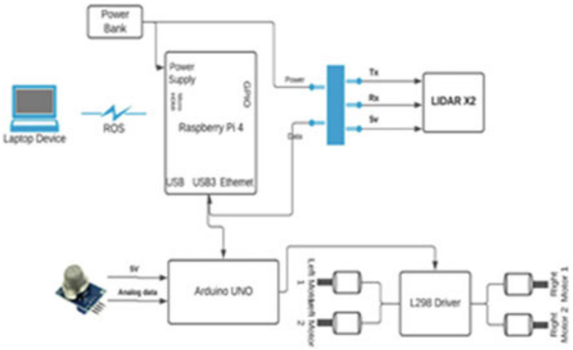


Table 1 Lidar data

Lidar data			
Measured angle	Measured range (m)	Actual range (m)	Error (cm)
359.2500006	0.5182499886	0.50	0.01824
270.484370	0.4942499995	0.50	0.00575
180.0312463	0.4772500098	0.50	0.02275
90.70312267	0.497249999	0.50	0.00276
0.046875001	0.5168499886	0.50	0.01684

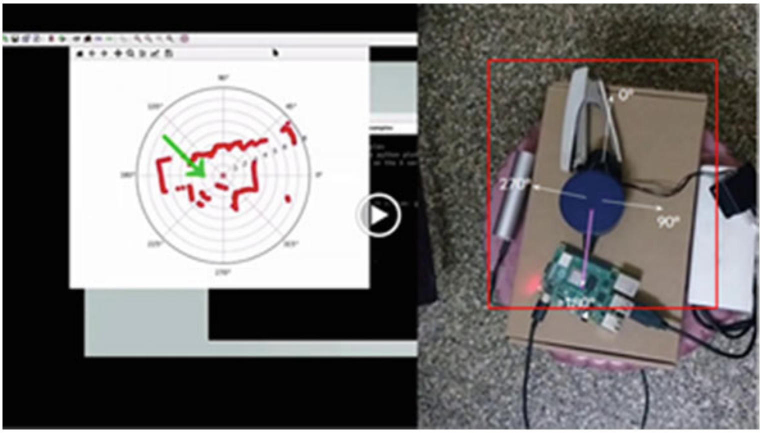


Fig. 3 Obtaining and mapping the LiDAR input

The Table 1 denotes the reading which is obtained from the lidar which is positioned as in Fig. 3.

% of Error = 1.52–4.6%

So the value obtained is considered an obstacle avoiding threshold value. So that error obtained can be neglected.

(3) Gas Sensor: MQ 2 gas sensor is used which is useful for gas leakage detecting. It can detect LPG, CH4, CO and alcohol connected with Arduino [14]. The range of MQ gas sensor concentration goes from 200 to 10000 ppm if the concentration value is les s than 200 ppm the smoke concentration is negligible.

(4) Arduino: Arduino Uno is used which is embedded with an ATmega328p microcontroller. The purpose of the arduino is to act as brain to raspberry pi to drive the robot to also detect the amount of gas produced in the surrounding environment with MQ5 sensor.

The robot has 2 compartments placed one after another. Under the first compartment 4 motors and l298 motor driver is attached above the first compartment Arduino Uno, Raspberry pi and a battery holder with 4 Li-ion batteries for powering the

Arduino and L298 driver. Above the second compartment the lidar is placed along with a powerbank with this the hardware setup is built.

3.2 Approach to Algorithm

Figure 2 block diagram depicts the obstacle avoidance setup. ROS melodic package, Ydlidar package and drivers along with arduino ROS support packages are installed on Raspberry pi 4 installed Ubuntu 18.04 LTS has its operating system. Arduino is used to control the h-bridge driver which deals with robot movement.

YD-Lidar X2 sensor is used to give data with a range of 360° and for every half angle −90–90° data used from the 360° scan data. In Fig. 4 flow diagram for the obstacle avoidance is shown. The point cloud data received was double checked for the calibrating purpose as the data shown in Fig. 3 and Table 1 [17].

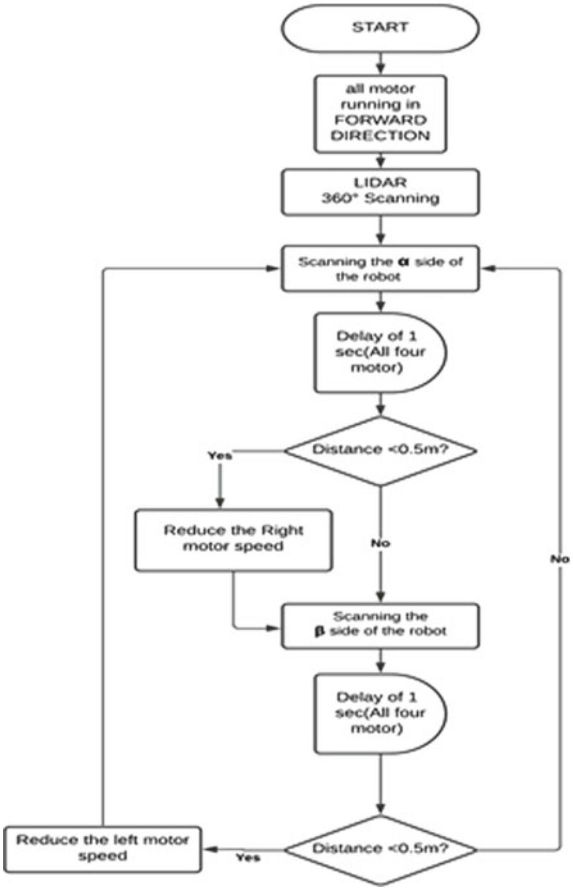

START
all motor running in FORWARD DIRECTION
LIDAR 360° Scanning
Scanning the α side of the robot
Delay of 1 sec(All four motor)
Distance <0.5m?
Yes
Reduce the Right motor speed
No
No
Scanning the β side of the robot
Delay of 1 sec(All four motor)
Reduce the left motor speed
Yes
Distance <0.5m?

Fig. 4 Flowchart for obstacle avoidance

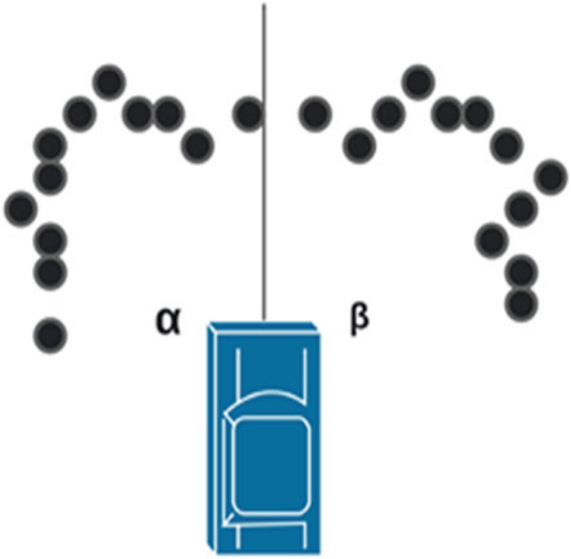


Fig. 5 Approach of Identification for the direction

If distance between the obstacle is less than 0.5 m on the α side which is −90–0° the robots try to avoid the obstacle by making the left side motors rotate in forward direction and right side motors doesn’t rotate to turn right, with a safe distance of 0.15 m between the robot and the obstacle and vice versa which is the threshold value for the robot. The above algorithm represents a fuzzy logic approach which could be useful for a mobile robot moving in an unknown environment [17] (Fig. 5).

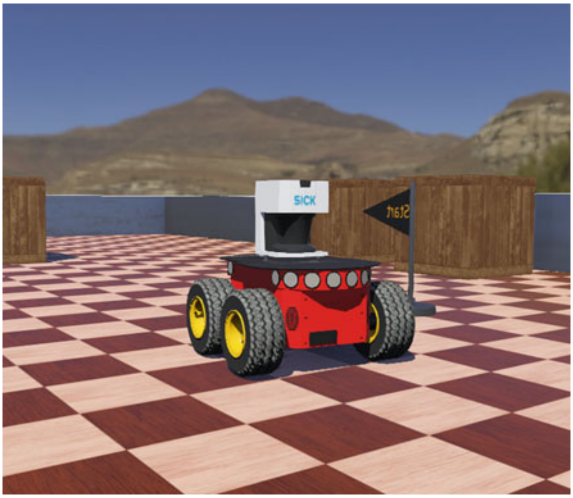
3.3 Simulation Verification

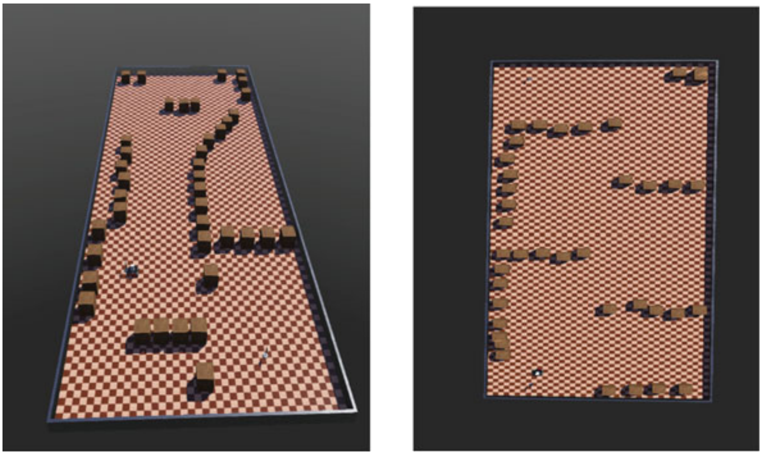
In order to verify the algorithm, the code built up is implemented in a simulation tool. Webots simulator is used, it is an open source virtual environment simulator. It is used to simulate and test our obstacle detecting algorithm. Then we made use of an inbuilt open source’s robot Pioneer 3AT (Fig. 6). Then a Conventional virtual lidar Sick LMS 291 Is used as ydlidar as x2 virtual is not available.

A basic rectangular box with desired area is used as our environment where some solid boxes are used as obstacles. By using different patterns, we tested our algorithm. A robot is placed in an initial position and allowed to run along the room without any destination to check whether our algorithm works in avoiding the obstacle placed before it. Our algorithm depicts the initial flowchart in Fig. 4.

Figure 7: type 1 and type 2 denotes different types of environment used in the simulation check-up, where the robot is checked to make sure it avoids an object without colliding with it. Braitenberg vehicle control strategy is implemented.

Fig. 6 Pioneer 3 AT robot





(a) Type 1

b) Type 2

Fig. 7 Different environment

4 Hardware

4.1 Design and Implementation

The Implementation process started by installing the suitable operating system for the Raspberry Pi 4 we chose Ubuntu 18.04 LTS, Then installing necessary drivers for lidar. Robot operating system ROS Melodic version is installed in Raspberry Pi

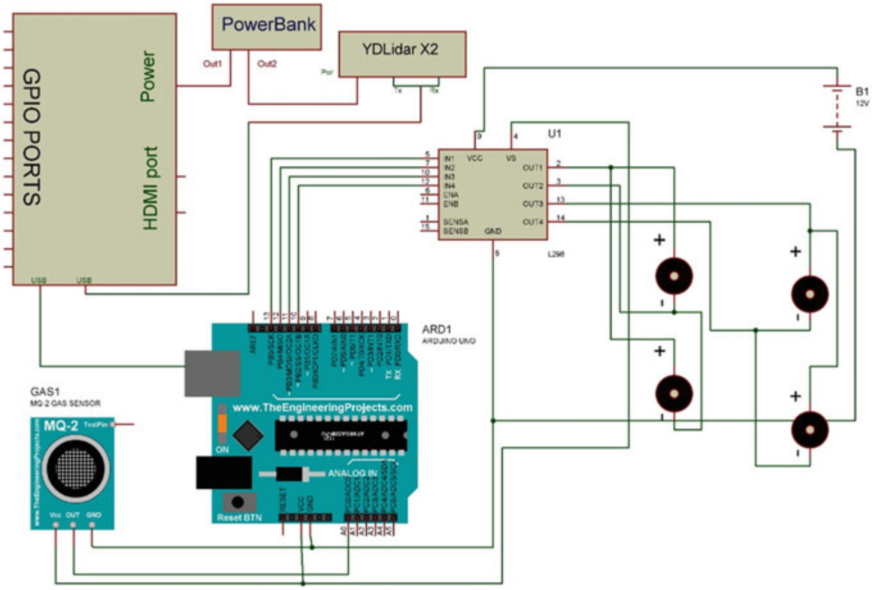


Fig. 8 Circuit diagram

then ROS support packages for both Arduino and Lidar are installed. A dual channel motor driver is used to operate the robot which is then connected with arduino. An arduino is developed which acts as a brain for L298 motor driver and gas sensor then sends data to the user via ROS. The Robot (Raspberry Pi) acts as the Master and the user (Laptop) act as the Observer. The data received from the gas sensor is shown in present time which can also be stored in a csv file and viewed later. The lidar which is directly connected so the data received from lidar is now separated based on Fig. 3 flow diagram then the robot decides whether to turn or proceed further this action is shown in real time using an ROS package called Rviz which also shows us the map being plotted then it can be exported for further use (Fig. 8).

5 Result

To inspect the proposed method the following verification is done separately. The first inspection is done for gas detection and then for obstacle avoidance.

Figures 9 and 10 shows the data received from a gas sensor and it is shown using the Arduino Ide serial monitor. Figure 10 represents the data received when there is no gas concentration, it usually detects and gives value 85 to 87 ppm. Then after giving an input they value hikes for the initial constant to a peak value as shown in Fig. 10. Similar analysis can be found in [9, 12].

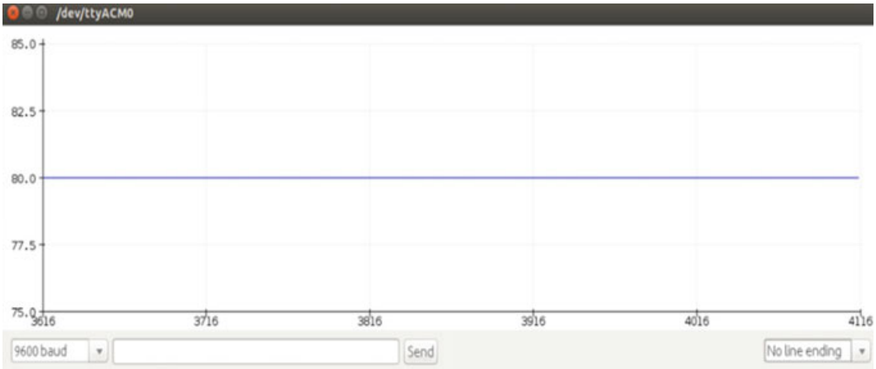


Fig. 9 Serial monitor graph when the gas detection is absent (*x*-axis = time, *y*-axis = ppm)

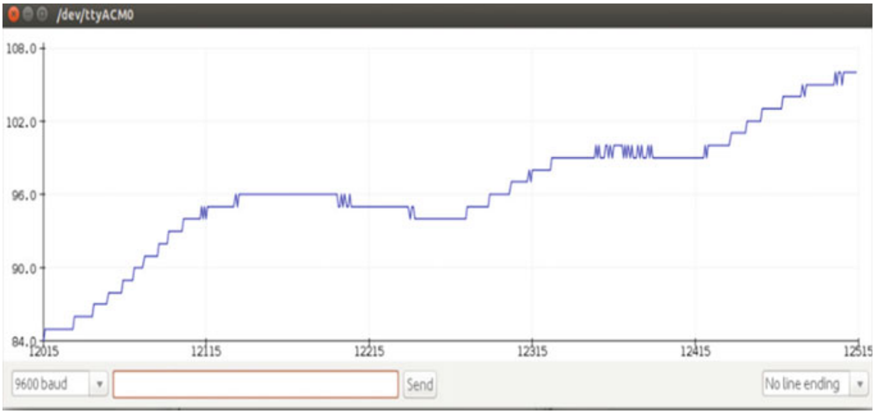
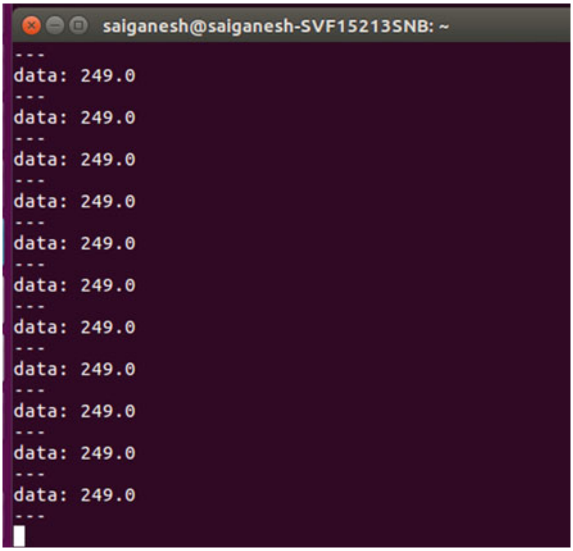


Fig. 10 Serial monitor graph when the gas detection is present (x-axis = time, y-axis = ppm)

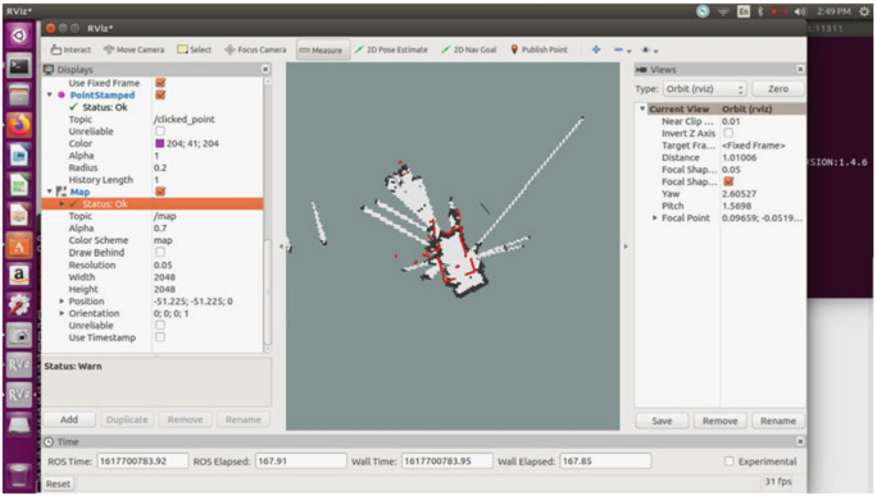
Figure 11 shows the value received from the gas sensor via ROS which is displayed in the Observers laptop/PC. There is no specific alerting system in the mobile robot the values received from the gas sensor which is shown via ROS in a terminal. If needed the concentration values can be graphed in a timely manner.

Figure 12 shows the real time mapping of the surrounding environment done by the robot via Rviz, an ROS package. A SLAM algorithm is used to map the surrounding environment by taking the lidar laser scan and plot it simultaneously based on the laser intensity, if the intensity is above a certain range it maps via black point and then forms a map with blackpoint clouds as shown in Fig. 12.

Fig. 11 ROS output of gas sensor on observer's interface



```
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
data: 249.0
---
```



Displays	
Use Fixed Frame	☑
▾ ● PointStamped	☑
✓ Status: Ok	
Topic	/clicked_point
Unreliable	□
Color	204; 41; 204
Alpha	1
Radius	0.2
History Length	1
▾ Map	☑
▸ ✓ Status: Ok	
Topic	/map
Alpha	0.7
Color Scheme	map
Draw Behind	□
Resolution	0.05
Width	2048
Height	2048
▸ Position	-51.225; -51.225; 0
▸ Orientation	0; 0; 0; 1
Unreliable	□
Use Timestamp	□

Status: Warn

Type: Orbit (rviz)

▾ Current View	Orbit (rviz)
Near Clip ...	0.01
Invert Z Axis	□
Target Fra...	<Fixed Frame>
Distance	1.01006
Focal Shap...	0.05
Focal Shap...	☑
Yaw	2.60527
Pitch	1.5698
▸ Focal Point	0.09659; -0.0519...

ROS Time: 1617700783.92 ROS Elapsed: 167.91 Wall Time: 1617700783.95 Wall Elapsed: 167.85 □ Experimental

31 fps

```
:11311
SION:1.4.6
```

Fig. 12 Rviz realtime map output

6 Conclusion and Future Scope

In this paper, we have done both simulation and hardware implementation (Fig. 1). The result shows that the proposed method works better in simulation and also in hardware implementation. The real time error of LiDAR was also identified and verified. Finally we have a robot that avoids the obstacles autonomously and mapping the surrounding environment within a range and also the robot detects the gas concentration level in the surrounding by ppm unit. Our robot still can be.

made robust so that it can be utilized in hazardous areas such as chemical industries and powerplant. Compared to conventional obstacle avoidance robot lidar based obstacle avoidance robots are more efficient and robust, the mobility of the robot is improved with the inclusion of a bluetooth module which helps the robot to move effortlessly by the user in a known environment. But the disadvantage of this robot is it uses a fuzzy logic method that can move in any number of ways if it is in the same environment. By ML (Machine Learning) the robustness, precision of movement could be increased significantly. In the upcoming days, the LiDAR can be used for several purposes instead of a traditional sensor. Now it's mostly used in mapping and in NASA as an instrumental in the satellite. As time progresses, unmanned vehicles will be using LiDAR.

References

1. D. Hutabarat, M. Rivai, D. Purwanto, H. Hutomo, *Lidar-based Obstacle Avoidance for the Autonomous Mobile Robot. 2019 12th International Conference on Information & Communication Technology and System (ICTS)* (2019). https://doi.org/10.1109/icts.2019.8850952
2. T.R. Madhavan, M. Adharsh, Obstacle detection and obstacle avoidance algorithm based on 2-D RPLiDAR, in *2019 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, Tamil Nadu, India, pp. 1–4 (2019). https://doi.org/10.1109/ICCCI.2019.8821803.
3. H. Dong, C-Y. Weng, C. Guo, H. Yu, I.-M. Chen, Real-time avoidance strategy of dynamic obstacles via half model-free detection and tracking with 2D lidar for mobile robots. IEEE/ASME Trans. Mech. 1–1 (2020). https://doi.org/10.1109/TMECH.2020.3034982
4. Y. Peng, D. Qu, Y. Zhong, S. Xie, J. Luo, J. Gu, The obstacle detection and obstacle avoidance algorithm based on 2-D lidar, in *2015 IEEE International Conference on Information and Automation*, Lijiang, pp. 1648–1653 (2015). https://doi.org/10.1109/ICInfA.2015.7279550
5. Z. Rozsa, T. Sziranyi, Object detection from a few LIDAR scanning planes. IEEE Trans. Intell. Veh. **4**(4), 548–560 (2019). https://doi.org/10.1109/TIV.2019.2938109
6. N. Baras, G. Nantzios, D. Ziouzios, M. Dasygenis, Autonomous obstacle avoidance vehicle using LIDAR and an embedded system, in *2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, Thessaloniki, Greece, pp. 1–4 (2019). https://doi.org/10.1109/MOCAST.2019.8742065
7. D. Hutabarat, M. Rivai, D. Purwanto, H. Hutomo, Lidar-based obstacle avoidance for the autonomous mobile robot, in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, Indonesia, pp. 197–202 (2019). https://doi.org/10.1109/ICTS.2019.8850952
8. A. Fenesan, T. Pana, D. Szöcs, W. Chen, Building an electric model vehicle and implementing an obstacle avoidance algorithm, in *2012 International Conference and Exposition on Electrical and Power Engineering*, Iasi, Romania, pp. 49–53 (2012).https://doi.org/10.1109/ICEPE.2012.6463599
9. R.K. Kodali, R.N.V. Greeshma, K.P. Nimmanapalli, Y.K.Y. Borra, IOT based industrial plant safety gas leakage detection system, in *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (2018). https://doi.org/10.1109/ccaa.2018.8777463

10. K. Song et al., Navigation Control Design of a Mobile Robot by Integrating Obstacle Avoidance and LiDAR SLAM, in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan, pp. 1833–1838 (2018). https://doi.org/10.1109/SMC.2018.00317
11. S.T. Padgett, A.F. Browne, Vector-based robot obstacle avoidance using LIDAR and mecanum drive, in *SoutheastCon 2017*, Concord, NC, USA, pp. 1–5 (2017).https://doi.org/10.1109/SECON.2017.7925312
12. S. Jamadagni, P. Sankpal, S. Patil, N. Chougule, S. Gurav, Gas Leakage and fire detection using raspberry Pi. in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (2019). https://doi.org/10.1109/iccmc.2019.8819678
13. T.R. Madhavan, M. Adharsh, Obstacle detection and obstacle avoidance algorithm based on 2-D RPLiDAR, in *2019 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2019, pp. 1–4. https://doi.org/10.1109/ICCCI.2019.8821803
14. V. Joshna, M. Kashyap, V. Ananya, P. Manitha, Fully autonomous robot to detect and degasify hazardous gas after flood disaster, in *2019 2nd International Conference on Power and Embedded Drive Control (ICPEDC)* (2019). https://doi.org/10.1109/icpedc47771.2019.9036703
15. M.A. Subramanian, N.S.R. Selvam, R. Mahalakshmi, J. Ramprabhakar, Gas leakage detection system using IoT with integrated notifications using Pushbullet–a review, in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)* (2020). https://doi.org/10.1109/icisc47916.2020.9171093
16. S. Geetapriya, N.R. Pillai, C.K. Aswin, M. Menon, Graph-Based algorithm for mobile robot navigation in a known environment, in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (2019). https://doi.org/10.1109/icoei.2019.8862775
17. N.P. Mayura, S. Veni, Building detection from LIDAR point cloud data, in *2019 International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, pp. 1416–1421 (2019). https://doi.org/10.1109/ICCES45898.2019.9002555
18. S. Shrestha, S. Shakya, Technical analysis of Zigbee wireless communication. J. Trends Comput. Sci. Smart Technol (TCSST) **2**(04), 197–203 (2020)
19. N.P. Varma, V. Aivek, V.R. Pandi, Intelligent wall following control of differential drive mobile robot along with target tracking and obstacle avoidance, in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)* (2017). https://doi.org/10.1109/icicict1.2017.8342539
20. A.D. Adhvaryu, S. Adarsh, K.I. Ramchandran, Design of fuzzy based intelligent controller for autonomous mobile robot navigation, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2017). https://doi.org/10.1109/icacci.2017.8125946
21. G. Ranganathan, An economical robotic Armplaying chess using visual servoing. J. Innov. Image Process(JIIP) **2**(03), 141–146 (2020)
22. R. Ramkumar, S. Adarsh, K.I. Ramachandran, Fusion of ultrasonic and RP Lidar 360 sensors using ANFIS for mobile robot navigation, in *2018 15th IEEE India Council International Conference (INDICON)*, Coimbatore, India, pp. 1–6, (2018)https://doi.org/10.1109/INDICON45594.2018.8987187

Implementation of the Modified Pre-trained DenseNet Model for the Classification of Grades of the Diabetic Retinopathy

Nitin Shivsharan and Sanjay Ganorkar

Abstract One of the world's most common causes of vision loss is diabetic retinopathy (DR); specific DR estimates are critical for planning and reviewing DR prevention and treatment strategies. The primary goal of this research is to construct an updated pre-trained model like dense convolutional network (DenseNet) and implement it for the classification of the DR severity stages like no DR, mild, moderate, severe, and proliferative. The implemented model uses the fundus photos graphs as input to distinguish different levels of the DR stages. From the freely accessible online dataset like "Asia Pacific Tele-Ophthalmology Society (APTOS) Blindness Detection", we have used fundus photographs to train the DensNet model. The essence of the dataset is an imbalance, and due to the imbalanced existence of the dataset, the training bias problem arises at the moment of training the deep learning models. The suggested system uses the pre-trained DenseNet-BC model which has been updated and trained to address the training bias problem with the weight method. The proposed model gives results in the form of performance parameters such as accuracy equal to 99.86%, recall equal to 98.02% and F1-score equal to 98.93%.

Keywords Diabetic retinopathy · Deep learning · Densenet model · Multi-class · Classification · Confusion matrix

1 Introduction

Diabetes is a form of disease in which a person does not produce enough insulin or does not respond to the release of insulin. Insulin is a hormone that controls how much

N. Shivsharan (✉)
Computer Engineering, SSPM's College of Engineering Kankavali, Sindhudurga 416602, India
e-mail: shivsharan.nitin@gmail.com

S. Ganorkar
Department Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Pune 411041, India
e-mail: srganorkar.scoe@sinhgad.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_30

glucose is in the blood. Diabetes refers to high blood sugar levels, which may result in rupturing of tiny blood vessels in the retina. And as fluid leaks from the ruptured blood vessels and spreads over the eye's surface, vision blurs, resulting in DR.

1.1 Diabetes and Diabetic Retinopathy Prevalence

In the near future, the predicted incidence of diabetes is seen in Table 1, which displays count of people with diabetes in the future. DR is now becoming one of the big vision-related complications caused by prolonged diabetes. And it has been noted that 19% world's population is suffering from the DR. Therefore, it is important to implement the periodic DR diagnosis process [2]. An automated DR screening method has been known for a long time, and previous attempts have yielded positive results through machine learning using images as a input data.

One of the goal of this work is to follow a pre-trained DenseNet model and change it in such a way as to provide better classification results. It is also suggested that the built model be used as a pre-screening method during the DR diagnosis phase. However, qualified and specialized individuals are required for the DR pre-screening process. Around 2015 and 2019, a study was conducted to determine the global number of ophthalmologists. According to the report, there were just 25 thousand ophthalmologists in 194 countries [3]. Due to this shortfall of the availability of experts required to at the time of mass DR screening, the computer-aided diagnostic (CAD) techniques are now gaining popularity as the most commonly used method. Three major issues arise during the automated grading of DR using CAD at the time of screening.

- The first is that most of the CAD programmes presently available assist to diagnosis the DR in just two grades, i.e. abnormal and the normal. But, the actual progression of DR happens through different five stages [4].
  The pictorial view of five different progression stages of the DR is shown in Fig. 1. The leftmost part of Fig. 1 depicts the first stage of DR development, i.e. no DR, while the rightmost part depicts the final stage of DR, i.e. PDR. The key purpose of the this work is to categorize fundus photos into the DR's various grades with improved accuracy.
- The optimum overall classification accuracy is the second issue with the multi-class classification problem.

Table 1 Projection of prevalence of diabetes in the near future [1]

Year	Projected prevalence in %	Actual count in millions
2019	9.3	463
2030	10.2	578
2045	10.9	700

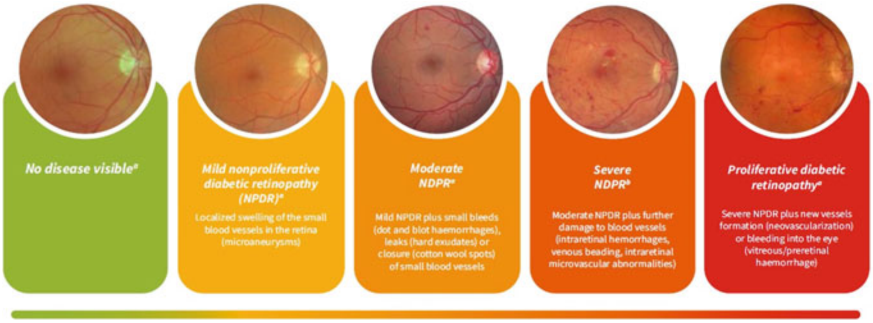

No disease visible[a]
Mild nonproliferative diabetic retinopathy (NPDR)[a]
Localized swelling of the small blood vessels in the retina (microaneurysms)
Moderate NDPR[a]
Mild NPDR plus small bleeds (dot and blot haemorrhages), leaks (hard exudates) or closure (cotton wool spots) of small blood vessels
Severe NDPR[b]
Moderate NPDR plus further damage to blood vessels (intraretinal hemorrhages, venous beading, intraretinal microvascular abnormalities)
Proliferative diabetic retinopathy[a]
Severe NPDR plus new vessels formation (neovascularization) or bleeding into the eye (vitreous/preretinal haemorrhage)

Fig. 1 Pictorial view of progression of the DR stages

- Problem third is related to the imbalanced property of dataset. The imbalanced dataset is one in which the number of samples in each class is not equal.

Also, on the basis of literature survey and literature review, it has been observed that there are relatively few papers concerned with classification of five stages of the DR using the convolutional neural network (CNN) method. In case of imbalanced dataset, it has to be ensured that a new developed model is still able to learn from the fundus images.

This article presents modified pre-trained DenseNet model for categorization various grades of DR. And the performance of the model presented is measured using performance criteria such as precision f1-score recall and average accuracy. The proposed system adopts three steps to implement the model like the fundus image pre-processing, the construction of DenseNet model and the performance evaluation.

The paper is structured as follows. Followed by the introduction section, the Section two explains the relevant work performed in this field by researchers. This section provides a summary of different methods used for detection and classification of DR. Section three describes various methods adopted and implemented to develop the proposed model. And results obtained have been elaborated in the Section four. The conclusion is the fifth and final section.

2 Related Work

The numerous methods for detecting and classifying the DR phases are discussed in this section.

Bhatia et al. [5] focus on detecting disease presence in the fundus image using an algorithm based on ensemble machine learning. The algorithm is applied to features derived from the results of various retinal image processing algorithms, such as optical disc distance, lesion specific, image level and quality assessment. In 2020, Gaurav Saxena et al. [6] addressed the various issues related to CNN technique used to detect diseases automatically using medical images. This article explores the use of publicly available databases, metrics used to deal with unbalanced datasets, related findings and their comparison with leading studies. In 2020, Ashish Bora et al. [7] developed and tested two variants of a deep learning framework. As the warning signs are monitored, the extent of seriousness of the condition must be validated in order to better take decisions about suitable care. The paper focuses on the concept of using a deep learning model to classify DR fundus images by severity level.

In the 2020 K. Shanka et al. [8] suggested a model like synergic deep learning (SDL) used to classify fundus images at various severity levels of DR. The results of the experiment show that the SDL model given provides better classification over current models. [9] focuses on the classification of DR fundus images using CNN technique according to the severity of the DR disease by using pooling, rectified linear activation unit (ReLU) and softmax, layers to achieve a high degree of accuracy. The output of the proposed algorithm was validated using the Messidor database. Deep learning is now a fast-track methodology for medical image processing. But deep learning algorithms are black boxes, like the popular ConvNets. A solution to produce heatmaps showing the pixels in images play a role in the fundus image grade predictions is suggested in the article [10]. Four forms of lesions were segmented manually in the DiaretDB1 dataset, and the output was also evaluated at the image level and at the lesion level.

The article [11] offers an introduction to recent deep learning models for the classification of different developmental stages of the DR. An approach to predict severity stages of the DR using ensemble-based deep learning model is suggested in the year 2020 by Reddy et al. [12]. The comparative analysis reveals that the methodology of ensemble machine learning outperforms the algorithms of individual machine learning. Comprehensive and automatic DR detection tools and methods were required. Previous approaches have shown promising results using image detection, pattern recognition and machine leaning. The study presented in [13] regarding the photography of the colour fundus images and the results suggest that the model has functional therapeutic potential. The extracted features of the DR dataset were identified using a principal component analysis-based deep neural network model with the Grey Wolf Optimization (GWO) algorithm [14]. And use of GWO allows for the best parameter selection for the model's training.

The research paper [15] has proposed an automatic system of DR identification and grading called DeepDR. DeepDR specifically senses the presence and strength of DR from fundus images by transfer and ensemble learning methods. In 2020 S Gayatri et al. [16], the proposed work compares performances of various traditional classifiers such as support vector machine (SVM), random forest (RF), random tree and J48 classifiers. And it has been found that the RF classifier is doing better compared to the remaining models with an improved overall accuracy for binary and multi-class classifications. A deep learning fundus imaging analysis for Macular Edema and the DR intensity stages was suggested by Jaakko Sahlsten [17]. In this study, they used a deep learning method to identify referable DRs.

An automated recognition of mild DR and multi-class DR levels using deep learning concepts was implemented in 2020 by Rubina Sarki et al.[18]. In 2019, Karthikeyan et al. [19] suggested a multi-class DR level classification model using artificial intelligence. One of the best aspects of this report is that it used limited samples to give training to the proposed CNN model. Also, in 2020, Li et al. [20] implemented a deep CNN model in the CAD system for detection of the DR. In 2012 [21], a popular approach such as weight method is adopted for dealing with imbalanced type of data. At the time of training the model, weighting samples in uncommon classes with high costs and then applying cost-sensitive learning principles were used. The proportion of samples of each class in the training set is generally used to calculate the weight of a class. In [22], three key components like microaneurysm detection, blended features candidate classification and DR prediction using merged image and lesion level features are implemented in an automated DR screening method based on colour fundus images. The [23] describes a quadrant ensemble automated DR grading method based on the Inception Resnet-V2 deep neural network architecture. For better network output, the presented model includes histogram equalization, optical disc localization and quadrant cropping, as well as a data augmentation stage. DR identification and classification using a reformed capsule network are discussed in [22]. The features from the fundus images are extracted using primary capsule layers and the convolution. The class capsule layer and softmax layer are used to determine the probability that the image belongs to a specific class.

2.1 Literature Review

Several methods for automated recognition and segmentation of retinal landmarks and pathologies have been developed in the past. However, recent breakthroughs in deep learning and modern imaging modalities in ophthalmology have opened up a whole new world of possibilities for researchers. The article [24] presents a state-of-the-art review of deep learning techniques for automatically classifying retinal landmarks, anatomy and disease using 2D fundus and 3D retinal images like optical coherence tomography (OCT) and 2D fundus images. This subsection of the paper provides an overall review study of the methods used to diagnose DR processes, the different datasets used and the methods used to detect lesions. Figure 2a. indicates the proportion of studies that use one or more publicly accessible datasets. And it has been found from the statistic that approximately 59% of studies use more than one dataset for the purpose of experimentation.

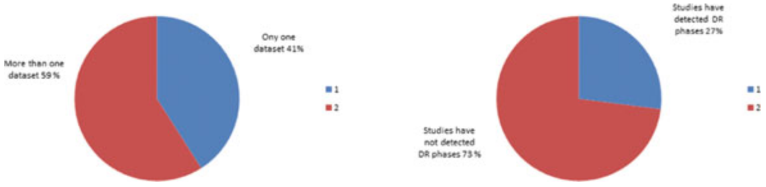
From Fig, 2b, it is identified that just 27% of studies recognized all five stages of the DR and 73% of studies only identify DR stages of the fundus images such as fundus image with DR and the no DR. The proportion of studies that uses methods to identify various lesions like Microaneurysm, Exudates, Hemhorages and Cottonwool Spots (CWS) in DR have been presented in Fig. 3.

Deep learning, particularly in the analysis and classification of medical images, has recently become one of the most successful techniques that has improved performance in many fields [25].

3 Proposed Work

3.1 Dataset

We used publicly available APTOS dataset [26] for training the model and evaluating its performance. The complete dataset is comprised of 18590 fundus photos, which organizers of the Kaggle competition separated into 3662. 1928 and 13000 training, validation and testing images, respectively. The training dataset consists of 3662 fundus images labelled with DR phases for individuals' left and right eye. The dataset is unbalanced by default, meaning that the class distribution between classes is not uniform. The distribution of images per DR class is shown in Fig. 4. And therefore, it is one of the issue in the multi- class classification system [27]. We used perplexity, which is probabilistic modelling, to present the imbalanced dataset multi-class classification problem. The distribution of the dataset using the theory of perplexity is demonstrated in Fig. 5.



(a) (b)

Fig. 2 [25] **a** The proportion of experiments using one or more public datasets. **b** The proportion of studies that have detected DR phases

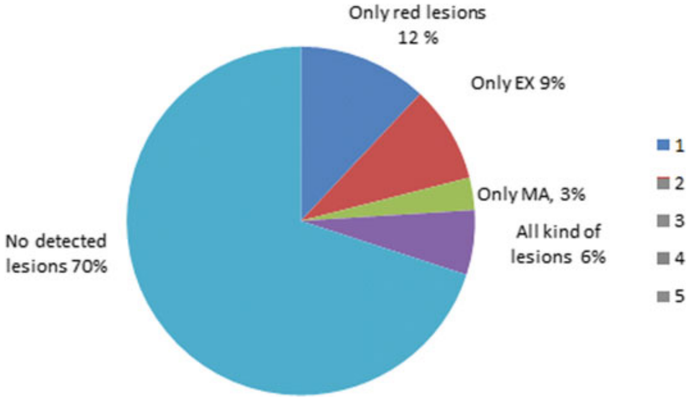


Fig. 3 The review of publications that found DR lesions [25]

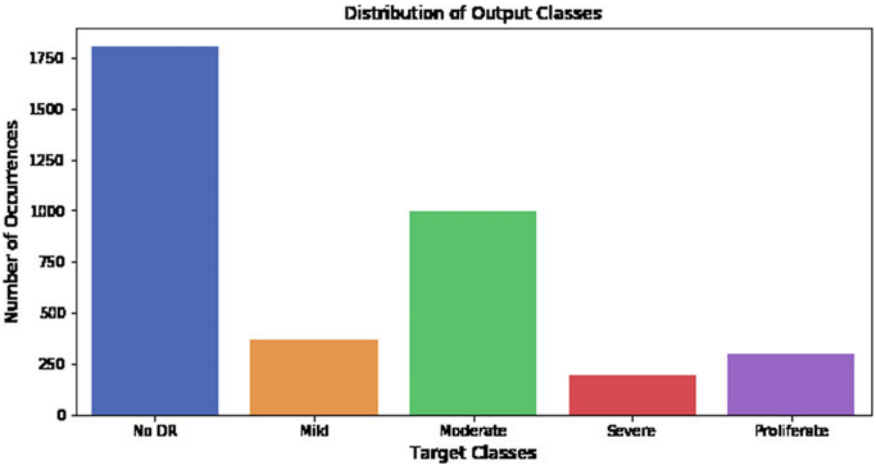


Fig. 4 No of images per DR class

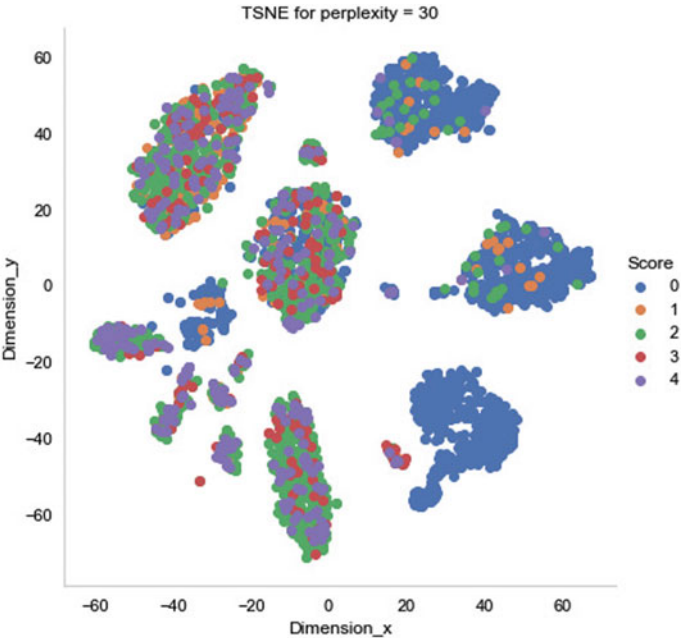


Fig. 5 Perplexity distribution for the APTOS blindness detection dataset

3.2 Methods

This section outlines the various methods proposed and implemented to apply the pre-trained DenseNet model to sort out fundus images into distinct levels of the DR. In order to enhance the accuracy of the current method, this paper suggested the modified pre-trained DenseNet model for the estimation of different phases of the DR. A large collection of retina images was developed using fundus photographing in a variety of lighting situations. We will find noisy data in the same way as we will in any dataset obtained. As a result, it is important to implement and use preprocessing techniques on image dataset. The following subsection presents preprocessing techniques used to develop the proposed model.

Preprocessing The Gaussian filter is used to smooth the input fundus images. In the subsequent step of preprocessing the fundus images, following the Gaussian filter technique, Ben Graham's approach is applied. The Ben Graham approach [28] is implemented on fundus images. Graham performed both adding and scaling operation on the image to a circular crop. A black segment around the actual representation of the retina image has an effect on the model's efficiency because it lacks detail features that are useful for learning. So, it is essential to remove black part of the image.

The size of the fundus images in the dataset varies as well. As a result, we used the radius of a 500-pixel image to make images of the same size. After applying all image preprocessing techniques like Gaussian blurring and Graham's proposed methods, the preprocessed image is shown in Fig. 6b.

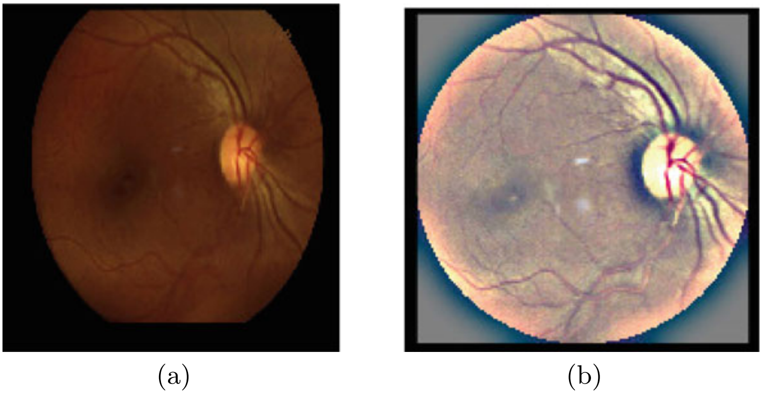

(a) (b)

Fig. 6 Preprocessing: **a** Input image **b** Image after applying all preprocessing steps

3.3 Building DenseNet Deep Learning Model

DenseNet is a network architecture that uses shorter links between layers to deepen deep learning networks while also allowing them more efficient to practice. It is a CNN with each layer connecting to the layers below it.

DenseNet DenseNets, contrary to popular belief, require fewer parameters than traditional CNNs because they do not be trained repetitive feature maps. Furthermore, some ResNet variants have shown that certain layers play a minor role and can be removed. Due to the information flow and gradients, one of the issues with very deep networks is the difficulty in training them. Since each layer has direct access to the gradients from the loss function and the original input image, DenseNets solves this problem. DenseNet is one of the latest visual object recognition discoveries in neural networks. With some basic variations, DenseNet is very similar to ResNet. A typical L-layer CNN has L-connections, but in case of DenseNet has $L(L+1)/2$ between each layer and its corresponding layer. ResNet uses an approach that is additive (+) that merges the identity layer with the future layer, while DenseNet connects the former and subsequent layers [29]. In Standard ConvNet, the input image is convolutioned several times to acquire high-level functionality. In standard ConvNet, the input image is convolutioned several times to acquire high-level functionality. Figure 7a depicts the sequence of convolution operations that occur in regular ConvNet. The original assumption that accuracy improves as the network grows deeper is incorrect: as the network grows deeper, consistency and accuracy problems arise. One of the most serious issues we have seen is the vanishing gradient problem. DenseNets and ResNets architectures attempt to solve this challenge, and they are, in our view, efficiency breakthroughs for deep CNN architectures. Figure 7b shows how the ResNet principle works. Each layer receives "collective intelligence" from the layers above it. In the DenseNet model, channel-wise concatenation is seen in Fig. 7c. The overall representation of the DenseNet model is shown Fig. 7d. Since each layer receives attribute maps from all preceding layers, the network can be thinner and more lightweight, resulting in fewer channels. The mathematical modelling of DenseNet model is as follows

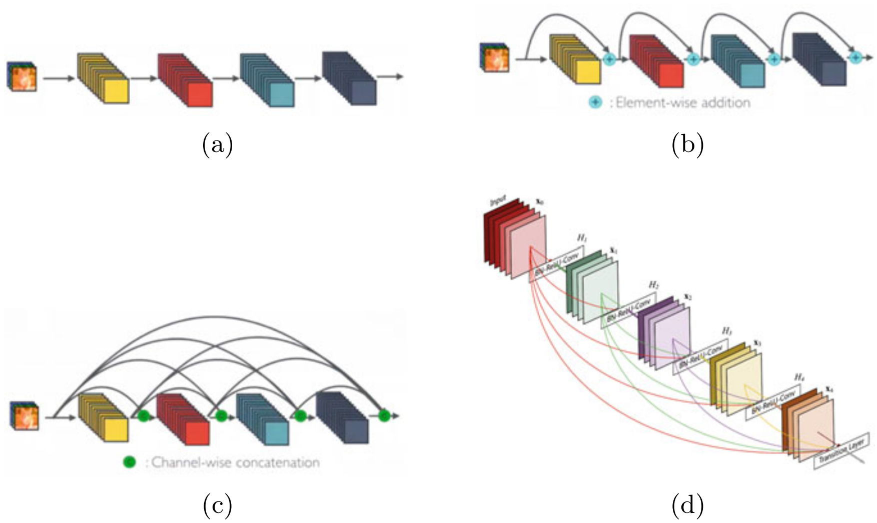
$$X_l = H_l(X_{l-1}) \tag{1}$$

Equation 1 is derived from the ResNet model. This behaviour was expanded by ResNets to include the skip relation, reformulating the Eq. 1 as:

$$X_l = H_l(X_{l-1}) + X_{l-1} \tag{2}$$

DenseNets concatenate the layer's output feature maps with the incoming feature maps rather than merging them. As a result, the Eq. 2 reshapes into:

$$X_l = H_l([x_0, x_1, x_{l-1}]) \tag{3}$$



(a) (b)

(c) (d)

Fig. 7 DenseNet model representation **a** Convolution layers in standard ConvNet **b** Element-wise addition in the ResNet architecture **c** Channel-wise concatenation in the DenseNet architecture **d** Pictorial view of DenseNet model [29]

Concatenating the feature maps have been represented mathematically as shown in Eq. 3.

DenseNets-BC The B appears after the term Bottleneck layer, which you may recognize from your ResNets work. To maximize model compactness much further, we may decrease the amount of feature maps at transformation layers. For those cases wherever the quantity of performance function maps would like to be limited, DenseNet-C is a very small incremental step to DenseNets-B. This reduction is calculated by the compression ratio θ. Instead of using m feature maps at a specific layer, we will have $\theta * m$. The value of θ ranges in between $[0 - -1]$. The code runs with the DenseNet-BC architecture, which has 1 × 1 convolutionary bottleneck layers, by design, and compresses the number of channels by 0.5 at each transfer layer. The pre-trained DenseNet model is adopted and trained using the class weight method.

Class Weight Method We present the updated pre-trained DenseNet model in this article, which links each layer in a feed-forward fashion to every other layer. To address the training bias problem that arises due to imbalanced datset, the class weight approach is implemented.

$$w_j = n_{\text{samples}}/(n_{\text{classes}} * n_{\text{samples}} j) \tag{4}$$

Here

- w_j: It is the weight of each class
- n_{samples}: The dataset's number of samples overall.

- n_{classes}: number of classes overall that are unique.
- $n_{\text{samples}} j$: number of classes in overall in the relevant class.

3.4 Pseudo Code

The following pseudo code depicts the overall execution of the proposed model.

Algorithm 1 Implementation of DenseNet model to grade the severity of DR stages

```
for Each fundus image do
    Apply the preprocessing steps
    Use Ben Graham approach
    Apply resize and cropping
end for
Use all preprocessed images to train the DenseNet model
Build the DenseNet model
Modify the training parameters of the DenseNet model
Use Class Weight method to train the model
Evaluate the model's performance
```

4 Result Analysis

The results of using the DenseNet model to grade the intensity of DR stages are discussed in this section. This section also includes explanation about the performance parameters to be used to assess results of the developed model and a distinguishing of the proposed model's results to that of the current model.

4.1 Metrics for Performance Evaluation

The terminologies used to describe performance metrics are as follows

- The case positive (P) is true positive instances that the data comprises.
- The case negative (N) is the true negarive instances that the data comprises.
- Performance parameters like true positive (TP), true negative (TN), false positive (FP) and the false negative (FN)

Using the above terminologies, the mathematical expression for performance metrics are as follows

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

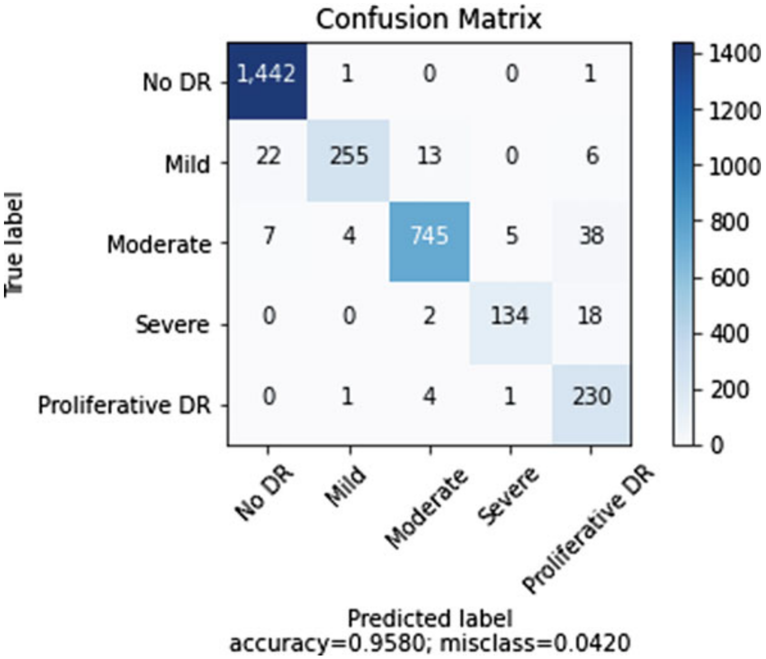
$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

The above equations express the mathematical description of the output parameters used to calculate the success of modified pre-trained DenseNet models

4.2 Results and Discussion

The findings obtained after implementing the constructed model for the classification of fundus images are described in this section.

At the training point, the model's success is measured using a confusion matrix as a performance metric. A confusion matrix is a table that can be used to simulate the prediction model's performance. Each entry in the confusion matrix represents the



	No DR	Mild	Moderate	Severe	Proliferative DR
No DR	1,442	1	0	0	1
Mild	22	255	13	0	6
Moderate	7	4	745	5	38
Severe	0	0	2	134	18
Proliferative DR	0	1	4	1	230

Fig. 8 Training instance confusion matrix

Table 2 The DenseNet model's efficiency

	Precision	Recall	F1-score	Support
No DR	0.9986	0.9802	0.9893	1471
Mild	0.8614	0.9770	0.9156	261
Moderate	0.9324	0.9753	0.9532	764
Severe	0.8701	0.9571	0.9115	140
PDR	0.9745	0.7849	0.8695	293
Accuracy	0.9580	0.9580	0.9580	0.9580
Macro avg	0.9274	0.9349	0.9278	2929
Weighted avg	0.9605	0.9580	0.9576	2929

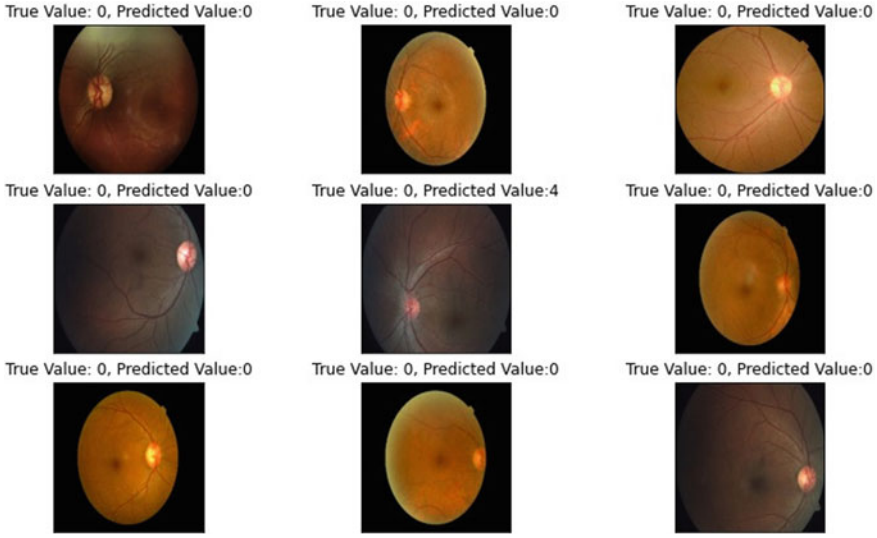

True Value: 0, Predicted Value:0
True Value: 0, Predicted Value:0
True Value: 0, Predicted Value:0
True Value: 0, Predicted Value:0
True Value: 0, Predicted Value:4
True Value: 0, Predicted Value:0
True Value: 0, Predicted Value:0
True Value: 0, Predicted Value:0
True Value: 0, Predicted Value:0

Fig. 9 Prediction of DR stages.

number of observations made by the model that correctly or incorrectly identified the groups. Figure 8 is a confusion matrix that depicts the performance of the built DenseNet model in predicting DR grades. Table 2 displays the DenseNet model's overall classification report. We have used output parameters like accuracy, recall and f1-score. From Table 2, it has been observed that the implemented model is showing accuracy equal to 99.86%, recall equal to 98.02% and f1-score equal to 98.93% for classifying fundus images into the no DR stage. The proposed model provides an overall multi-class classification accuracy equal to 95.80%, while classifying various phases of the DR.

Finally, the modified pre-trained DenseNet model implemented on the actual fundus image dataset to forecast the phases of intensity of the DR. The final output

Table 3 Comparison of methods used to diagnose the grades of the DR

References	Methods	Performance metrics
Wei Zhang et al.[15]	CNN (ResNet50, InceptionV3, Inception ResNetV2, Xception and DenseNets)	Accuracy= 96.5 % Sensitivity= 98.1% and a Specificity= 98.9%
Tao Li et al. [30]	DenseNet-121, CNN (GoogLeNet, VGG-16, ResNet-18, and SE-BN-Inception)	area under curve (AUC)=82.84%
Harry Pratt et al. [31]	CNN	Accuracy=75%, Sensitivity=30 %, Specificity=95%
Yi Peng et al. [32]	CNN (WP-CNN, ResNet, SeNet and DenseNet)	Sensitivity=90.94 %, Specificity=95.7%
Proposed model	DenseNet-BC	Accuracy=99.86%, Recall= 98.02% and f1-score = 98.93%

of the developed model, which takes a fundus picture as input and predicts its DR stage, is shown in Fig. 9.

Table 3 compares several ways for diagnosing the grades of the DR using various kinds of performance metrics. The proposed model outperforms existing models, according to the comparison analysis.

5 Conclusion

Currently, the CAD-based mass DR screening approaches are most commonly used. This paper proposes a modified pre-trained DenseNet model that can be used to describe the five stages of DR, including no DR, mild, moderate, severe and PDR. The DenseNet model used in this research to diagnose DR severity stages takes a fundus picture as input and applies an established model to it in order to forecast the seriousness of the various phases of DR.

On the fundus picture, Ben Graham's preprocessing method is used. And, when it comes to model training, the class weight approach is used to reduce training bias caused by the dataset's imbalance. DenseNet has one significant benefit over traditional deep CNNs: information passed across several layers will not be washed out or disappear by the time it reaches the network's end.

Performance measures such as precision, accuracy, recall, macro- average and weighted average have been used to put the model to the test. The proposed model's an overall multi-class classification accuracy is 95.80%, while classifying various phases of the DR.

The developed model's future scope is to create a model that uses the transfer learning principle to improve classification results. One of the future goal of this research is to introduce a DenseNet model for hierarchical classification of the DR levels.

References

1. P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A.A. Motala, K. Ogurtsova, J.E. Shaw, D. Bright, R. Williams, Global and Regional Diabetes Prevalence Estimates for 2019 and Projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Research and Clinical Practice **157**, 107843 (2019). https://doi.org/10.1016/j.diabres.2019.107843, https://www.sciencedirect.com/science/article/pii/S0168822719312306
2. R. Raman, P.K. Rani, S.R. Rachepalle, P. Gnanamoorthy, S. Uthra, G. Kumaramanickavel, T. Sharma, Prevalence of diabetic retinopathy in India: Sankara Nethralaya diabetic retinopathy epidemiology and molecular genetics study report 2. Ophthalmology **116**(2), 311–318 (2009). https://doi.org/10.1016/j.ophtha.2008.09.010
3. S. Resnikoff, V.C. Lansingh, L. Washburn, W. Felch, T.M. Gauthier, H.R. Taylor, K. Eckert, D. Parke, P. Wiedemann, Estimated number of ophthalmologists Worldwide (International Council of Ophthalmology Update): will we meet the needs? British J. Ophthalmol. **104**(4), 588–592 (2020). https://doi.org/10.1136/bjophthalmol-2019-314336, https://bjo.bmj.com/content/104/4/588
4. S.D. Solomon, M.F. Goldberg, ETDRS grading of diabetic retinopathy: still the gold standard? Opthalm. Res. (2019). https://doi.org/10.1159/000501372
5. K. Bhatia, S. Arora, R. Tomar, Diagnosis of diabetic retinopathy using machine learning classification algorithm, in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)* (2016) pp. 347–351. https://doi.org/10.1109/NGCT.2016.7877439
6. G. Saxena, D.K. Verma, A. Paraye, A. Rajan, A. Rawat, Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets. Intell.-Based Med. **3–4**, 100022 (2020)
7. A. Bora, S. Balasubramanian, B. Babenko, S. Virmani, S. Venugopalan, A. Mitani, G. de Oliveira Marinho, J. Cuadros, P. Ruamviboonsuk, G.S. Corrado, L. Peng, D.R. Webster, A.V. Varadarajan, N. Hammel, Y. Liu, P. Bavishi, Predicting the risk of developing diabetic retinopathy using deep learning. Lancet Dig. Health **3**(1), e10–e19 (2021). https://doi.org/10.1016/S2589-7500(20)30250-8, https://www.sciencedirect.com/science/article/pii/S2589750020302508
8. K. Shankar, A.R.W. Sait, D. Gupta, S. Lakshmanaprabu, A. Khanna, H.M. Pandey, Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recogn. Lett. **133**, 210–216 (2020), https://doi.org/10.1016/j.patrec.2020.02.026, https://www.sciencedirect.com/science/article/pii/S0167865520300714
9. T. Shanthi, R. Sabeenian, Modified alexnet architecture for classification of diabetic retinopathy images. Comput. Electr. Eng. **76**, 56–64 (2019)
10. G. Quellec, K. Charrière, Y. Boudi, B. Cochener, M. Lamard, Deep image mining for diabetic retinopathy screening. Dical Image Anal. **39**, 178–193 (2017). https://doi.org/10.1016/j.media.2017.04.012, https://www.sciencedirect.com/science/article/pii/S136184151730066X
11. S. Sengupta, A. Singh, H.A. Leopold, T. Gulati, V. Lakshminarayanan, Ophthalmic diagnosis using deep learning with fundus images–a critical review. Artif. Intell. Med. **102**, 101758 (2020)
12. G.T. Reddy, S. Bhattacharya, S. Siva Ramakrishnan, C.L. Chowdhary, S. Hakak, R. Kaluri, M. Praveen Kumar Reddy, An ensemble based machine learning model for diabetic retinopathy classification, in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pp. 1–6 (2020). https://doi.org/10.1109/ic-ETITE47903.2020.235

13. Y.S. Kanungo, B. Srinivasan, S. Choudhary, Detecting diabetic retinopathy using deep learning, in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)* 801–804 (2017). https://doi.org/10.1109/RTEICT.2017.8256708
14. T.R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P.K.R. Maddikunta, G. Srivastava, Deep neural networks to predict diabetic retinopathy (2020), pp. 1868–5145. https://doi.org/10.1007/s12652-020-01963-7
15. W. Zhang, J. Zhong, S. Yang, Z. Gao, J. Hu, Y. Chen, Z. Yi, Automated Identification and grading system of diabetic retinopathy using deep neural networks. Knowl.-Based Syst. **175**, 12–25 (2019). https://doi.org/10.1016/j.knosys.2019.03.016https://www.sciencedirect.com/science/article/pii/S0950705119301303
16. S. Gayathri, A.K. Krishna, V.P. Gopi, P. Palanisamy, Automated binary and multiclass classification of diabetic retinopathy using haralick and multiresolution features. IEEE Access **8**, 57497–57504 (2020). https://doi.org/10.1109/ACCESS.2020.2979753
17. J. Sahlsten, J. Jaskari, J. Kivinen, L. Turunen, E. Jaanio, K. Hietala, K. Kaski, Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. Sci. Rep. 2–11 (2019). https://doi.org/10.1038/s41598-019-47181, https://www.nature.com/articles/s41598-019-47181-wciteas
18. R. Sarki, K. Ahmed, H. Wang, Y. Zhang, Automated detection of mild and multi-class diabetic eye diseases using deep learning. Health Inf. Sci. Syst. (2020). https://doi.org/10.1007/s13755-020-00125-5, https://link.springer.com/article/10.1007/s13755-020-00125-5citeas
19. S. Karthikeyan, K.P. Sanjay, R.J. Madhusudan, S.K. Sundaramoorthy, P.K. Namboori, Detection of multi-class retinal diseases using artificial intelligence: an expeditious learning using deep CNN with minimal data. Biomed Pharmacol J. (2019)
20. Y.H. Li, N.N. Yeh, S.J. Chen, Y.C. Chung, Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network. Mob. Inf. Syst. (2020). https://doi.org/10.1155/2019/6142839
21. W. Huang, G. Song, .M.W., Xie, K.: Adaptive weight optimization for classification of imbalanced data
22. G. Kalyani, B. Janakiramaiah, A. Karuna, L.N. Prasad, Diabetic retinopathy detection and classification using capsule networks. Complex Intell. Syst. (2021). https://doi.org/10.1007/s40747-021-00318-9
23. B. Charu, J. Shruti, S. Meenakshi, *Deep Learning-Based Diabetic Retinopathy Severity Grading System Employing Quadrant Ensemble Model* (J. Digit, Imag, 2021)
24. M. Badar, M. Haris, A. Fatima, Application of deep learning for retinal image analysis: a review. Comput. Sci. Rev. **35**, 100203 (2020)
25. W.L. Alyoubi, W.M. Shalash, M.F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: a review. Inf. Medi. Unlocked **20**, 100377 (2020)
26. American Academy of Ophthalmology, https://www.kaggle.com/ratthachat/aptos-eye-preprocessing-in-diabetic-retinopathy
27. A. Brownlee, Imbalanced classification with python better metrics, balance skewed classes, cost-sensitive learning. Machine Learning Mastery Pty. Ltd., PO Box 206, Vermont Victoria 3133, Australia (2020)
28. E.E. Reber, R.L. Michell, C.J. Carter, *Kaggle Diabetic Retinopathy Detection Competition Report* (Tech. rep, Kaggle, 2015)
29. G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243
30. T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, H. Kang, Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Inf. Sci. **501**, 511–522 (2019). https://doi.org/10.1016/j.ins.2019.06.011. https://www.sciencedirect.com/science/article/pii/S0020025519305377

31. H. Pratt, F. Coenen, D.M. Broadbent, S.P. Harding, Y. Zheng, Convolutional neural networks for diabetic retinopathy. Proced. Comput. Sci. **90**, 200–205 (2016). https://doi.org/10.1016/j.procs.2016.07.014, https://www.sciencedirect.com/science/article/pii/S1877050916311929, 20th Conference on Medical Image Understanding and Analysis (MIUA 2016)
32. Y.P. Liu, Z. Li, C. Xu, J. Li, R. Liang, Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. Artif. Intell. Med. **99**, 101694 (2019)

Performance Analysis of Filter Bank Multicarrier System for 5G Networks

I. S. Akila, C. Elakkiya, M. Mohana Priya, B. Nivedha, and S. Yadhaarshini

Abstract 5G technology aims to provide higher data rates, low latency, improved system capacity and, increased reliability for its users. The sync shaped spectrum in Orthogonal Frequency Division Multiplexing (OFDM) leads to large out of band emissions and reduces spectral efficiency. Filter Bank Multicarrier (FBMC) is one of the new waveforms best-suited candidates for 5G technology. The FBMC is well localized in the time and frequency domain. FBMC filters in a subcarrier basis using a basic pulse shaped prototype filter and it is insensitive to Carrier Frequency Offset (CFO). This paper presents the benefits of FBMC compared to OFDM using MATLAB simulation across the parameters such as Bit Error Rate (BER) and Power Spectral Density (PSD). FBMC was also generated using SMW 200A Vector Signal Generator and analyzed its performance using FSW43 Signal Spectrum analyzer.

Keywords FBMC · SMW 200A Vector Signal Generator · FSW43 Signal · Spectrum Analyzer · PSD · BER · OFDM

1 Introduction

The emerging 5G wireless technology that contributes significant achievements in the history of wireless communications has to be designed in such a way that it meets the needs of the growing population of mobile consumers and must be

I. S. Akila (✉) · C. Elakkiya · M. Mohana Priya · B. Nivedha · S. Yadhaarshini
Department of ECE, Coimbatore Institute of Technology, Coimbatore 641014, India
e-mail: akila@cit.edu.in

C. Elakkiya
e-mail: 1804081ece@cit.edu.in

M. Mohana Priya
e-mail: 1904211ece@cit.edu.in

B. Nivedha
e-mail: 1804103ece@cit.edu.in

S. Yadhaarshini
e-mail: 1804126ece@cit.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_31

capable of providing robust service against dynamic constraints.5G technology will be enforced with exciting features such as a user data rate of around 100 Mbps, around 1ms decreased latency time, and a peak data rate of around 20 Gbps. These features render users with good coverage and high speed.5G technology can make changes in mobile communication by producing ultra-high-speed data rates. The extraordinary attributes such as low latency connection, good audio-video refinement of multimedia choices, and added gaming features will attract people across the world towards this technology. The three main use case categories of 5G are enhanced Mobile Broad Band (eMBB), massive Machine-Type Communications (mMTC), and Ultra-Reliable Low Latency Communications (URLLC) having 20Gbps peak rate, 106/km^2 device density and less than 1 ms latency. To achieve this, a flexible air interface is required. The main component of any air interface is a waveform. Hence the waveform has to be designed in such a way to facilitate such flexibility. The essential design criteria to meet these requirements of 5G are High spectral efficiency, Low latency, High reliability, Massive asynchronous transmission, Low device complexity, Low device complexity, and High energy efficiency.

In 5G technology, FBMC is one of the efficient modulation schemes in which multicarrier techniques are employed. Multicarrier techniques are free from Inter Symbol Interference (ISI) and they are also resistant to multipath fading. In addition to this, FBMC has the ability to perform more efficiently and systematically than OFDM which is a waveform in 4G technology.

2 Related Works

One of the most popular information security approaches is Symmetric key cryptography which is employed to transmit data along with the internet safely [1]. In [2], To meet the next-generation mobile communication network's requirement of high data rates with optimal use of spectrum many research works are performed to design a new waveform. In [3], a wireless mesh network with a wideband CDMA technique was implemented and it can be used in military and consumer-based applications. In [4], Spectral efficiency in OFDM gets decreased when there was a decrease in the number of subcarriers, but FBMC had increased spectral efficiency even with a reduced number of subcarriers. In [2], a comparison was made with OFDM, FBMC, and F-OFDM waveforms and inferred that FBMC has less BER and it has less sensitivity to CFO. In [4], Throughput in FBMC gets improved when compared to OFDM because FBMC had large bandwidth availability and usability and also FBMC did not employ CP overhead as in OFDM, thus Out of Band (OOB) emission in FBMC was very low when compared with CP–OFDM. In [2], A drawback is that OFDM has good Multiple Input Multiple Output (MIMO) compatibility whereas FBMC has less MIMO compatibility. In [5], FBMC–QAM system was proposed with two prototype filters in which MIMO and channel estimation schemes can be utilized as in OFDM. In [6], discusses vehicle to vehicle communication, to ensure road safety, delay and outage critical data are the key factors considered. To achieve high bandwidth and network capacity in cars, the design of the new waveform in 5G is essential. Fifth generation (5G) mobile networks are supported with higher channel bandwidth, low latency, high-capacity architecture, and cost-effectiveness. An efficient way of fulfilling the above requirements is to make use of an optical-fiber-based infrastructure where multiple wireless services can be used in the same fiber to Remote Radio Head (RRH) sites. By the usage of Generalized Frequency Division Multiplexing and Universally Filtered Orthogonal Frequency Division Multiplexing, simultaneous 4G and 5G transmission are achieved as is shown in [7]. The work in [8], provides a comparison among Filter Bank Multicarrier, Generalized Frequency Division Multiplexing, Universal Filtered Multicarrier, and Resource Block Filtered Multicarrier for 5G communication. An overview regarding the generation of the above waveforms, their advantages and disadvantages are given. The performance of the above waveforms is analyzed by considering certain factors like time and spectral efficiency, robustness to time-frequency misaligned users, numerical complexity, and resilience to power amplifier non-linearity. FBMC is said to be one of the best waveforms for 5G by performing the analysis. In the work presented in [9], different types of multicarrier modulation techniques, like high power amplifier (HPA), Non-Linear Distortion (NLD), and the Bit Error Rate (BER) are introduced. The BER performance of OFDM and FBMC/OQAM modulation was analyzed in the presence of Additive White Gaussian Noise (AWGN). By using the simple HPA model technique, it gives a soft envelope limiter, and also easy to compute the BER theoretical expression. The FBMC based Massive MIMO technique is one of the popular techniques used in 5G communication technology [10]. By using this Massive MIMO technique, the number of subcarriers required in the system can be reduced and provides a better understanding of waveform with emphasis on FBMC based Massive MIMO networks. The paper [11, 12] depicts, Filter Bank Multicarrier with offset quadrature amplitude modulation (FBMC/OQAM) as one of the popular non-orthogonal waveform techniques employed in 5G transmissions. FBMC/OQAM is a multicarrier technique based on a set of filter bank filtering and offsets quadrature amplitude modulation. FBMC has a guard carrier protection with the same performance as CP-OFDM.FBMC is more suitable for 5G uplink multi-user asynchronous transmission when we compared it to CP-OFDM.

3 OFDM

Present wireless and telecommunications systems employ OFDM modulation scheme which uses encoding digital data on multiple carrier frequency approach. OFDM scheme yields high data rates and minimizes multipath fading and thus becomes a desirable choice in wideband digital communication, 4G technology, and audio broadcasting. OFDM uses a greater number of closely spaced subcarriers that are orthogonal to each other with overlapping spectra in which data are transmitted in parallel. Here each subcarrier is modulated at a low symbol rate. This makes OFDM less sensitive to frequency selective fading.

A Guard interval is inserted between OFDM symbols to eliminate ISI and the need for a pulse-shaping filter. It is also used to reduce the sensitivity to time synchronization problems. During this guard interval, a cyclic prefix is allowed to transmit over the channel. Cyclic Prefix (CP) used in OFDM tends to decrease the spectral efficiency.

Figure 1 shows the transmitter and receiver block of OFDM. At the transmitter side, serial to parallel conversion takes place to load the symbols on the subcarrier. *N* point IFFT is taken for the loaded symbols to have '*k*' sample.

$$x(k) = [1/N] * \sum_{l=0}^{N-1} X(l) e^{j2\pi kl/N} \tag{1}$$

Equation (1) represents the IFFT expression where *k* is the number of samples, '*l*' is the subcarrier and '$X(l)$' is the data symbol.

The CP is added to eliminate ISI effects. The samples are then fed to the parallel to serial converter and propagated over the channel. Equation (2) represents the OFDM transmitted signal from (3)

$$x(t) = \sum_{k=0}^{M-1} x(k)(l + k(N + Ncp)) \tag{2}$$

Where '*l*' is the number of subcarriers, '$X(l)$' is the data symbol, '*M*' is the number of OFDM symbols, 'Ncp' is the number of cyclic prefix and *N* is the number of data symbols transmitted.

On the receiver side, inverse operation of transmitter is performed to obtain the original signal 'S_n' as represented in Eq. (3).

$$S_n = h_n * x_n + n \tag{3}$$

where x_n is the data transmitted over the channel h_n and *n* is the white Gaussian noise in the time domain.

Orthogonality in OFDM is maintained by having subcarrier spacing as *f=i*/*T* Hz, where '*T*' is the symbol duration which is equal to the size of the receiver size window, and '*i*' is the integer which is equal to 1. Also, the frequency deviation in subcarriers leads to inter carrier interference. To maintain the orthogonality between

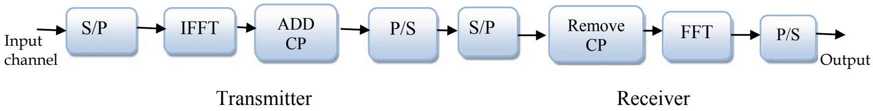


Fig. 1 OFDM transceiver

the subcarriers, there should be a synchronization between time and frequency, and they are highly sensitive to CFO which leads to ISI and ICI. Hence OFDM cannot meet the requirements of 5G. Therefore, an alternative waveform must be designed for 5G technology.

4 FBMC

FBMC is a type of multicarrier modulation that has been developed to overcome the drawbacks of OFDM. The redundancy in OFDM reduces the throughput of the system and power wastage during transmission process. The FBMC employs a bank of filters where each subcarrier is filtered resulting in the reduction of side lobes and cleaner subcarriers. Filtration of each subcarrier is done using a filter that has an amplitude and phase response. The design of the filtering should emphasize the orthogonality of waveforms so as to reduce ISI and ICI.

In FBMC, the prototype filter is designed for zero frequency carriers and the other filters are deduced from it through frequency shifts. The number of multicarrier symbols that overlap in the time domain is defined as the overlapping factor 'K'. The prototype filters are characterized by the overlapping factor and the filter order can be chosen as $2{*}K - 1$ where K takes the values from 2 to 4 as per the PHYDYAS project.

Figure 2 shows the Transmitter block of FBMC. The data bits are given as inputs. The symbol mapping is done using QAM modulation. Offset Quadrature Amplitude Modulation (OQAM) is used to attain lower sidelobes and utilization of total channel capacity. But this requires complex receivers for broadband channels. Here, real and imaginary parts of the data are not transmitted simultaneously while a delay of half symbol duration is introduced to overcome ISI. Upsampling is performed based on the overlapping factor K. Frequency spread spectrum is done to convert a narrowband signal to wideband spectrum which results in occupation of larger spectral bandwidth. The entire frequency band is subdivided into smaller frequency bands and the signals change their carrier frequency among the distinct frequencies. Frequency spreading helps in the reduction of interference and makes jamming difficult. This makes the spread spectrum less sensitive to narrow band interference. Thus, this is used in military applications where safety is the main concern because the interpretation of the signal or the information becomes more difficult. Then IFFT is computed for the transmitted symbols which are then added and transmitted over a channel. Frequency spreading followed by IFFT provides the following benefits low PAPR, low complexity, frequency domain equalization, and good time localization.

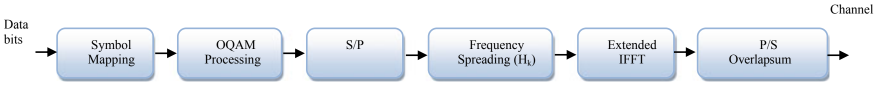


Fig. 2 FBMC Transmitter

Eq. (4) represents the FBMC channel output:

$$C_i(t) = \int_{-\infty}^{\infty} h(t-\alpha) f_i(\alpha) \tag{4}$$

where '$h(t)$' is the channel impulse response, '$f_i(\alpha)$' is the filter impulse response For ISI free transmission two symbols are shifted in time, in the same subcarrier as represented in Eq. (5).

$$\int_{-\infty}^{\infty} C_i(t)C_i(t-Kt)\mathrm{d}t = 0, \quad \text{for } \forall\, i \text{ and } K = \pm 1, \pm 2, \ldots \tag{5}$$

Where K is the time shift

Equation (6) should be satisfied for ICI free transmission. The ICI is calculated between different subcarriers and it must be zero.

$$\int_{-\infty}^{\infty} C_i(t)C_j(t-Kt)\mathrm{d}t = 0, \text{ for } \forall\, i, j, i \neq j \text{ and } K = \pm 0, \pm 1, \pm 2 \ldots \tag{6}$$

For ISI and ICI free transmissions filter coefficients are presented in Eq. (7).

$$C_i^2(f)H^2(f) = A_i + Q_i(f) > 0, \text{ for } fi - fs < f < fi + fs \tag{7}$$

$C_i(f)$ is the amplitude response
$H(f)$ is the transfer function of the filter
A_i is the arbitrary constant
Q_i is the Shaping function
Q_i must have odd symmetries about $fi - \frac{fs}{2}$ and $fi + \frac{fs}{2}$

Figure 3 shows the receiver structure of FBMC. The FFT operation is performed. The data are recovered by performing a weighted dispreading operation followed by OQAM post-processing and symbol demapping. The FBMC receiver performs the inverse operation done at the transmitter side. FBMC exhibits improved throughput, high spectral efficiency, and decreased out of band emissions. FBMC has higher spectral efficiency and thus it can provide higher data rates which are essential for 5G communication.

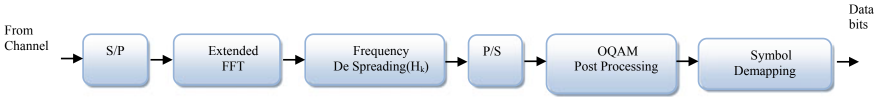


Fig. 3 Shows the receiver structure of FBMC

Table 1 Simulation parameters

Parameter	Values
Number of FFT	1024
Number of Guards	212
Overlapping Factor (K)	2, 3, 4
Number of symbols	1000
Modulation scheme	2:4QAM 4:16QAM 6:64QAM 8:256QAM
SNR dB	−5 to 50

5 Simulation Parameters

Table 1 presents the simulation parameters used in MATLAB Simulation of the FBMC.

6 Simulation and Results Analysis

The FBMC transmitter and receiver structure is simulated using MATLAB for the aforementioned simulation parameters and is compared with OFDM for Bit Error Rate (BER) & Power Spectral Density (PSD). Figures 4 and 5 illustrate the simulated results of PSD plots of FBMC and OFDM and is inferred that out of band emissions of OFDM begins at −9dB, for FBMC begins at nearly −100dB. There, a −91dB

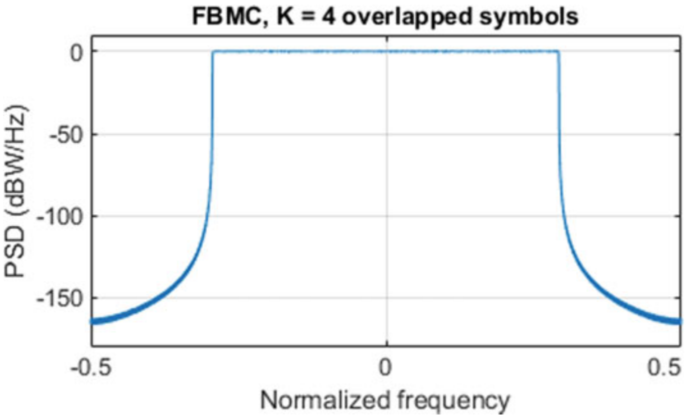
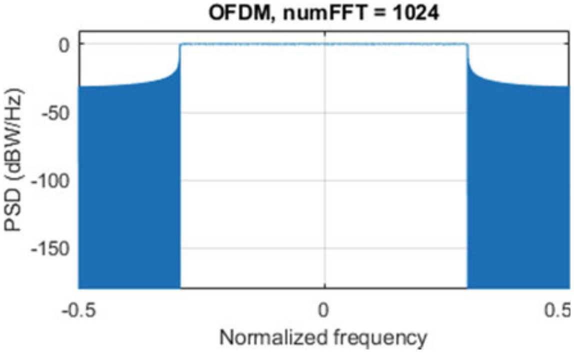


Fig. 4 Power spectral density of FBMC

Fig. 5 Power spectral density of OFDM



difference has been observed, and also it is inferred that reduced side lobes of FBMC that lead to improved utilization of the available spectrum. Thus, higher spectral efficiency achieved in FBMC in comparison with OFDM has been shown through the results.

Bit Error Rate (BER) is the other metric taken for the analysis of the FBMC system for M-ary QAM modulation schemes and it is the number of errors bits relative to the total number of bits received. BER is generally expressed as a negative power of 10. A BER of 10^{-3}indicates that out of 1000 bits transmitted 1 bit has an error. The transmitted and received bits are compared and the total number of error bits is calculated. The SNR Vs BER plot is plotted for different values of modulation schemes like 4QAM, 16QAM, 64QAM, and 256QAM. From Fig. 6 it is observed

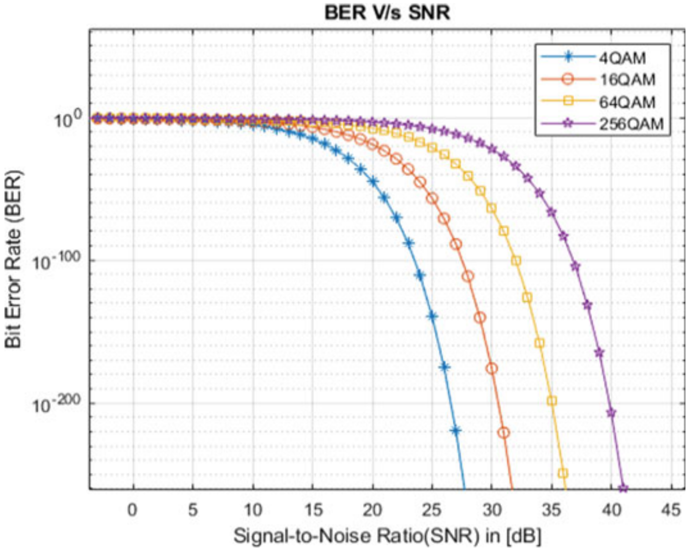


Fig. 6 Bit error rate—FBMC

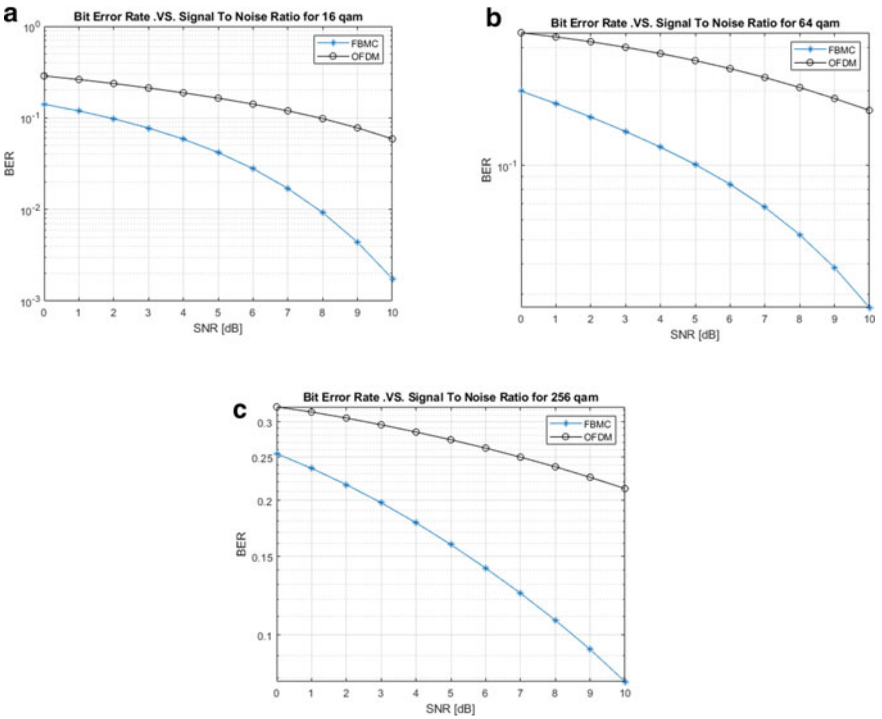


Fig. 7 **a** BER plot–FBMC versus OFDM for 16QAM, **b** BER plot–FBMC versus OFDM for 64QAM, **c** BER plot–FBMC versuss OFDM for 256QAM

that as the Signal to Noise Ratio increases, BER decreases as less noise is observed in the channel. On comparing the BER plots for M-ary QAM modulation schemes, it was inferred that as the value of the M-ary keying in the QAM increases, BER also increases, and it has been observed that 4QAM has the least BER. Figure 7 exhibits that, FBMC has less BER when compared to OFDM for 16QAM, 64QAM, and 256 QAM modulation schemes respectively.

7 System Parameters

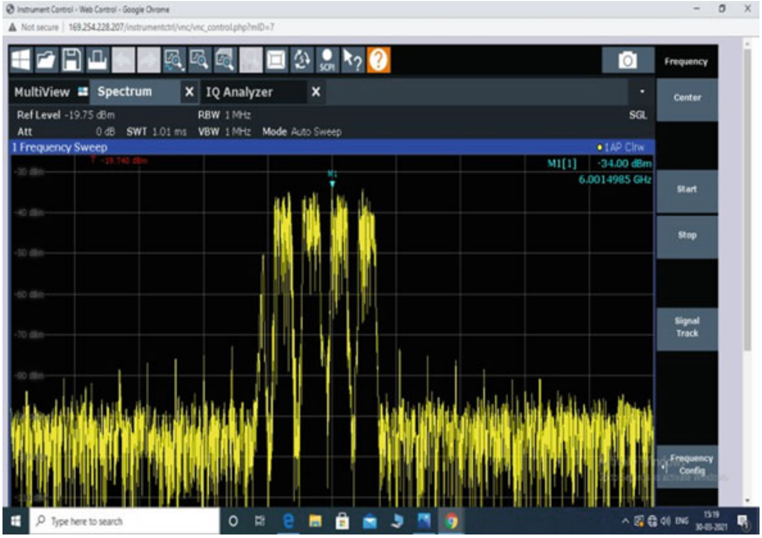
Table 2 presents the parameters used in the implementation of FBMC under two scenarios using SMW 200A Vector signal generator and FSW 43 Signal Spectrum analyzer.

Table 2 Parameters used in hardware implementation

Parameter	Scenario I	Scenario II
Number of subcarriers	64	64
Subcarrier spacing	15 kHz	312.5 kHz
Cyclic prefix length	16 samples	16 samples
Occupied subcarriers	53	53
Sequence length	10 symbols	10 symbols
Occupied bandwidth	150 kHz	16.563 MHz
Overlapping factor(K)	4	4
Number of right side guard bands	27	5
Number of left side guard bands	27	6
Filter	Raised cosine	Raised cosine
Sampling rate	960 kHz	20 MHz

8 Implementation and Result Analysis

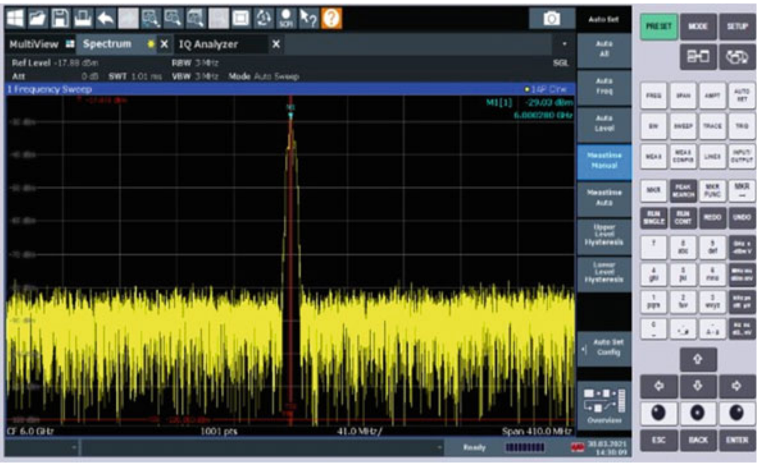
Using SMW 200A Vector Signal Generator, FBMC has been generated and the signal is analyzed with FSW-Spectrum Analyzer. Based on the choice of Scenario I& II as shown in Table 2, the spectral plot of FBMC is represented in Figs. 8 and 9. The IQ-tar file has been extracted from FSW-Spectrum Analyzer and is given as the input



Ref Level -19.75 dBm RBW 1 MHz SGL
Att 0 dB SWT 1.01 ms VBW 1 MHz Mode Auto Sweep

M1[1] -34.00 dBm
6.0014985 GHz

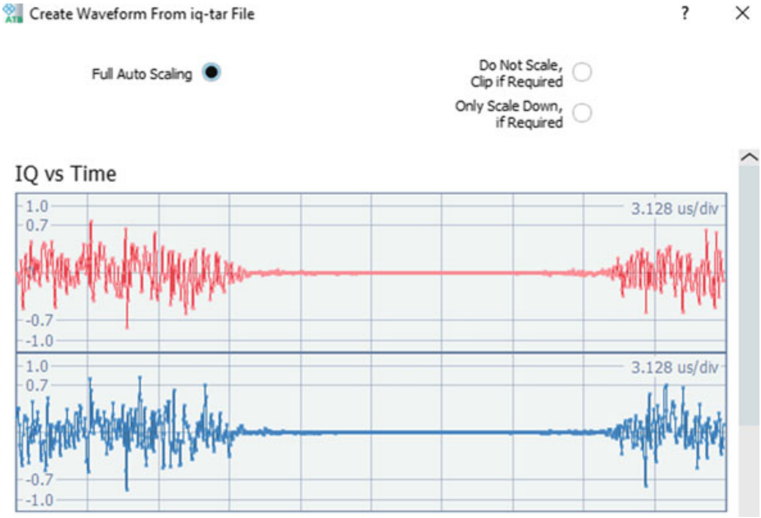
Fig. 8 FSW—Spectral Plot-Scenario 1



Ref Level -17.88 dBm RBW 3 MHz
Att 0 dB SWT 1.01 ms VBW 3 MHz Mode Auto Sweep
M1[1] -29.03 dBm
6.000280 GHz
CF 6.0 GHz 1001 pts 41.0 MHz/ Span 410.0 MHz
30.03.2021
14:30:09

Fig. 9 FSW—Spectral Plot-Scenario II

to the ARB toolbox to generate the plots of IQ, Spectrum, and its constellation as shown in Figures 9, 10, 11 and 12 for scenario I and scenario II.


Create Waveform From iq-tar File

Full Auto Scaling ●

Do Not Scale, Clip if Required ○

Only Scale Down, if Required ○

IQ vs Time

Fig. 10 IQ plot—Scenario I

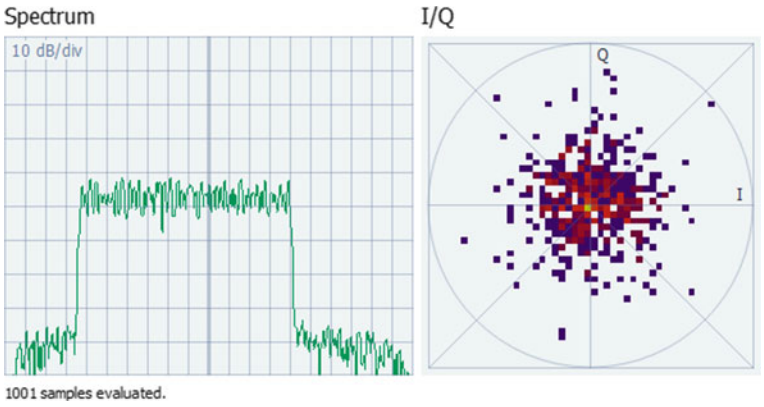


Fig. 11 Spectrum and IQ Constellation—Scenario I

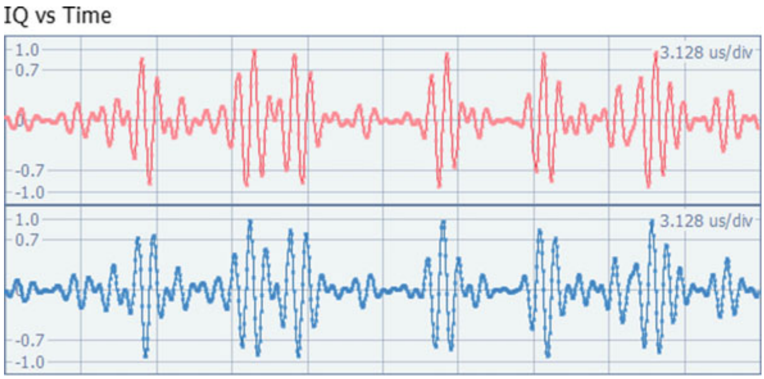


Fig. 12 IQ plot—Scenario II

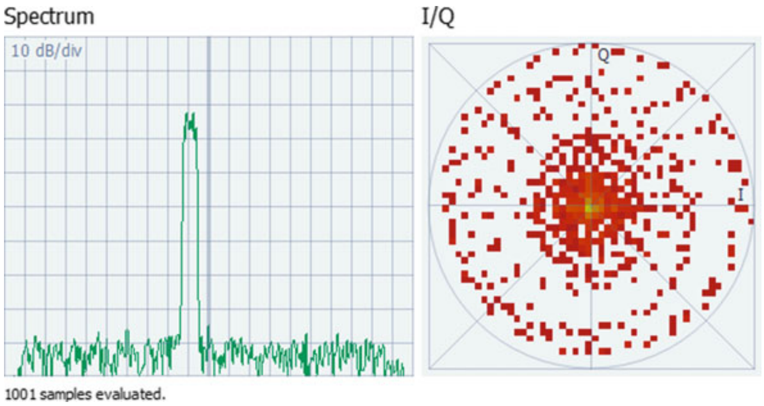


Fig. 13 Spectrum & IQ Constellation—Scenario II

9 Conclusion

Implementation of 5G technology in the field of communication brings a wide variety of benefits for its users. To accommodate the large scope of desirable features, proficiency in 5G technology has to be largely improved when compared to former technology. Since the transmitter and receiver side of FBMC is employed with a filter bank, desirable 5G technology features such as high data rate, high reliability, and low latency are established efficiently in FBMC systems, whereas in OFDM systems, these features cannot be achieved. OFDM is highly sensitive to ISI and ICI. FBMC systems are very less sensitive to Carrier Frequency Offset. Hence ISI and ICI are eliminated in FBMC. Also, it has been observed that FBMC shows better spectral efficiency than OFDM under the tested scenarios. The results were analyzed for a range of SNR values, and it is seen that a higher SNR rate leads to a decrease in BER. Also, by examining the plots relating BER and PSD values, FBMC makes an effective and efficient technology for 5G communications. Thus, FBMC waveform is the best-suited waveform for 5G communication which meets all the requirements of 5G communications.

Acknowledgments We would like to acknowledge the support and guidance provided by "TEQIP III Sponsored 5G Innovation Centre", Department of ECE, Coimbatore Institute of Technology, Coimbatore for the utilization of SMW 200A, Vector Signal Generator & FSW Signal, and Spectrum Analyzer to carry out our research work in FBMC.

References

1. R. Chatterjee, R. Chakraborty, J.K. Mandal,Design of cryptographic model for end- to-end encryption in FPGA based systems, in *IEEE Proceedings of 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (2019), pp. 459–465
2. A. Bedoui, M. Et-tolba, A comparative analysis of filter bank multicarrier (FBMC) as 5G multiplexing technique, in *IEEE Proceedings of International Conference on Wireless Network & Mobile Communication (WINCOM)* (2017). https://doi.org/10.1109/WINCOM.2017.8238200
3. D. Sivaganesan, Improvisation of mesh network with wideband code division multiple access. IRO J. Sustain. Wireless Syst. **03**, 198–205 (2019)
4. R. Nissel, S. Schwarz, M. Rupp, Filter bank multicarrier modulation schemes for future mobile communications. IEEE J. Sel. Areas Commun. **35**(8), 1768–1782 (2017)
5. H. Nam, M. Choi, S. Han, C. Kim, S. Choi, D. Hong, A new filter-bank multicarrier system with two prototype filters for QAM symbols transmission and reception. IEEE Trans. Wireless Commun. **15**(9), 5998–6009 (2016)
6. S. Schwarz, T. Philosof, M. Rupp, Signal processing challenges in cellular assisted vehicular communications. IEEE Signal Process. Mag. **34**(2), 47–59 (2017)
7. Michailow, A. Festag, 5G now: Non-orthogonal, asynchronous waveforms for future mobile applications. IEEE Commun. Mag. **52**(2), 97–105 (2014)
8. M. Van Eeckhaute, A. Bourdoux, P. de Doncker, F. Horlin, Performance of emerging multicarrier waveforms for 5G asynchronous communications. J. Wireless Commun. Netw. (Springer) (2017) https://doi.org/10.1186/s13638-017-0812-8

9. H. Bouhadda, H. Shaiek, D. Roviras, R. Zayani, Y. Medjahdi, R. Bouallegue, Theoretical analysis of BER performance of nonlinearly amplified FBMC/OQAM and OFDM signals. EURASIP J. Adv. Signal Process https://doi.org/10.1186/1687-6180-2014-60 (2014)
10. A. Farhang, J.P. Marchetti, N. Figueiredo, F. Miranda, Massive mimo and waveform design for 5th generation wireless communication systems, in *1st International Conference on 5G for Ubiquitous Connectivity (5GU)*, vol. 28, no. 3 (2014), p. 7075
11. D. Liu, Y. Liu, Z. Zhong, D. Miao, Z. Zhao, H. Guan, 5G uplink performance of Filter Bank Multi carrier, in *The Proceeding of IEEE International Conference on the Wireless Network and Mobile Communication.* 978/liu2016/5090/1345 (2016)
12. Y.H. Yun, C. Kim, K. Kim, Z. Ho, B. Lee, J.-Y. Seol, A new waveform enabling enhanced QAM-FBMC systems, in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (2015), pp. 116–120

Challenging the Network Privacy of Beaker Browser

Aishvarya Nair, P. P. Amritha, and V. Sarma

Abstract The Internet was introduced as decentralized platform for users to share data and communicate. However, the information on Internet is now handled and controlled by some global corporations, which gives them complete control of user data. The key to this problem was a decentralized web with client-only architecture that would give user more control over their information. Beaker Browser is one such initiative that allows peers to communicate and share websites without a central server intervention. Beaker works on a peer-to-peer technology and the websites hosted on it are digitally signed using a private key.The public key is then used as a URL. If the public key is shared only within a closed group, the website will remain private. While the web technology is on the cusp of being renewed, it has become all the more necessary to understand the threats associated with it. In this paper, we survey the features of Dat protocol and Beaker Browser. We also introduce a proof of concept that challenges the network privacy of the websites hosted on Beaker.

Keywords Dat protocol · Beaker Browser · Peer-to-peer · Decentralization · Network privacy · Semantic graph

1 Introduction

Although centralized web browsing was conceptualized to make data sharing easier, it has done so over the expense of user's data privacy and security. The proliferation of cyber threats and innovative ways used by malicious attackers has led to the question of whether it is time to adapt decentralized web browsing on a larger scale than it already is. An easier process to set up, centralized web browsing consists

A. Nair (✉) · P. P. Amritha (✉) · V. Sarma
TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2cys19009@cb.students.amrita.edu

P. P. Amritha
e-mail: pp_amritha@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_32

of architecture build around a single server supervising the processing, storage, and network functions accessible to the user. The central server accepts requests from the smaller workstations and processes the information accordingly. This architecture makes a web browsing facility that is consistent, efficient, and affordable. It required less IT management time to be devoted to it since there is only one server to maintain. Additionally, it also makes it easier to track and collect data since all of it is required to flow through one place. However, from a security point of view, the same reasons which make centralized web browsing efficient have made it unfit to tackle the constant cyber threats which threaten the security and privacy of users. The use of a single server also makes the architecture vulnerable since a malicious attacker only needs to take control of a single server to cause problems in the whole web browsing system. Additionally, the corporations controlling these servers can gain complete access to user data and use it for purposes suitable to their needs. The data passing through is often used to adjust the user experience and only share information deemed right by the companies, robbing users of their freedom of choice. Hence, a migration to the decentralized network has been deemed to be a feasible solution.

The main principle behind decentralized web browsing is to remove the dependency on one central server. The architecture of this system consists of multiple independent machines that are interconnected to provide a pool of resources for the user to access data from. Each machine can set its own set of instructions and restrictions for accessing data. Additionally, with the absence of a central server, there is no single potential point of failure or vulnerability for attackers to exploit, making it a more efficient system than centralized web browsing. The security system of decentralized web browsing consists of firewalls set up by each independent machine which would not compromise any other component of the system in the case of being hacked.

Although an ideal replacement for centralized web browsing, products with the concept of decentralized web browsing have been in scarcity. One such product is Beaker Browser. Launched in 2016, this peer-to-peer web product consists of multiple sets of experimental technologies integrated to give more control over the web. The main principles behind the product were to allow anyone to become a server, serve the same site with multiple computers, and mitigate the need for a back end. The client-server model of web browsing that is currently being used has concomitant threats to the user. The exposure of confidential data, link rots, and a massive volume of confidential information is collected and utilized to improve customer experience, whereas the obligation of protecting user information is not yet practiced effectively. The increasing cost of hosting a website also expropriates the liberty of users. Beaker Browser is considered the future of web browsers and emancipates the technology from centralized tyranny. It provides the functionality of sharing the website URL within a closed network. In this paper, we introduce an approach that can reveal private URLs. An increasing number of applications being designed on Beaker Browser leverage the benefits of peer-to-peer browsing. This new way of browsing addresses many issues of centralization and allows the users to share websites among themselves.

2 Related Work

Blockchain technology is expected to revolutionize the finance and banking sectors around the world. Sankar et al. [1] Although it is considered the epitome of decentralization, we need a lightweight technology without consensus. Dat is one such protocol that has some innovative features for an open web. An increasing number of applications being designed on Beaker Browser leverage the benefits of peer-to-peer browsing. There are several kinds of research that have concentrated on creating decentralized web applications. As emerging innovations continue to evolve, projects on peer-to-peer networks are a constant research subject.

Ogden et al. [2] dat protocol is an ongoing research work that is being designed to sync folders even though they are larger or constantly changing. It utilizes cryptographic hashes signature and encryption to securely transfer information. These properties make Dat a cryptographically secure protocol. The protocol takes inspiration from existing systems like IPFS, Git, and BitTorrent. The paper [3] introduces a prototype to communicate scholarly research modules in a peer-to-peer network. To benefit from the features of Dat, this prototype uses a decentralized register that is immutable for evaluating consuming and discovering research. Additionally, it facilitates constructive certification discussions and allows anyone with Internet access to participate. The decentralization of Dat protocol requires less dependency on institutions to maintain key data stores of related information and replaces it with widespread and distribution of that information. Jannes et al. [4] this paper aims to present key application scenarios and trends in various business domains that require a more client and data-centric web middleware for web applications in the edge that are decentralized and peer-to-peer. They define a set of key requirements for data operations in such middleware and elaborate them with some application cases. Furthermore, it discusses the current limitations of a browser as a peer-to-peer communication platform and also as complex decentralized applications. Finally, we conclude with a performance assessment of our prototype middleware for data and peer-to-peer applications. Naik et al. [5] A systematic characterization of next-level P2P (NL P2P) like Chord, Pastry, Kademlia, CAN, BitTorrent, Gnutella, Dat, etc., along with the examination of their key concepts is done in this paper. In this work, they study the different aspects of P2P overlay frameworks like routing, security, query, adaptation to non-critical failure, and so forth dependent on developed conventions. Further, they explore some upcoming challenges with NL P2P frameworks. A structured performance correlation of the protocol performance is also carried out. Artback [6] a comparison of Dat protocol and HTTP is carried out in this paper to understand which would be a good fit for video file delivery. The differences in scalability and bitrate and the challenges with implementing such a solution are explored. They have concluded with a study of Dat protocol showing good scalability and performance especially on a larger number of clients compared to HTTP [7] In this paper, they are proposing a blockchain framework with smart contracts to protect social media contents using IPFS, a modern decentralized file storage system. This shows that IPFS is immutable and also has additional features that can be used to enhace security.

3 Dat Protocol

The data is primarily stored and retrieved by a centralized infrastructure, and the client has to request it from a server. These client-server architectures are based on location-based addressing which further leads to link rots and content drifts. The address of this content points to the location where it is stored and whoever owns the location has control over the data. This means that if the content changes, there is no resort to access the older version, and if the location of the data is changed, the user is left with a broken link. Dat is a peer-to-peer protocol for sharing data between different computers that ameliorates the existing technology. This protocol eliminates the role of a central server by providing a distributed network for file sharing. Dat supports content-based addressing and version control which allows users to search for a specific content instead of searching a specific location. Many domains are also generated by attackers in client-server web technology to create the huge domain name system (DNS) traffic [8]. Dat efficiently deals with such problems as every information stored is identified with a unique key. This key can be used to locate the content on the network. When the data is shared by multiple peers, even if a user stops seeding it, the data will be available on the network. This p2p network reduces bandwidth costs on frequently shared files as it is distributed across all available computers. Comparison of Dat and HTTP protocol is shown in Table 1. Dat influences other intelligent techniques of offline data sharing, free publishing of data, and self- archiving. Dat is known to make file sharing faster easier and cheaper and is therefore used in various other applications. The key properties of Dat are (1) Content Integrity (2) Incremental versioning (3) Network privacy. Refer Fig. 1 to understand the architecture of Dat protocol.

3.1 Content Integrity

The potential of Dat protocol is to verify the version of data received is termed as content integrity. Additionally, content addressability is referring to a file by the hash of its content [2]. This lets users not only validate whether the version of data received is the one they require or not, but also lets people cite specific versions of the data by referring to the specific hash. The data uploaded to the Dat network are stored in

Table 1 Comparison of Dat and HTTP protocol

Properties	Dat protocol	HTTP protocol
Architecture	Peer-to-peer	Client-server
Bandwidth	Distributed	Depends on server capacity
Link rots/Content drift	Low chances (If no peer hosts it)	High chances
User privacy	High	Low

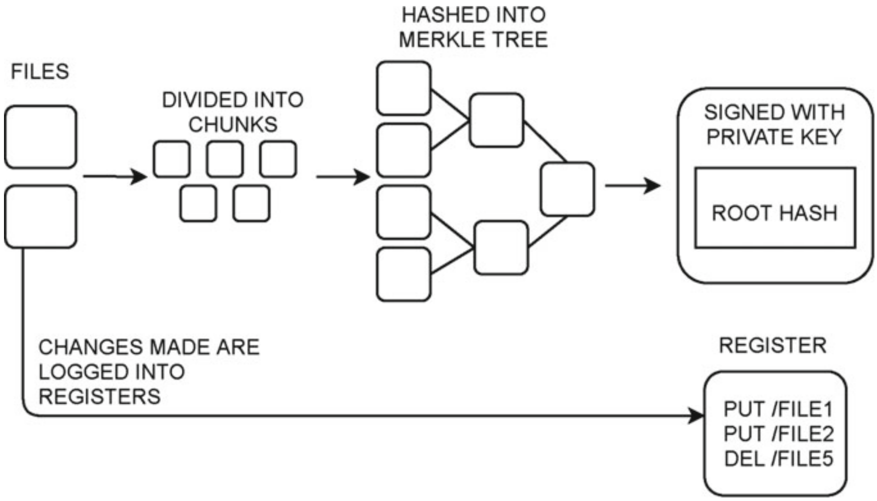


Fig. 1 Dat architecture

files on a file system called hyperdrive. Each of these files gets split up into some number of chunks. Refer Fig. [3] which are subsequently hashed into a Merkle tree where each child node has all the child nodes. The top hash is then signed using a private key. In Dat architecture as shown in Fig. 1, each filesystem receives a unique 32 bytes public key that can be used to identify the file on the network. This public key is used in Dat link to retrieve content and is in the form of dat://public key//suffix.

3.2 Incremental Versioning

In Dat, each file has a persistent public key that remains unaffected by bit-level changes [3]. Other peer-to-peer protocols such as InterPlanetary File System (IPFS) receives a different public key upon bit-level changes, and this creates the need to resharing of the keys after every change. The file system of Dat is based on append-only-log registers. Blockchain technology also provides a distributed, shared, and append-only ledger [9]; but the difference is that in blockchain, new records are added through consensus and smart contracts, and this protocol regards all its actions as commands to the filesystem, allowing all operations on it to be regarded as actions that are appended to a register [5] as shown in Fig. 1.

3.3 Network Privacy

During transport, all messages in the Dat protocol are encrypted and signed using the private key, refer Fig. 1, which ensures that unless one knows the public key, they will not have access to any members of the swarm for that Dat and can neither communicate with them. Anyone with the public key can verify whether the messages were created by a holder of the private key or not [2]. Every Dat repository has a corresponding private key which is stored in your home folder and never shared. The protocol never exposes your public or private key over the network. During the discovery phase, the BLAKE2b hash of the public key is used as the discovery key. We can calculate the discovery key easily from the public key [2]; however, if you only know a discovery key, you cannot work backward to find the corresponding public key. This prevents eavesdroppers from learning of Dat URLs by observing network traffic [10]. This protocol is also known as Hypercore protocol but in this paper, we will be referring to it as Dat.

4 Beaker Browser

Beaker Browser is an application of Dat Protocol and is an experimental peer-to-peer web browser [11]. In the client-server model used by current web browsers, the user agent stores all the browsing information. The users are obliged to rely on a central server for the safety of their data. Beaker Browser, thus, gives users more control over the web and significant rights on their online browsing information. Beaker Browser is built on decentralized technology and gives users the leverage of being a server and a client at the same time. The advantage of hosting the same website on multiple systems distributes the control among different peers and thereby minimizes the chances of broken links.

4.1 Dat Protocol in Beaker

Beaker Browser uses Dat protocol instead of HTTP to share the websites and, therefore, derives all the properties of Dat, i.e., the URL identifier is dat:// or hyper:// is used instead of http://. However, users can also visit HTTP websites on Beaker. The Dat-based websites hosted on Beaker are stored as files in the hyperdrive. The source files of these websites can be viewed and downloaded by anybody. The crux is to provide an open web for all users and, thus, build communities based on trust. The web pages are encrypted, hashed, and signed using a private key. The corresponding 32 bytes public key is then used in the URL to access the website. The web pages can be uploaded on the hyperdrive, and the hyper:// link can be shared around. The benefit here is how this enables private sharing. When you create a hyperdrive and share the link, then only people with the link can view the site. Beaker Browser, thus, utilizes the content integrity and network privacy property of Dat protocol to host and share the website among peers.

4.2 Hyperdrives

Traditionally, the websites are hosted on servers but Beaker makes the process extremely simple by using hyperdrives. Hyperdrive is built on the hypercore protocol and is called self-hosting servers. They are folders that can be hosted directly from any system. All the web pages are uploaded to hyperdrives, and every hyperdrive has a unique hyper URL as shown in Fig. 2. The concept of co-hosting helps in temporarily sharing other hyperdrives and contributing to the bandwidth. When we visit a website in Beaker, we share the bandwidth for the time being. Similar to Bit Torrent, this helps in keeping the websites up for the service and the peer count indicates the number of peers hosting the website. Beaker Browser, thus, completely removes the intervention of a central server. Beaker automatically creates a private "System Drive" for storing your private information. It contains your saved hyperdrives.

4.3 Fork and Copy of Websites

Only the owner of the websites can make changes to it but users can create their version of the website. Since hyperdrives are folders, we can create a copy of these websites and host it again. This gives the user the freedom to modify the interface and the functionalities as per their requirement. The two ways a website can be replicated are Fork and Copy. The forked version is linked to the parent website, whereas the copy creates an independent clone. During development, the forked version can be useful for deploying changes and merging them with the parent website. The fork

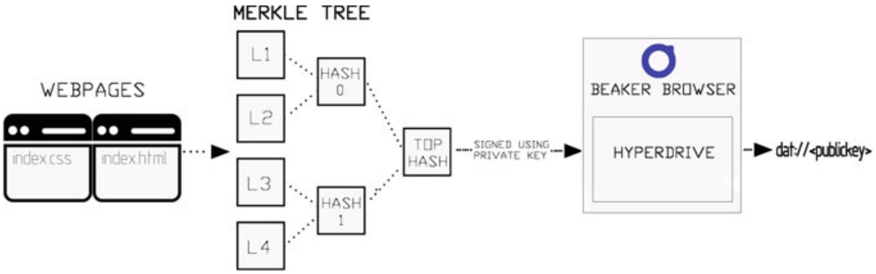


Fig. 2 Websites stored as hyperdrives

and the copy of any website have a unique URL like all other websites. This means the copy is signed with a different public key.

5 Proposed Work

Beaker Browser provides network privacy to the users by creating a unique URL for every website. Since there is no authentication mechanism used by this browser, if the user has the public key, they are considered eligible to view the content. The URL is unique to the content, and every website is independent of the other, but Beaker Browser stores the information of forked websites and their parent. We have proposed an approach that takes the input of this information and creates a graph of the URL's linking the parent and forked websites to each other. This semantic graph will help reveal the URLs of connected websites, thereby challenging the network privacy of users.

6 Implementation

The system drive is a private drive in Beaker that stores the public key of all the websites hosted. It also preserves the link of forked websites to its parents. If Beaker Browser replaces the traditional web browsers, there will be a huge number of websites used by the peers, and if the system drive is compromised in any attack, the data stored there can be used to create a massive semantic net of public keys. With no other authentication mechanism used by Beaker, the public keys can be used to access private websites. Our approach is to prove that if the data of the system drive is given as an input to the model, it will reveal the public keys of other systems as well. The proposed model takes system drive data as input. The data is stored in an SQL database and the model creates a semantic graph of all the public keys available. The model uses the python library netwokx to link these public keys.

For explaining the working of this approach, we have considered five systems. In each system, we have installed Beaker Browser. Now, we have created a root website 'Website 1' in 'System A.' Two forks of 'website 1' are created as 'website2' and 'website3.' 'website2' is shared with 'system B and 'website 3' is shared with 'system C.' Now a fork of website2 and website3 is created as 'website4' and 'website 5' and shared with 'system D' and system E, respectively. Refer Fig. 3.

7 Result Analysis

To demonstrate the breach of network privacy, we have taken the system drive information only of system A, system D, and system E and given it as an input to the

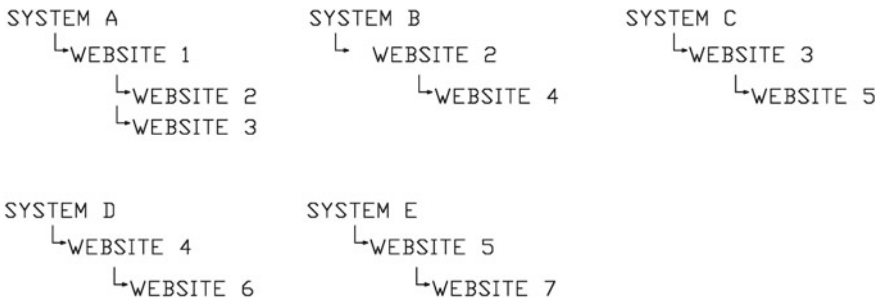


Fig. 3 Websites corresponding to the system

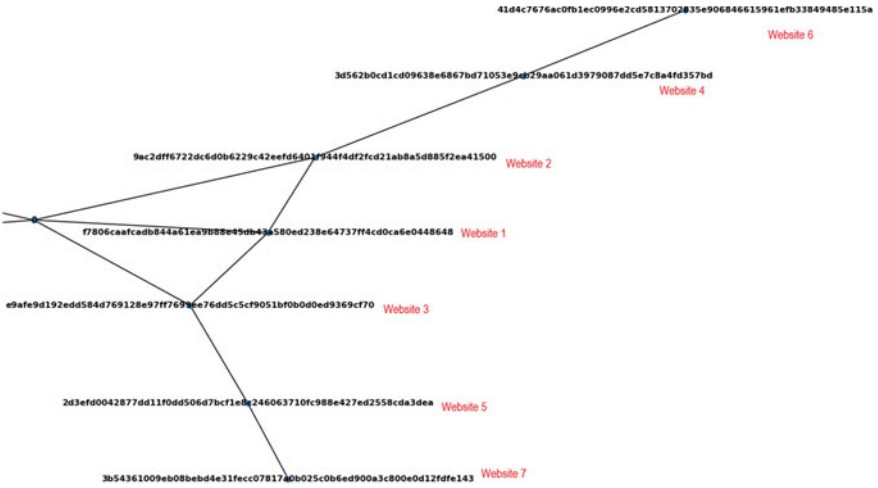

41d4c7676ac0fb1ec0996e2cd5813702835e906846615961efb33849485e115a
Website 6
3d562b0cd1cd09638e6867bd71053e9cb29aa061d3979087dd5e7c8a4fd357bd
Website 4
9ac2dff6722dc6d0b6229c42eefd6401f944f4df2fcd21ab8a5d885f2ea41500
Website 2
0
f7806caafcadb844a61ea9b88e45db43a580ed238e64737ff4cd0ca6e0448648
Website 1
e9afe9d192edd584d769128e97ff7699ee76dd5c5cf9051bf0b0d0ed9369cf70
Website 3
2d3efd0042877dd11f0dd506d7bcf1e8e246063710fc988e427ed2558cda3dea
Website 5
3b54361009eb08bebd4e31fecc07817a0b025c0b6ed900a3c800e0d12fdfe143
Website 7

Fig. 4 Semantic graph of public keys

model. The model creates a semantic web of all the public keys stored in the database along with the public keys of systems A, D, and E. Now if the public key of ‘website4’ is queried, it will give the semantic structure of all public keys related to ‘website 4.’ This is a breach of other public keys. Ideally, every system can only view the immediate parent of the website. For example, if we consider website4 that is shared with system D, the user will only be able to view that website 4 is a fork of website2. There will be no trace of website1 to the user in system D but with this method we can trace it to the root website as shown in Fig. 4. In this figure, you can see the tree of websites related to one another. The periodical updating of the root key performed [12] by IPFS would help us reduce the level of risk.

The idea is to explain the possibility of URL exposure even if the data from two of the systems were not used. The success rate of the model will be directly proportional to the number of system drive data uploaded. The more system drives are uploaded the better information the semantic graph will provide. The forked website information, therefore, limits the privacy of public keys. This approach will not be successful if an independent copy of the website is created as the system drive doesn't store any link to the copied website. This claims that the network privacy of Beaker Browser can be maintained only if an independent copy of the website is created instead of the forked copy.

8 Mitigation Approach

IPFS(InterPlanetary File System) is similar to the Dat protocol which is used by Beaker Browser. In IPFS, we can create a private IPFS by using swarm keys. While accessing the public key of the data, the IPFS node checks for the swarm key. Only the nodes with the swarm key will be able to view the content. A similar mechanism can be implemented in Beaker to prevent users from accessing private websites. With swarm key authentication, even if the public key is exposed, the website content will be limited to the peers within the closed network.

9 Conclusion and Future Work

Traditional methods of web browsing have advanced user friendly features and provide mitigations for many risks, whereas new age browsers are still experimental. In conclusion, we state that the model could be easily used by adversaries to create a repository of public keys. The existing features of Dat protocol are not competent to prevent outside users from accessing a website. Although Beaker Browser has some amazing client-centric features, there is no central authority responsible for the security of information.Therefore, the client model must have efficient properties to preserve the privacy of websites. It is still experimental and needs a lot of improvement before it can replace our traditional technology. The features from existing peer-to-peer technology can be used to modify Dat protocol into a more secure version. As future work, we plan to perform an active attack on Beaker Browser to attain the system drive data. This will give more insight into how the storing of public keys in a plaintext format is not the best way of storing information on the browser.

References

1. L.S. Sankar, M. Sindhu, M. Sethumadhavan, Survey of consensus protocols on blockchain applications, in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)* (IEEE, 2017)
2. M. Ogden, K. McKelvey, M.B. Madsen, Dat-distributed dataset synchronization and versioning. Open Sci. Framework **10** (2017)

3. C. Hartgerink, Verified, shared, modular, and provenance based research communication with the dat protocol. Publications **7**(2), 40 (2019)
4. K. Jannes, B. Lagaisse, W. Joosen. The web browser as distributed application server: towards decentralized web applications in the edge, in *Proceedings of the 2nd International Workshop on Edge Systems, Analytics and Networking* (2019)
5. A.R. Naik, B.N. Keshavamurthy, Next level peer-to-peer overlay networks under high churns: a survey. Peer-to-Peer Network. Appl. **13**(3), 905–931 (2020)
6. J. Artback, Comparison of video file transmission: over dat protocol and hypertext transfer protocol (2019)
7. M. Kripa, et al., Blockchain framework for social media DRM based on secret sharing, in *International Conference on Information and Communication Technology for Intelligent Systems*. (Springer, Singapore, 2020)
8. P. Karunakaran, Deep learning approach to DGA classification for effective cyber security. J. Ubiquitous Comput. Commun. Technol. (UCCT) **2**(04), 203–213 (2020)
9. P.S. Sangeerth, K.V. Lakshmy, Blockchain based smart contracts in automation of shipping ports, in *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (IEEE, 2021)
10. https://datprotocol.github.io/how-dat-works/
11. https://beakerbrowser.com/
12. V. Suma, W. Haoxiang, Optimal key handover management for enhancing security in mobile network. J. Trends Comput. Sci. Smart Technol. (TCSST) **2**(04), 181–187 (2020)

Self-Organizing Deep Learning Model for Network Traffic Classification

U. Prabu and V. Geetha

Abstract Network traffic classification is one of the active research streams in the field of network management. Network traffic classification plays a crucial role in providing quality of service (QoS), management, forecasting future trends, and detecting potential security threats. A network operator must provide services based on QoS for each application in a network. Incorporating intelligence in network traffic systems plays a significant role in providing QoS. Deep learning prevents the need to select features by a domain expert as it automatically selects features through training. On concerning the above things, a self-organizing deep learning model for network traffic classification is proposed. In the proposed system, the features are extracted using deep learning models. The extracted features are then fed into the proposed self-organizing model. The self-organizing model is trained with the labelled data for clustering which in turn will cluster the unlabelled data. After further refinement, the cluster tends to update its centroid in the optimal position. Then, the process of dimensionality reduction takes place where the dimensions of the features are reduced based on the filter and wrapper methods. Finally, the performance of these methods is given to the classification algorithms which classify the features. The work would be further extended by developing a feature extraction model using deep learning which is expected to provide an efficient traffic classification while comparing with the state-of-the-art network traffic classification models.

Keywords Network traffic · Feature identification · Deep learning models · Classification algorithms · Dimensionality reduction · Self-organizing model

U. Prabu (✉)
Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India

V. Geetha
Department of Information Technology, Pondicherry Engineering College, Puducherry, India
e-mail: vgeetha@pec.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_33

1 Introduction

In contemporary networks, network traffic classification (TC) has become one of the essential tasks. It is important to identify various kinds of applications making use of network resources for the proper management [1]. Subsequently, for the advanced management tasks such as quality of service (QoS) and anomaly detection, TC is considered as one of the major prerequisites. Even, TC has gained attention in academia, research, and industry [2–4]. TC techniques classify the network traffic in accordance with the network protocols present in application layers which brought the attention of many researchers towards it [5].

Traditionally, network traffic classification has been done using port information of protocol and specification of it [6–8]. But, due to the advancement of new applications, traditional techniques are unable to classify the traffic as there in no documented and fixed format for port number [9, 10]. In this circumstance, machine learning techniques are used by the researchers. Through these techniques, specific patterns of protocols are mined and used for creating a classifier. Numerous supervised and unsupervised techniques are used in the recent years that have achieved higher accuracy [11].

Also, deep learning techniques are analysed recently since its performance is better than other machine learning techniques [12, 13]. The current deep learning-based TC techniques classifies the traffic based on the packet payloads and header information as learning features. But, these techniques are applicable for unencrypted traffic and have a greater computational cost. As a result, the techniques based on flow statistics are introduced. These techniques depend on time series and statistical features. It enables to handle both encrypted as well as unencrypted traffic.

The following are the objectives of the research:

1. To develop a self-organizing model for data clustering using supervised learning method.
2. To design a swarm-based dynamic dimensionality reduction model with consistency perspective.
3. To develop a feature extraction model using deep learning method to improvise the accuracy in classification schema.

2 Literature Survey

Shu et al. [14] proposed a deep learning-based novel method for network traffic classification. Naive Bayesian, random forest, and deep learning neural network algorithm are used for classification. The values from the real-time network are considered as the dataset. The network flow attributes such as WWW, mail, FTP data, and P2P are taken as features. Accuracy is considered as the metric. In this work, comparison has been done between random forest, naïve Bayesian, and deep learning algorithms. The proposed method has greater accuracy and improved stability.

Millar et al. [15] proposed a deep neural network classifier to achieve higher accuracy for application protocols. UNSW-NB15 is taken as dataset. Application protocols are considered as the feature for traffic classification. Precision, recall, and degrees of freedom are taken as the metrics. Feedforward reverse-propagation neural networks are used to analyse performance of deep learning. Traffic classification has been done, and malicious classes have been found. The result implies that the proposed classifier provides an effective classification of network traffic and detection of malicious traffic.

Li et al. [16] proposed a novel byte segment neural network (BSNN). Recurrent neural network is used for network traffic classification. The network datagrams are given as input to the BSNN, and the classification results are obtained directly. Datagrams are considered as network traffic feature. STMP, DNS, Cloudmusic, PPLive, QQ, BitTorrent, Amazon, Yahoom, 360, and Foxmail are taken as dataset for traffic processing. Precision, recall, F1-measure, and accuracy are considered as metrics. BSNN maps the datagram to application protocol. The results show that it is an efficient classifier which supports the traditional and complex protocols.

Lim et al. [17] proposed a packet-based technique for network traffic classification. The payload data are considered as network traffic feature. Labelled packet capture traces (PCAP) provided by UPC's Broadband Communication Research Group are taken as dataset. Precision, recall, and F1-score are considered as metrics. CNN extracts characteristics of data and grasps patterns of features. Meanwhile, residual network architecture performs better than CNN. The traffic classification is done with higher accuracy, and better QoS for network is also provided.

Lotfollahi et al. [18] proposed a novel scheme called deep packet. It manages traffic characterization as well as application identification. Packets are considered as network traffic feature. ISCX VPN—non-VPN—is taken as dataset. Precision, recall, and F1-score are considered as metrics. Convolutional neural network is used for traffic characterization, and stacked auto-encoder neural network is used for application identification. The results depict that deep packet outperforms the similar works.

Shi et al. [19] proposed an efficient feature optimization approach (EFOA). It is based on feature selection (FS) and deep learning (DL). It produces the robust and discriminative features and searches the optimal features. Real-world traffic traces (Cambridge and UNIBS) are taken as dataset. From the dataset, flow classes and applications are mapped and considered as features. WWW and P2P are the flow classes. BitTorrent and Skype are the applications. Byte overall ccuracy, flow overall accuracy, flow g-mean, byte g-mean, and F-measure are considered as metrics. The result shows that irrelevant features are removed, robust and discriminative features are generated, and optimal features are selected. The proposed approach has better runtime performance.

Lyu et al. [20] proposed a method to classify the traffic of media. Multilayer perceptron (MLP) and convolutional neural network (CNN)-based traffic classification methods are designed to accurately classify the target traffic (audio, picture, text, and video) into different categories. A TC system is designed that uses both the packet and flow level features to improve the classification performance. Packet level as well as flow level features are considered. The values from the real-time network are taken as dataset. Precision, recall, and F1-score are considered as metrics. Audio, video, text, and images are classified efficiently.

Wang [21] proposed a method based on deep learning and neural networks. The automatic feature learning is accomplished through feature extraction and feature selection. Feature extraction is achieved through artificial neural networks (ANN) where labels are necessary and stacked auto-encoder (SAE) when labels are not necessary. Feature selection is accomplished through ANN which finds discriminative features and auto-encoder which finds something in common in one category. Protocol classification has been done through either ANN or SAE.

3 Proposed Model

As there are many concerns in the network traffic classification, a self-organizing deep learning model for network traffic classification is proposed in this article.

The research is planned to carry out in three phases as such:

Phase 1: Self-organizing model.

Phase 2: Dynamic dimensionality reduction.

Phase 3: Feature identification using deep learning model.

3.1 Phase 1: Self-organizing Model

The self-organizing model makes use of supervised and unsupervised learning algorithms. In supervised learning, the labelled traffic data is trained to map the unlabelled data. In unsupervised learning, the unlabelled data get clustered. After the creation of cluster, a cluster head is found based on the centroid value. Then, the process of self-organization is carried out.

3.2 Phase 2: Dynamic Dimensionality Reduction

In this phase, the feature identification has been done. The dimensions of the features are reduced based on the filter and wrapper methods. After that, the performance is analysed. Then, it is given as input to the various classification algorithms to classify the features. If the classification accuracy is greater than the threshold value (Θ), the classified features are checked with the performance metrics. Otherwise, the data are sent back to the self-organizing model for further optimization and self-organization of the clusters.

3.3 Phase 3: Feature Extraction Using Deep Learning Model

The traffic data obtained as input are either labelled or unlabelled. The feature extraction is carried by using one of the deep learning models such as deep neural network, convolutional neural network, and deep belief networks. These models are believed to classify the network traffic very efficiently. Thus, we can attain the better performance while comparing to the earlier two phases.

The proposed model for network traffic classification is shown in the Fig. 1.

4 Significance

The following are the significance of the research:

1. The features of traffic data will be efficiently clustered using the self-organizing model when compared to earlier research.
2. The dimensionality reduction model helps to reduce the dimension of features which are extracted, and it provides the clear picture based on the requirements.
3. The feature extraction model using deep learning may provide the optimal and robust features for traffic classification.
4. The model may provide good balance on the overall performance.

5 Conclusion

In today’s scenario, network management plays a vital role. Especially, traffic classification has a major concern in network management. Traditional traffic classification techniques are not up to the level in classifying the network traffic because of the changes in the port number format and recently progressing protocols with various formats. By considering the above-mentioned problems, a self-organizing model for network traffic classification has been proposed. The model is expected to provide better performance than the existing models. In future, the work will be implemented with an extended idea.

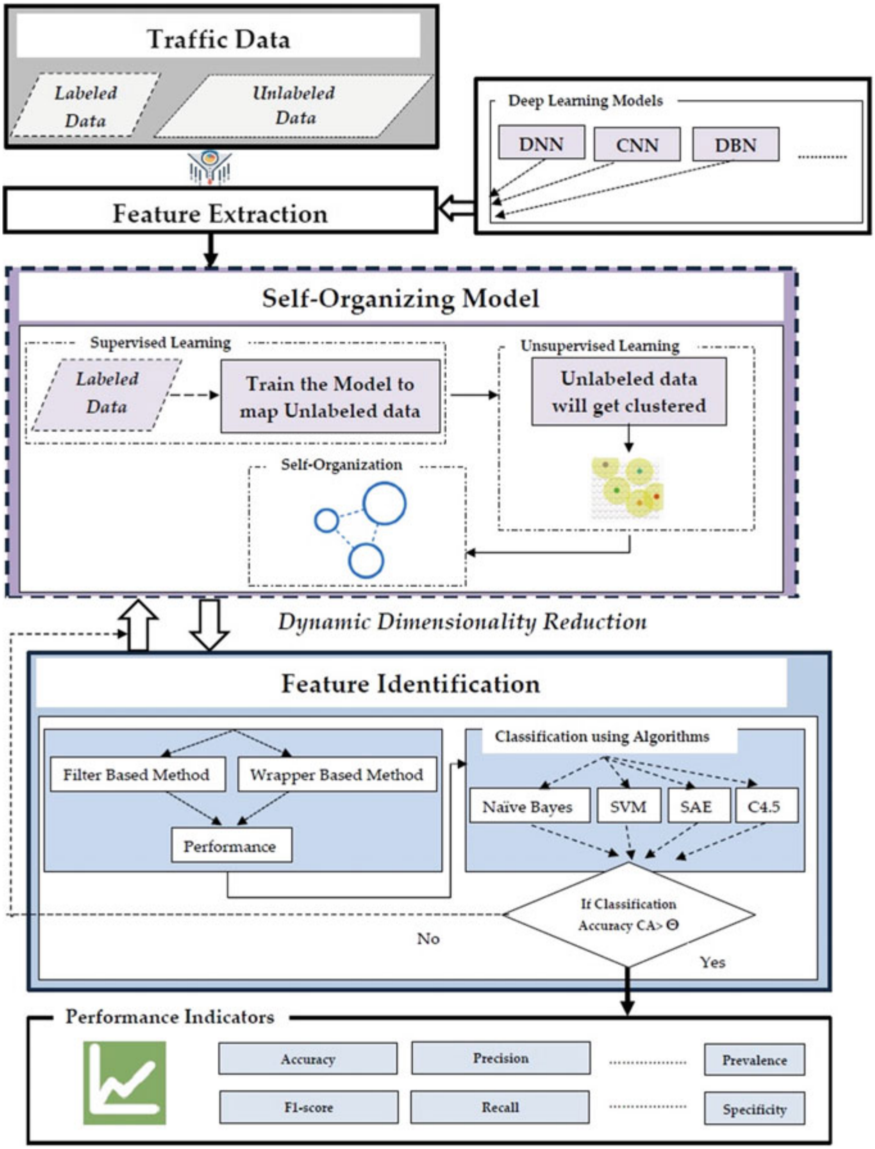


Fig. 1 Proposed model for network traffic classification

References

1. S. Bagui, X. Fang, E. Kalaimannan, S.C. Bagui, J. Sheehan, Comparison of machine-learning algorithms for classification of VPN network traffic flow using time-related features. J. Cyber Security Technol. **1**(2), 108–126 (2017)

2. A. Dainotti, A. Pescape, K.C. Claffy, Issues and future directions in traffic classification. IEEE Netw. **26**(1), 35–40 (2012)
3. M. Finsterbusch, C. Richter, E. Rocha, J.-A. Muller, K. Hanssgen, A survey of payload-based traffic classification approaches. IEEE Commun. Surv. Tutorials **16**(2), 1135–1156 (2013)
4. P. Velan, M. Čermák, P. Čeleda, M. Drašar, A survey of methods for encrypted traffic classification and analysis. Int. J. Netw. Manage. **25**(5), 355–374 (2015)
5. H. Kim, K.C. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, K.Y. Lee, Internet traffic classification demystified: myths, caveats, and the best practices, in *Proceedings of the 2008 ACM CoNEXT Conference* (2008), pp. 1–12
6. Y. Wu, G. Min, K. Li, B. Javadi, Modeling and analysis of communication networks in multicluster systems under spatio-temporal bursty traffic. IEEE Trans. Parallel Distrib. Syst. **23**(5), 902–912 (2011)
7. C.S. Sastry, S. Rawat, A.K. Pujari, V.P. Gulati, Network traffic analysis using singular value decomposition and multiscale transforms. Information Sci. **177**(23), 5275–5291 (2007)
8. T. Karagiannis, K. Papagiannaki, M. Faloutsos, BLINC: multilevel traffic classification in the dark, in *Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (2005), pp. 229–240
9. Z. Li, G. Xia, H. Gao, Y. Tang, Y. Chen, B. Liu, J. Jiang, Y. Lv, Netshield: massive semantics-based vulnerability signature matching for high-speed networks. ACM SIGCOMM Comput. Commun. Rev. **40**(4), 279–290 (2010)
10. R. Pang, V. Paxson, R. Sommer, L. Peterson, Binpac: a yacc for writing application protocol parsers, in *Proceedings of the 6th ACM SIGCOMM Conf. Internet Meas.* (2006), pp. 289–300
11. T.T.T. Nguyen, G. Armitage, A survey of techniques for internet traffic classification using machine learning. IEEE Commun. Surv. Tutorials **10**(4), 56–76 (2008)
12. W. Wang, M. Zhu, X. Zeng, X. Ye, Y. Sheng, Malware traffic classification using convolutional neural network for representation learning, in *2017 International Conference on Information Networking (ICOIN)*, IEEE, pp. 712–717 (2017)
13. M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, J. Lloret, Network traffic classifier with convolutional and recurrent neural networks for Internet of Things. IEEE Access **5**, 18042–18050 (2017)
14. J.H. Shu, J. Jiang, J.X. Sun, Network traffic classification based on deep learning. J. Phys. Conf. Ser. **1087**(6), 062021 (IOP Publishing) (2018)
15. K. Millar, D. Smit, C. Page, A. Cheng, H. Chew, C. Lim, *Looking Deeper: Using Deep Learning to Identify Internet Communications Traffic* (2017)
16. R. Li, X. Xiao, S. Ni, H. Zheng, S. Xia, Byte segment neural network for network traffic classification, in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)* (2018), pp. 1–10
17. H.-K. Lim, J.-B. Kim, J.-S. Heo, K. Kim, Y.-G. Hong, Y.-H. Han, Packet-based network traffic classification using deep learning, in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, IEEE (2019), pp. 046–051
18. M. Lotfollahi, M.J. Siavoshani, R.S.H. Zade, M. Saberian, Deep packet: a novel approach for encrypted traffic classification using deep learning. Soft Compu. **24**(3) 1999–2012 (2020)
19. H. Shi, H. Li, D. Zhang, C. Cheng, X. Cao, An efficient feature generation approach based on deep learning and feature selection techniques for traffic classification. Comput. Netw. **132**, 81–98 (2018)
20. Q. Lyu, X. Lu, Effective media traffic classification using deep learning, in *Proceedings of the 2019 3rd International Conference on Compute and Data Analysis* (2019), pp. 139–146
21. Z. Wang, The applications of deep learning on traffic identification. BlackHat USA **24**(11), 1–10 (2015)

Artificial Intelligence-Based Rubber Tapping Robot

T. S. Angel, K. Amrithesh, Karthik Krishna, Sachin Ashok, and M. Vignesh

Abstract Rubber tapping is the process of extracting latex from rubber trees. Rubber tree (*Hevea brasiliensis*) is the most prevalent plantation crop in Kerala, India, and it was discovered to be the main source of income for the over 11.5 lakh people who live there. Because of its outstanding qualities, natural rubber is often used in everyday life. Manual tapping is currently the most popular method of accessing natural rubber. Rubber tree tapping is considered a skilled labor-intensive task, and the availability of such labor is dwindling. This is considered to be one of the factors affecting natural rubber production in Indian plantations. As a solution to the problems faced by the rubber industry, an AI-based rubber tapping robot is proposed that can automatically move from one tree to another in a rubber plantation and tap each tree.

Keywords Artificial neural network (ANN) · Artificial intelligence (AI) · Rubber tapping robot · Adaptive fuzzy-PI · Speed control · DC motor · Path planning

1 Introduction

Rubber trees produce sticky, white latex, which is harvested and refined into natural rubber. Natural rubber and its derivatives are widely used in various industries, shipping, national security and medical care. Natural rubber's excellent properties can no longer be matched by synthetic materials [1, 2]. In the world, India is one of the largest producers of natural rubber.

Rubber tree tapping is the practice of extracting latex from a rubber tree by creating a controlled incision on the tree bark [3]. When a rubber tree is first tapped, it can generate latex for up to 25 years. The three layers of the rubber tree are the inner wood, the middle layer which is known as cambium, and the outer layer called the bark [3]. During the rubber tapping, take care not to let the blade in the tapering tool go deep into the bark, causing the cambium to be damaged [2].

T. S. Angel (✉) · K. Amrithesh · K. Krishna · S. Ashok (✉) · M. Vignesh
Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India
e-mail: angelts@am.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_34

Rubber tapping is a skill-based activity that necessitates the tapper's tool handling skills along with proper hand–eye coordination. Taping usually begins as the tree trunk exceeds a diameter of around 50 cm at a height of 1 m from the earth [3]. The helical pattern is scribed on the tree surface with a pointing tool, and the label is then deepened against the wood with tapping equipment. To avoid latex from spilling out of the cut, the helical cut is made at an angle toward the wood region [3]. The blade could then be used to scrape a thin layer of bark right above the previous cut, and latex begins to flow into the collecting cup into the cut. Nowadays, the major problems faced by the rubber industry are the dwindling in professional labor availability, the cost of labor has risen over time, worker's inefficiency and poor tapping abilities will cause harm to the tree [1]. Furthermore, rubber tapping workers are plagued with a slew of health issues. Backache was the most common musculoskeletal concern. Cuts were the most common health threat, followed by eye injuries, chemical injuries and snake bites.

The automation in agriculture is the main concern and the emerging subject across the world and farmers' conventional methods of production were insufficient. Automation has protected crop yields from a variety of influences such as climate change, population development, labor challenges and food security concerns. There is an exponential increase in the research and development in the field of robotics for various applications [4, 5]. The robots were developed to climb the trees and harvesting, seed sowing, rubber tapping, etc. [6–9].

Semi-automated rubber tapping machines with vibrational sensors for improved cuts need manual assistance for tapping each tree, and thus it consumes time [3]. Image processing is used to detect the tapping path which results in an automated rubber tapping process [1]. In this chapter, an artificial intelligence-based fully automated rubber tapping robot is discussed which would ease the rubber tapping process by reducing labor requirements to a minimum. Proposed rubber tapping robot uses an ANN to detect the tapping depth and length to be tapped from ground level from age and the circumference of the tree and tapping count. The movement of the proposed robot is controlled by fuzzy based adaptive-PI controller. Path planning is based on the coordinates of the rubber trees. The machine is designed to automatically mimic and generate the same pattern as a traditional technique.

The content is organized as follows. AI-based rubber tapping process is given in Sect. 2. Path planning of the robot is depicted in Sect. 3. Section 4 deals with the prediction algorithm of the rubber tapping robot. In this section, it predicts the tapping depth and length to be tapped from ground level using an ANN, based on the field data collected from various rubber plantations in Kerala, India. In Sect. 5, it deals with the comparison of a PI controller and an adaptive fuzzy-PI controller for the speed control of DC motor which aids the robot movement. Finally, Sect. 6 concludes the work.

2 AI-Based Rubber Tapping Process

The entire process flow of the AI-based rubber tapping robot is depicted in Fig. 1. As the initial step, the robot reads all the coordinates of the rubber trees in the rubber plantation. After that the robot will find the distance between the current position and the next position of rubber trees. The next step is to find the reference speed for dc motor 1 to cover the distance in a particular time, and the robot moves from the current position to the next position. Speed of the dc motor is controlled using an adaptive fuzzy-PI controller. After reading the inputs (age, circumference, tapping iteration), a trained ANN model predicts the tapping depth and length to be tapped from the ground level. The robot will find the reference speed of dc motor 2 with respect to length to be tapped from ground level that is predicted from trained artificial neural network model to cover the distance in a particular time, and it will move to desired height from the current height. Afterward, the robot will find the reference speed of dc motor 3 to complete the tapping in a particular time. Then, the tapping process of the rubber tree takes place at a predefined angle and predicted tapping depth using dc motor 3. Finally, the robot will check whether all the rubber trees are tapped or not. If all rubber trees are not tapped, then the process is repeated. Thus, combined ANN and fuzzy based intelligent technique are used for the proposed rubber tapping robot, and path planning is based on coordinates of the rubber tree.

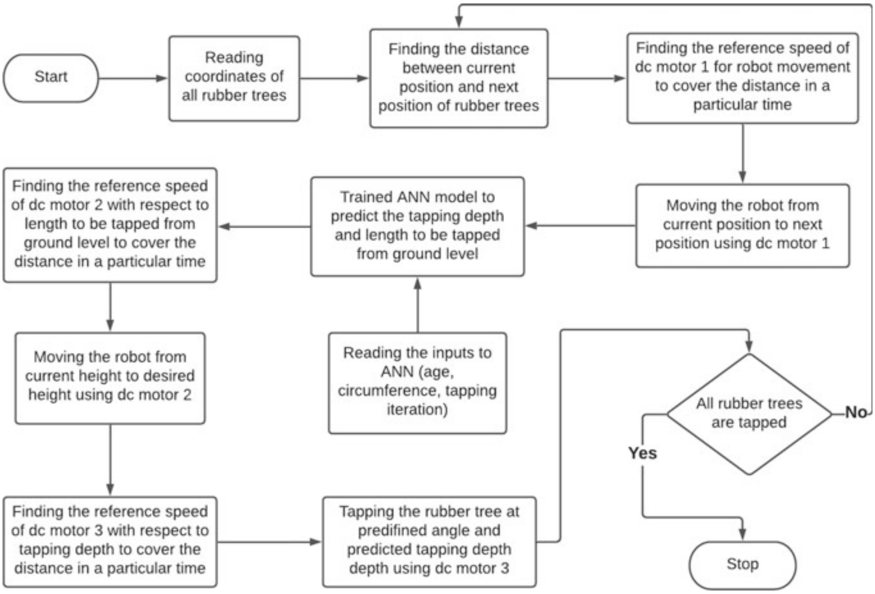

Start
Reading coordinates of all rubber trees
Finding the distance between current position and next position of rubber trees
Finding the reference speed of dc motor 1 for robot movement to cover the distance in a particular time
Moving the robot from current position to next position using dc motor 1
Trained ANN model to predict the tapping depth and length to be tapped from ground level
Reading the inputs to ANN (age, circumference, tapping iteration)
Finding the reference speed of dc motor 2 with respect to length to be tapped from ground level to cover the distance in a particular time
Moving the robot from current height to desired height using dc motor 2
Finding the reference speed of dc motor 3 with respect to tapping depth to cover the distance in a particular time
Tapping the rubber tree at predifined angle and predicted tapping depth depth using dc motor 3
All rubber trees are tapped
No
Yes
Stop

Fig. 1 Process flow of the AI-based rubber tapping

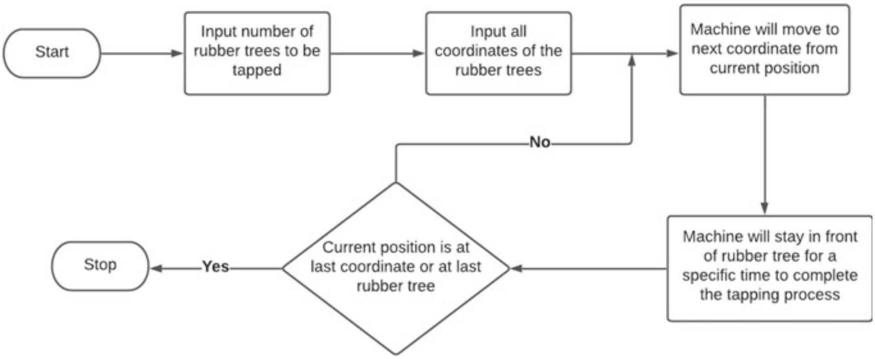

Start
Input number of rubber trees to be tapped
Input all coordinates of the rubber trees
Machine will move to next coordinate from current position
Machine will stay in front of rubber tree for a specific time to complete the tapping process
Current position is at last coordinate or at last rubber tree
No
Yes
Stop

Fig. 2 Process of path planning

3 Path Planning of the Robot

Rubber tapping robot must be able to stop in front of rubber trees to perform the tapping process, and this process needs to be continued for the entire rubber trees in the plantation. Hence, the path planning of the robot is one of the key steps in this work. It ensures that the robot goes to each rubber tree in the plantation [6]. The process of path planning is shown in Fig. 2.

The path planning algorithm is simulated in MATLAB software. For convenience, the total number of rubber trees to be tapped is given as 8 (4 rubber trees grouped in 2 columns), and the time taken for each tapping process is given as 8 s in the simulation. The rubber tapping robot is marked by a circular (O) symbol, while the rubber trees to be tapped are indicated by a star (*) mark. The rubber tapping robot will switch from one rubber tree to the next with a delay that represents the amount of time it takes to tap each rubber tree. Analogously, the rubber tapping robot will switch from one row to the next after tapping all of the rubber trees in that row. Rubber tapping robot progress through each rubber tree after each tapping process is depicted in both Figs. 3 and 4.

4 Prediction Algorithm for Cutting Depth and Length to Be Tapped from Ground Level

Cutting depth and length to be tapped from ground level is predicted using artificial neural networks. As a first step, the field data from the rubber plantations were collected such as the circumference of the rubber tree, age in years, tapping thickness and length to be tapped from ground level. A back propagation network is trained using age and circumference as inputs and tapping thickness and length to be tapped from ground level as outputs, respectively.

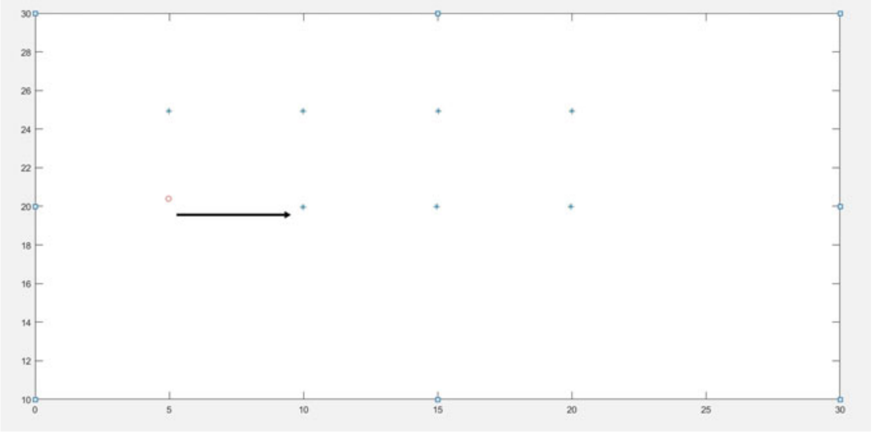


Fig. 3 Robot's progress through each rubber tree

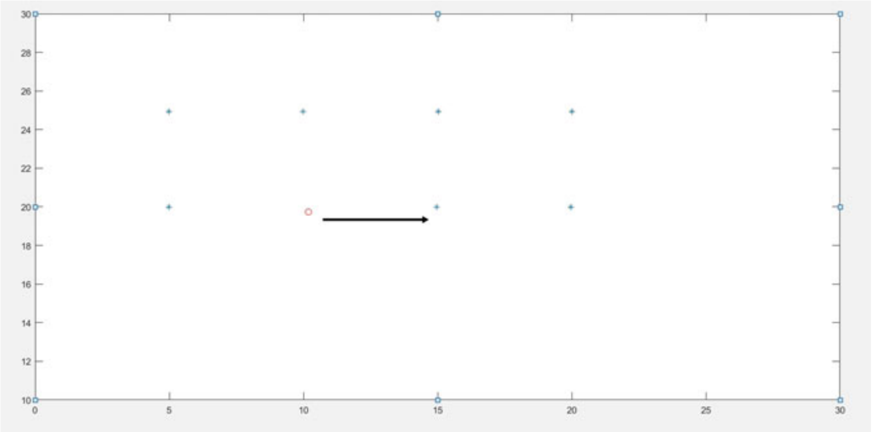


Fig. 4 Robot's progress through each rubber tree

ANN is modeled in MATLAB and verified the prediction. From the field data, 85% is used for training, 5% for validation and 10% for testing purposes. Levenberg-Marquardt algorithm is used for training the ANN. The regression plot of the ANN training, validation, and testing is given in Fig. 5. The regression plot shows that the training, testing, validation, and overall accuracy of the model are good and acceptable. To check the adaptability of the trained ANN model, it is tested using unseen data. ANN model with testing results is shown in Fig. 6. The accuracy of the predictions is computed. Analysis of the data obtained from the trained model with that of the collected field data is shown in Table 1. Error percentage of the tapping depth prediction by ANN is given in Table 1 and the maximum error percentage is 3.75. It is clear that the ANN predicts the tapping depth with a tolerance of ±4%.

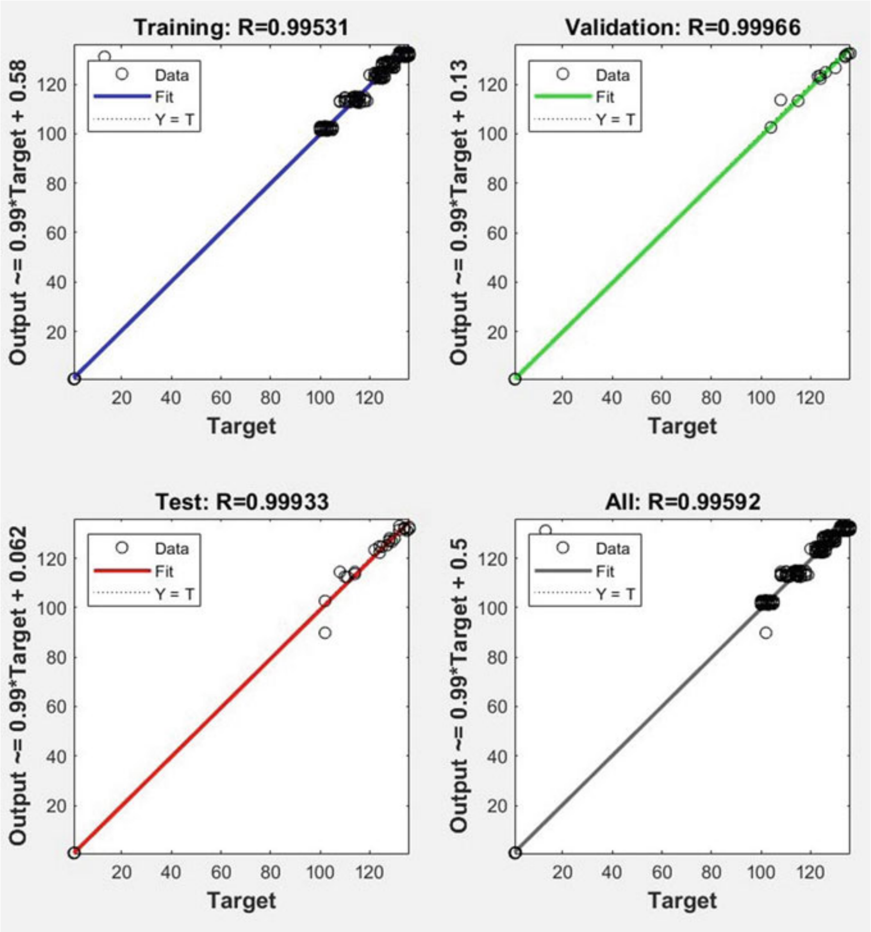


Fig. 5 Regression plot of ANN

Similarly, the error percentage of prediction of the length to be tapped from ground level by ANN is given in Table 1, and the maximum error percentage is 3.69. The ANN predicts the length to be tapped from ground level with a tolerance of ±4%. So the model is an efficient one that predicts the output which is similar to the actual output. This model predicts the length to be tapped from ground level and taping depth when inputted age, circumference and tapping iteration as shown in Fig. 6.

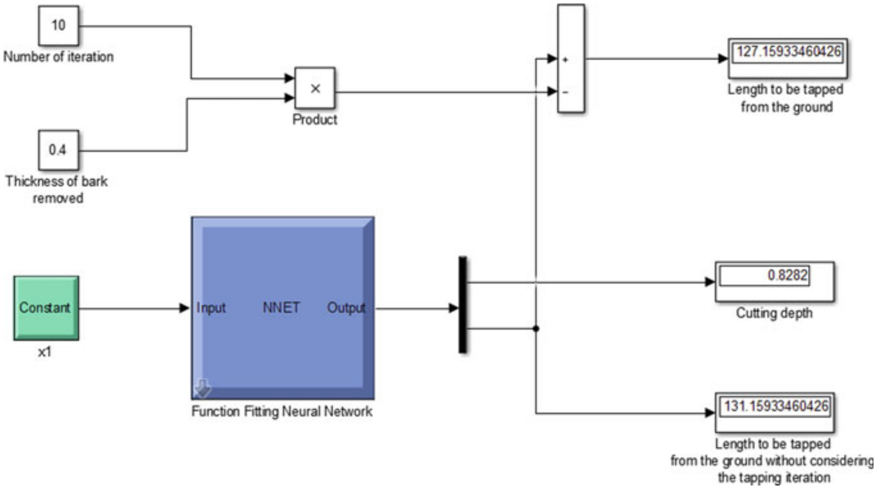

10
Number of iteration
0.4
Thickness of bark removed
×
Product
+
-
127.15933460426
Length to be tapped from the ground
Constant
x1
Input
NNET
Output
Function Fitting Neural Network
0.8282
Cutting depth
131.15933460426
Length to be tapped from the ground without considering the tapping iteration

Fig. 6 Testing of the ANN model

Table 1 Error percentage of the prediction by ANN

Age in years	Circumference in cm	Tapping depth (from field data) in cm	Tapping depth from ANN model in cm	Error % of tapping depth prediction	Length to be tapped from ground (from field data)	Length to be tapped from ground level from ANN model in cm	Error % of length to be tapped from ground level prediction
8	84.7	0.7	0.7	0	103	101.7	1.26
8	86.2	0.8	0.77	3.75	127	123.8	2.51
8	83.4	0.8	0.79	1.25	121	124.4	2.80
7	79.4	0.8	0.79	1.25	115	114.9	0.08
9	87.4	0.7	0.69	1.42	129	129.2	0.15
8	80.9	0.7	0.69	1.42	103	101.4	1.55
9	88.2	0.7	0.69	1.42	128	129.3	1.01
7	78.9	0.9	0.88	2.22	110	113.6	3.27
10	89.6	0.8	0.79	1.25	138	132.9	3.69
10	91.3	0.8	0.79	1.25	129	132.6	2.79

5 Movement Control of Robot Using PI and Adaptive Fuzzy-PI Controllers

Block diagram representation of the movement control of robot is given in Fig. 7. A DC motor can be used for the movement of the robot from one location to another. After reading the location coordinates of trees, the distance between them is calculated, and the reference speed for the motor is determined to cover the pre-calculated distance in a specified time. A timer is used for the operation of the motor and a speed control mechanism to ensure the reach of the robot at the destination accurately. Distance between the rubber trees is obtained using (1). Reference speed to cover the calculated distance in a specified time is obtained using (2).

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

In (1), (x_1,y_1) represents the coordinates of the first rubber tree, and (x_2,y_2) represents the coordinates of the second rubber tree.

$$N = \frac{60D}{\pi dt} \tag{2}$$

In (2), D represents the distance between two rubber trees, t represents the time in seconds needed to cover the distance, and d represents the wheel's diameter coupled with a DC motor.

Block diagram representation of the dc motor used for the analysis is shown in Fig. 8. In Fig. 8, Ra is armature resistance, L_a is armature inductance, J is rotational inertia, B is a viscous function, K_m is motor constant, K_r converts rad/s into rpm, and K_b is back emf constant, respectively [10]. Two different controllers are studied for the speed control of the dc motor, the PI controller and an adaptive fuzzy-PI controller [11, 12]. In an adaptive fuzzy-PI controller, the proportional and integral constant of the PI controller are not constant rather varying. Normalized error (e) and change in error (ec) of the speed are the input to the fuzzy Mamdani inference

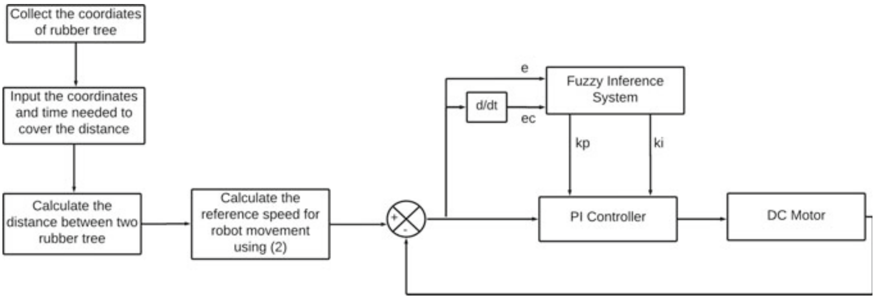

Collect the coordiates of rubber tree
Input the coordinates and time needed to cover the distance
Calculate the distance between two rubber tree
Calculate the reference speed for robot movement using (2)
+
-
e
d/dt
ec
Fuzzy Inference System
kp
ki
PI Controller
DC Motor

Fig. 7 Block diagram representation of the movement control of robot

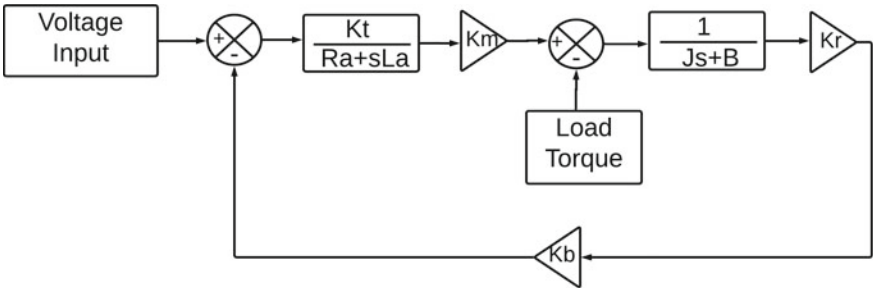


Fig. 8 Block diagram of DC motor

system and outputs are the weighting parameters for the proportional and integral constant of the PI controller.

Let K_{p} and K_{i} are the proportional and integral constants of the conventional PI controller then the proportional and integral constant of the adaptive fuzzy-PI controller are given by (3) and (4), respectively.

$$K_{\mathrm{p\_ fuzzyPI}} = \mathrm{WK_p} \times K_{\mathrm{p}} \tag{3}$$

$$K_{\mathrm{i\_ fuzzyPI}} = \mathrm{WK_i} \times K_{\mathrm{i}} \tag{4}$$

In (3), $\mathrm{WK_p}$ is the weighting parameter of proportional constant and in (4), $\mathrm{WK_i}$ is the weighting parameter of integral constant, respectively, and that are obtained from the fuzzy inference system. The membership functions of the input variables and the output variables are given in Figs. 9 and 10, respectively. Seven linguistic variables are used for defining the fuzzy sets of each variable and that are NB, NM, NS, Z, PS, PM and PB, respectively [10, 12]. The rule base for the fuzzy inference system relating the input variables and each of the output variables is given in Fig. 11.

Response of the dc motor for extreme control conditions is given in Figures 12 and 13, respectively. The peak overshoot with conventional PI controller is considerably reduced with the adaptive fuzzy-PI controller. The settling time of both the controllers is comparable and they attain the steady-state at 0.5 s. Depending upon the distance between the rubber trees the motor might be operated in the speed range of 50–500 rpm and the adaptive fuzzy-PI controller gives a steady performance in this

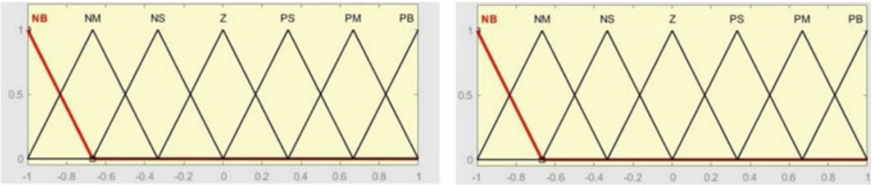


Fig. 9 Membership function of **a** error(e) and **b** change in error(ec)

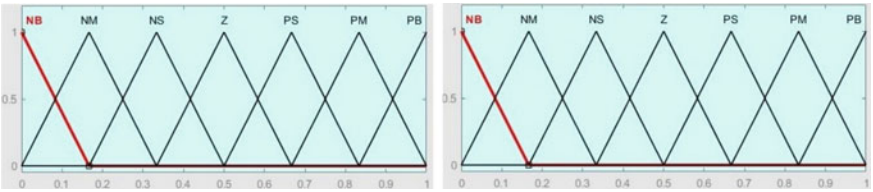
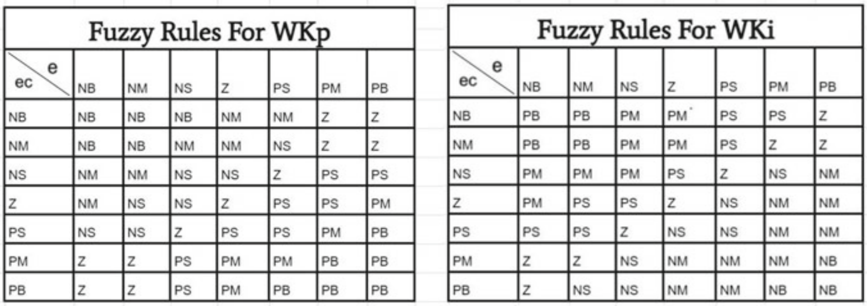


Fig. 10 Membership function of **a** WK_i and **b** WK_p



Fuzzy Rules For WKp

ec \ e	NB	NM	NS	Z	PS	PM	PB
NB	NB	NB	NB	NM	NM	Z	Z
NM	NB	NB	NM	NM	NS	Z	Z
NS	NM	NM	NS	NS	Z	PS	PS
Z	NM	NS	NS	Z	PS	PS	PM
PS	NS	NS	Z	PS	PS	PM	PB
PM	Z	Z	PS	PM	PM	PB	PB
PB	Z	Z	PS	PM	PB	PB	PB

Fuzzy Rules For WKi

ec \ e	NB	NM	NS	Z	PS	PM	PB
NB	PB	PB	PM	PM	PS	PS	Z
NM	PB	PB	PM	PM	PS	Z	Z
NS	PM	PM	PM	PS	Z	NS	NM
Z	PM	PS	PS	Z	NS	NM	NM
PS	PS	PS	Z	NS	NS	NM	NM
PM	Z	Z	NS	NM	NM	NM	NB
PB	Z	NS	NS	NM	NM	NB	NB

Fig. 11 Fuzzy rule base for **a** WK_i and **b** WK_p

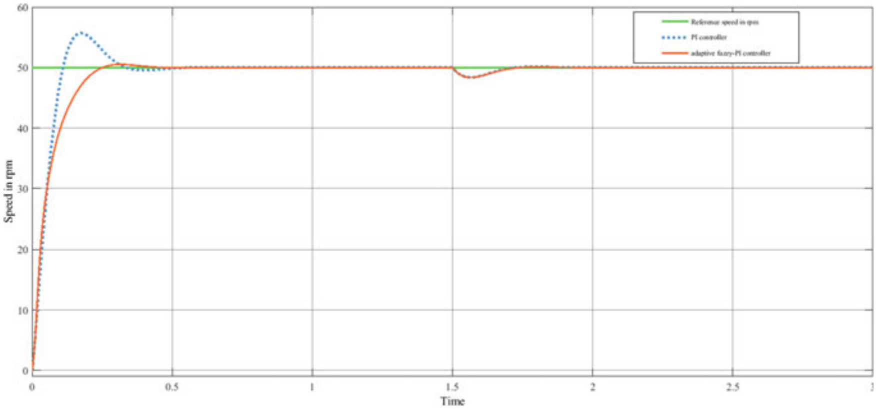


Fig. 12 Response of the dc motor for a reference speed of 50 rpm

wide range of speed control. Effect of load torque variations is considered in the simulation by incorporating a load torque variation from 20 to 50 Nm at 1.5 s, and as observed in Figs. 12 and 13 the controller response is faster, and steady-state is attained accurately.

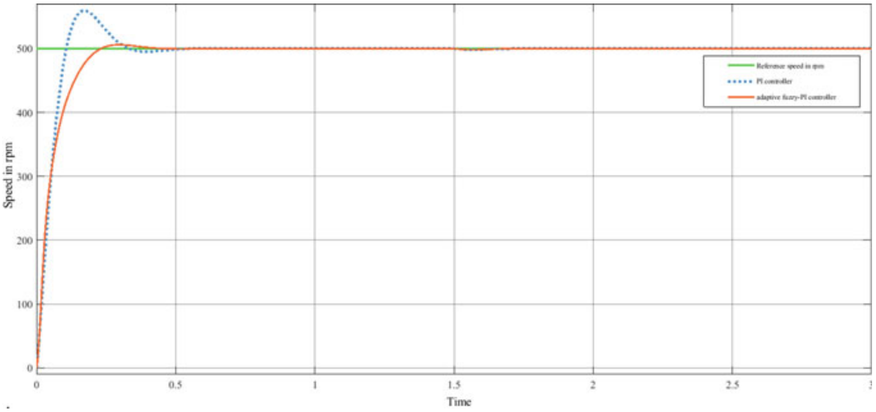


Fig. 13 Response of the dc motor for a reference speed of 500 rpm

6 Conclusion

In this chapter, simulation of an AI-based rubber tapping robot is discussed. The proposed AI-based scheme is adaptive in nature, suitable for any plantation since the tapping depth and height is predicted using an ANN which is trained using field data, and thus, it exhibits generalization capability. From the above discussions, it is clear that the ANN predicts the tapping depth and length to be tapped from ground level accurately with a tolerance of ±4%. Movement control of the robot was analyzed using a conventional PI controller and an adaptive fuzzy-PI controller. It is observed that the peak overshoot is considerably reduced with the adaptive fuzzy-PI controller, and the performance is stable in a wide range of speed control and under varying load conditions. Thus, the combined ANN and fuzzy-based intelligent rubber tapping robot make the rubber tapping process much easier and faster with less human effort. Thus, it is an effective option to replace the problems faced by the rubber industry. Study was limited to the simulation of the intelligent algorithms for path planning, tapping depth prediction and movement control. Mechanical design of the robot and the cutting tool path planning is not considered in this study.

Appendix

Resistance of armature: 0.5 Ω	Torque constant: 1.6 Nm/A
Inductance of armature: 1.0 mH	Back emf constant: 1.6 Vs/rad
Rotational inertia: 5 kg/m^2	Proportional constant, K_{p}:2
Viscous friction constant: 0.01 Nm/rad/s	Integral constant, K_{i}: 50

References

1. Y.A.I. Yatawara, W.H.C. Brito, M.S.S. Perera, D.N. Balasuriya, *Appuhamy*—The Fully Automatic Rubber Tapping Machine, vol. LII, no. 02 (2019), pp. 27–33
2. S. Zhang, C. Zhang, J. Zhang, T. Yuan, W. Li, D. Wang, F. Zhang, Design and experiment of suspension-typed rubber tapping device. Int. Agricultural Eng. J. **27**(4) (2018)
3. R. Nair Arjun, S.J. Soumya, R.S. Vishnu, R.R. Bhavani, Semi automatic rubber tree tapping machine. in *International Conference on Robotics and Automation for Humanitarian Applications (RAHA)* (2016)
4. A. Antony, P. Sivraj, Food delivery automation in restaurants using collaborative robotics, in *Proceedings of the IEEE International Conference on Inventive Research in Computing Applications, ICIRCA-2018* (2019), pp. 111–117
5. G. Ranganathan, An economical robotic arm playing chess using visual servoing. J. Innov. Image Process. (JIIP) **2**(03), 141–146 (2020)
6. R.K. Megalingam, G.V. Vivek, S. Bandyopadhyay, M.J. Rahi, Robotic arm design, development and control for agriculture applications, in *International Conference on Advanced Computing and Communication Systems (ICACCS -2017)*, 06–07 Jan 2017
7. T. Narayanan, R.S. Vishnu, R.R. Bhavani, V.A. Vashista, Cable driven parallel robot for coconut farm, in *International Conference on Advances in Computing, Communications and Informatics-ICACCI 2017*, pp. 864–870, January-2017
8. P.V.S. Jayakrishna, M.S. Reddy, N.J. Sai, N. Susheel, K.P. Peeyush, Autonomous seed sowing agricultural robot, in *International Conference on Advances in Computing, Communications and Informatics, ICACCI-2018*, art. no.-8554622, pp. 2332–2336, December 2018
9. R.K. Megalingam, S.K. Manoharan, S.M. Mohandas, S.R.R. Vadivel, R. Gangireddy, S. Ghanta, K.S. Kumar, P.S. Teja, V. Sivanantham, Amaran: an unmanned robotic coconut tree climber and harvester. IEEE/ASME Trans. Mechatron. **26**(1), 288–299 (2020)
10. Y.A. Almatheel, A. Abdelrahman, Speed control of DC motor using fuzzy logic controller, in *Proceedings of International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, Khartoum, Sudan (2017)
11. S.M. Sam, T.S. Angel, Performance optimization of PID controllers using fuzzy logic, in *Proceedings of IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, ICSTM 2017*, pp. 438–442, October 2017
12. Kartik Sharma, Dheeraj Kumar Palwalia, "A modified PID control with adaptive fuzzy controller applied to DC motor", in proc. International conference on information, communication, instrumentation and control,ICICIC 2017.

Unmanned Aerial Vehicles (UAV) Jammer

Fat'hi Salim Said AL-Ghafri and Lavanya Vidhya

Abstract Unmanned aerial vehicles, or industrial drones as they are more widely known, have sparked questions about confidentiality of private assets. Drones are machines that fly without an operator, and they are controlled and monitored from the platform by an engineer. Drones come in a wide range of designs, configurations, sizes, and features. Drones can be used for a number of purposes, including video recording, spying, firing weapons, and dropping bombs. Due to the coordinator of communication in the Sultanate of Oman's Telecommunication Regulatory Authority's (TRA) protective regulation of adopting drones and jammers, there have been no programs or discoveries regarding drones or jammers in the Sultanate of Oman. As a result of the lack of organized regulations for the use of drones in Oman, several drones have been illegally imported into the country over the last 3 years. TRA was aware that Oman's drone legislation will be released shortly, posing a significant challenge in determining how to secure Oman's diverse and delicate organizations from unauthorized drone use in the absence of jammer authorizations. This framework reinforces a jammer system that complies with the TRA specification of disrupting drone band frequencies of 2.4 and 5.8 GHz, with GPS as an alternative. The jammer's suggested range is 500 m. This paper gave a simple representation of three main configurations of signal production analysis for jamming signals. Since jammers that block (2.4 and 5.8 GHz) frequencies are unlawful in the Sultanate of Oman, this model is intended for academic purposes and is not for business or private use. In order to develop and test a jammer system for this framework, a TRA permission letter was received.

Keywords Jammer · Signals · Drone · TRA · Communication · Regulations · Legislation · 2.4 GHz · 5.8 GHz · Frequencies

F. S. S. AL-Ghafri (✉)
Middle East College, Muscat, Sultanate of Oman

L. Vidhya
Department of E.C.E, Middle East College, Muscat, Sultanate of Oman
e-mail: vidhya@mec.edu.om

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_35

1 Introduction

UAVs have progressed from military to customer status, and they are now commonly utilized in activities like photographs from an aircraft and drone shooting, as well as in enterprises like cinematography. Numerous researchers are investigating their potential applications in a range of fields. For example, using UAVs to install tiny base station cells are proving to be a successful approach for delivering wireless networks to ground consumers in a number of situations. Among 2016 and 2020, Goldman Sachs defense and aerospace scientific analysts predict a $100 billion potential market for drones [2]. The increased accessibility of buyer, UAVs has created major surveillance and security problems. Drones should be used with caution in places like nuclear power stations, jails, airports, national boundaries, and armed-forces regions, where they could lead to incidents or be exploited for unlawful reasons. Regulated UAVs, on the other hand, such as those used to provide wireless services to base users, can be hijacked and must, therefore, be operated safely.

Unmanned aerial vehicles (UAVs) have varying level of control. They either are controlled remotely or include a predetermined flight path that they obey using global navigation satellite system (GNSS) signals and associated devices. In Oman, there are currently no laws or rules regarding the use of drones, but until the Public Authority for Civil Aviation publishes such regulations [2], TRA must include strategies to secure vulnerable companies, such as enabling the use of jammers in such locations, since without these strategies, delicate enterprises can be threatened at any time. The use of these types of jammer systems to preserve confidentiality is essential for a variety of reasons, including:

- A lack of detailed legislation governing the application of unmanned aircraft systems, as well as their strict implementation.
- There is a scarcity of appropriate education or training for users of unmanned aircraft systems, as well as their privileges and obligations.
- Current unmanned aircraft systems sensors and avionics devices have limited functionality in the Oman sector.
- There is a scarcity of primary facilities for mapping and preventing unmanned aerial systems.

The 2.4 and 5.8 GHz bands are intended for Zigbee, wireless fidelity, Bluetooth, sound / visual transmissions, and certain remote controls, but this spectrum is not widely utilized in drone/UAV remote-control systems. As a result, the system introduced in this project would be beneficial in reducing the impact on the performance of other communications networks in this band range. This concern restricts the forms of jamming approaches that can be utilized to jam drones, not only in terms of effectiveness, but also in terms of the impact on certain communication networks caused by the jamming algorithm. In this paper, we modeled and demonstrated results for the band ranges 2.4 and 5.8 GHz [2], which are widely used in modern drones and unmanned aerial vehicles and are often regarded as critical factors or parameters. The device obstructs communication between the operator (RC) and the testing drone.

The noise signal interferes with communication, causing the testing drone to land in the similar place when the battery is depleted. The jamming distance, which is defined as 500 m, and the power output of the jammer system, which is determined by the geographic area aimed for blocking, are both important parameters in the model.

2 Related Work

There are many trading systems for tracking and jamming wirelessly piloted unmanned aerial vehicles on the market. However, there is not a lot of literature about how to neutralize UAV remote-control systems. To our knowledge, no works on concord-aware jamming of unmanned aerial vehicles remote-control systems have been published in the open literature.

The efficacy of available commercially low-cost jammers against unmanned aerial vehicles has been investigated, with the conclusion that GNSS signal transmission can be jammed from a reasonable range (a few 100 m away from the UAV) [4]. However, even though the blocker is much nearer to the unmanned aerial vehicles than the operator with the wireless controller, the research concluded that jamming of the remotely controlled signals using the deemed jammers is ineffective.

A responsive tracking and jamming platform for intervening with WiMAX and Wi-Fi networks have also been developed using a different approach. The mechanism scans the range for focused signals and performs jamming specifically if one is found [4]. The digital signal processing (DSP) techniques were applied in a field-programmable gate array (FPGA) in another framework based on an SDR platform. The method, on the other hand, is designed for UAVs and includes a flexible jammer that is utilized to compare the efficacy and efficiency of concord-aware jamming to various jamming approaches.

A method for developing and implementing a dual-band global mobile communication jammer that works at both GSM (900 and 1800) frequencies was also suggested [1]. The developers suggested a range of 0–5 m for the handheld smartphone jammer, which is cranking three kinds of GSM 900 mobile devices in India: Aircel, Tata Docomo, Vodafone, or Reliance, which operate at 935–960 MHz.

3 Radio Frequency Radiations

Radio frequency is the lowest frequency of the electromagnetic radiation, and it is used in wireless networks (digital and analog). It has a frequency range of 3–300 kHz. Military radio, mobile networks, aircraft navigation, analog radio, television broadcasting, amateur radio, and satellite systems are examples of broadcasting systems that run in the RF range [4]. The International Telecommunications Union (ITU) has

Table 1 Classifications of frequency bands

F	λ	Band	Description
30–300 Hz	10^4–10^3 km	ELF	Extremely low frequency
300–3000 Hz	10^3–10^2 km	VF	Voice frequency
3–30 kHz	100–10 km	VLF	Very low frequency
30–300 kHz	10–1 km	LF	Low frequency
0.3–3 MHz	1–0.1 km	MF	Medium frequency
3–30 MHz	100–10 m	HF	High frequency
30–300 MHz	10–1 m	VHF	Very high frequency
300–3000 MHz	100–10 cm	UHF	Ultra-high frequency
3–30 GHz	10–1 cm	SHF	Super high frequency
30–300 GHz	10–1 mm	EHF	Extremely high frequency

classified the radio frequencies into the following groups and areas of usage (Table 1).

4 Forms of Jamming

The proposed technique is based primarily on signal jamming, which prevents drone remote-control signals from communicating and being hijacked. Drone jamming aims to generate sufficient interference in the drone's RF transmitter to prevent it from responding to remotely controlled commands (RC). Based on the signals being sent, there are three forms of jamming:

- **Reflection jamming:**

Non-emitting machines that mirror back signals to generate false aim indicators are referred to as reflection jamming, which is also known as mechanical jamming. Chaff, corner reflectors, chaff rope, and decoys are the most basic types of reflection jamming.

- **Electronic Jamming:**

It is a form of electronic dispute in which jammers intervene with data transmission signals directed at an opponent's radar system by blocking the receiver with strong coordinated power signals. There are two main classifications of technique (repeater and noise technique). There are three kinds of noise jamming: spot, barrage, and sweep jamming.

- **Inadvertent Jamming:**

In certain situations, cordial sources may trigger mechanical and electronic jamming. This form of jamming could be avoided with careful preparation and sensing, but it is not always possible.

5 Techniques for Jamming

There are three popular methods for blocking or jamming radio frequency signals:

- **Technique for Spoofing:**

When a UAV or drone is detected, the jammer sends a signal to disintegrate it, or in some situations [10], an adaptive device sends an alert notification to the drone's owner whether the drone has actually been recorded in the database to travel sufficiently far from the confined area.

- **Technique for Shielding Attacks:**

It is an electro-motive force (EMF) defensive technique for drone/UAV signals, in which UAVs (receivers) are unable to interact with the sender, and the secured region is shielded against every illegitimate drone intrusion.

- **Denial-of-Service (DOS) Technique:**

In this method, a jammer device transmits a noise signal at the same frequency response as the drone in order to reduce the drone's signal-to-noise ratio (SNR) to the lowest possible level [10].

6 Signal Jammer Design Parameters

According to the techniques used during jamming, the drone jammer system is linked to the denial-of-service technique, in which the system transmits relatively close signal frequencies of 2.4 and 5.8 GHz. To allocate the system configuration, it is necessary to concentrate on parameters within the interface. The following are the key parameters:

- **The Distance (*D*) to be Jammed:**

The gap between the jammer and the spectrum that needs to be blocked or jammed is critical in the design since the amount of jammer power output is dictated by the spectrum that needs to be blocked or jammed. The relationship among *O*/*P* power and distance (*D*) will be clarified in the next step. The proposed design for this project has a 500 M allocation.

- **The radio frequency:**

The project's radio frequency is set to match that of the most recent drone technology. The most popular frequencies used by drone manufacturers are 2.4 and 5.8 GHz. As a matter of fact, the project's aim was to increase the frequency of this form of reducing drone technology.

7 Concept of Jamming

The aim of signal jammers is to prevent or disrupt communication between two or more parties. The main aim of the device in this case is to obstruct contact between the operator and the tasted drone. The basic jamming case is depicted in Fig. 1 by a triangle.

The jammer system sends out radio frequency fitting the frequency utilized by the tasted drone operator in order to obstruct. Both transmissions (jammer and operator) must be received by the tasted drone, with the higher frequency taking precedence, which appears to mean that if the jammer's strength is greater than the tasted drone operator's power, all interaction or distribution among the tasted drone and its operator would be disrupted. This project streamlines all relevant points in order to provide a clear understanding of any signal interference. However, in order to comprehend the jamming method [4], further parameters will be introduced to define the signal levels of two specific ties. First and foremost, one of such critical parameters is the signal-to-noise ratio (SNR). The signals from the jammer to the tasted drone (receiver) could be lesser or lower than the communication signal among the operator and the tasted drone in certain situations, but it will still be dominant. The precise explanation of this scenario is that, in order to interfere with one of the signals at its input, the tasted drone (receiver) usually needs a specific distance between the two signals (which is generally the stronger). However, in complicated communication systems, this difference can be negative when the transmission signal is much lesser

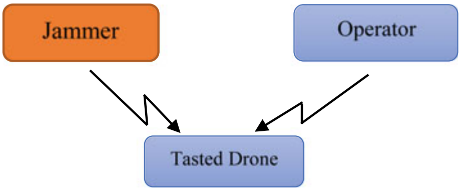


Fig. 1 Jamming triangle

than the jammer device's signal, and the tasted drone (receiver) is still capable to communicate with the signal of the tasted drone operator. In this project, we would not consider SNR and instead make the assumption that the higher signal at the tasted drone's receiver will always benefit.

8 System Design

The key block diagram of the jammer device is shown in Fig. 2. The system is divided into four stages, each of which complements the others. Power supply, intermediate frequency (IF), radio frequency (RF), and jamming signal are the components in order. The first stage generates power, which is then combined with noise and then saw tooth signals in the mixer, until the signal is sent to the amplifier and ready for transmission to the target.

- **Power Supply:**

The drone jammer parts receive DC power from the power supply in order to operate the competency level. The transformer, rectifier, filter, and regulator are the four main components of power supply, as seen in the diagram (Fig. 3).

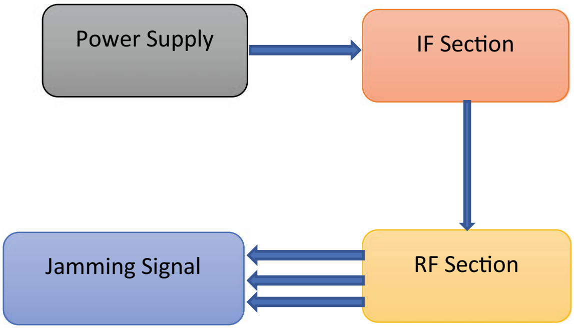


Fig. 2 Block diagram of jamming system

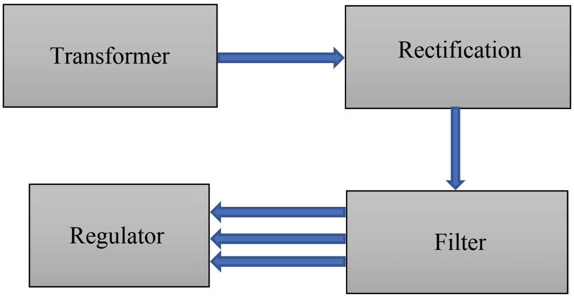


Fig. 3 Components of power supply

- **Intermediate Frequency (IF) Component:**

As an input to the VCO in the RF component, the intermediate frequency (IF) segment generates a tuning signal. This section is divided into three categories:

- **Noise Generator:**

This produces the desired amount of noise. A LM2931 amplifier, a resistor, and a Zener diode make up the noise generator. Since the Zener diode in the noise generator circuit is in reverse mode, it produces big band noise. The big noise signal should then be amplified, which will be accomplished in two steps. The NPN transistor will be used for the first level of amplification, and the LM2931 integrated circuit will be used for the second stage.

- **Ramp Generator:**

The ramp generator circuit generates a triangle waveform that is used to tune the voltage driven oscillator in the RF section. It is made up of an integrated circuit (LM2931), capacitors, and resistors. The LM2931 generates squared signal from pin 1 that is fed into pin 6, which is inverting the input through a resistor. After that, pin7 will produce the triangle signal [4]. A resistor connected to the output side of the circuit can also be used to change the ramp level.

- **Mixer:**

In a drone jammer, the mixer circuit combines the triangle and noise signals to produce the necessary output for the RF component. In this application, the LM358 operational amplifier serves as a mixer, with pin 2 receiving signals from the ramp and noise parts. As a result, the circuit's output is a tuning signal (Fig. 4).

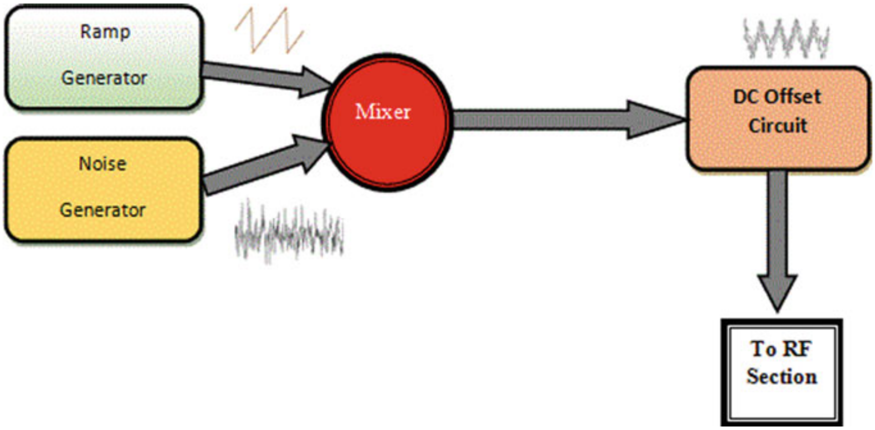


Fig. 4 Workflow of intermediate frequency component

- **RF Component:**

The RF section is the most important part of the mobile jammer; it consists of the voltage controlled oscillator (VCO), RF power amplifiers, and the antenna. These components were selected according to the desired specification of the jammer such as the frequency range and the coverage range. It is divided into three parts:

- **Voltage Controlled Oscillator:**

This system generates RF signals by sweeping the IF segment's output signals from the least desired frequencies to the max desired frequencies. As the applied input voltage changes, the output frequencies change as well, and the two have a direct proportional relationship. Along with other components, the Voltage Regulated Oscillator is a key component in the circuit of a jammer system. It is possible to refer to it as the drone jammer's core. Two VCOs were used in this section, one for the 2.4 GHz band and the other for the 5.8 GHz band, since the drone adapters use these two frequencies.

- **Power Amplifier:**

For the two frequencies of 2.4 and 5.8 GHz, the power amplifier receives signals from the VCO and amplifies them up to 33 dbm. It is used to maximize the jammer's coverage area while also increasing the signal jamming capacity. Furthermore, there are two types of tuning circuits: open loop and closed loop. The closed loop is straightforward, requiring only a few op-amps and a few passive components. This type of tuning circuit generates VCO transition from lowest to highest frequency through a saw tooth wave. This form uses a phase locked loop (PLL) to continuously control the frequency of VCOs. This is the most common form used in radio receivers. The open loop style was used in this project.

- **Antenna:**

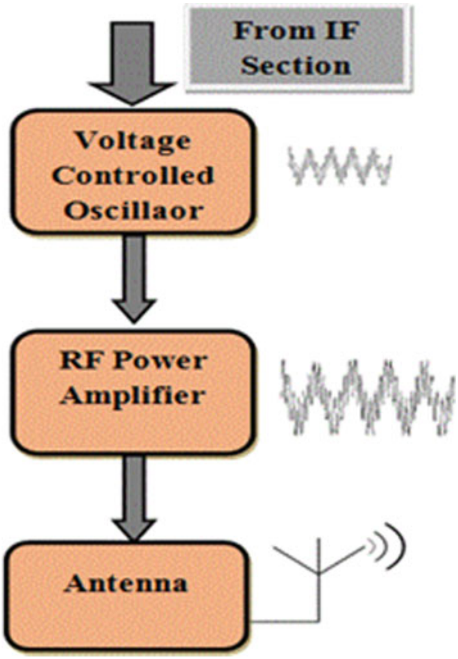
  Two antennas are proposed to be used to relay the jammer signals:

- Parabolic 2.4 GHz Antenna with 24 dB Gain.
- Parabolic 5.8 GHz Antenna (this antenna will not be used in this project due to its high cost and long delivery time).

9 Illustration of a Drone Jammer

This project's model was created with the help of the Proteus 8 technical software (student version). The complete system configuration of drone jammer devices is shown in Fig. 6. Even though this project is known as a professional project based on its concept and nature, the academic edition has drawbacks and is not intended to fulfill business requirements. The workflow of Radio Frequency is shown in Fig. 5.

Fig. 5 Workflow of RF component



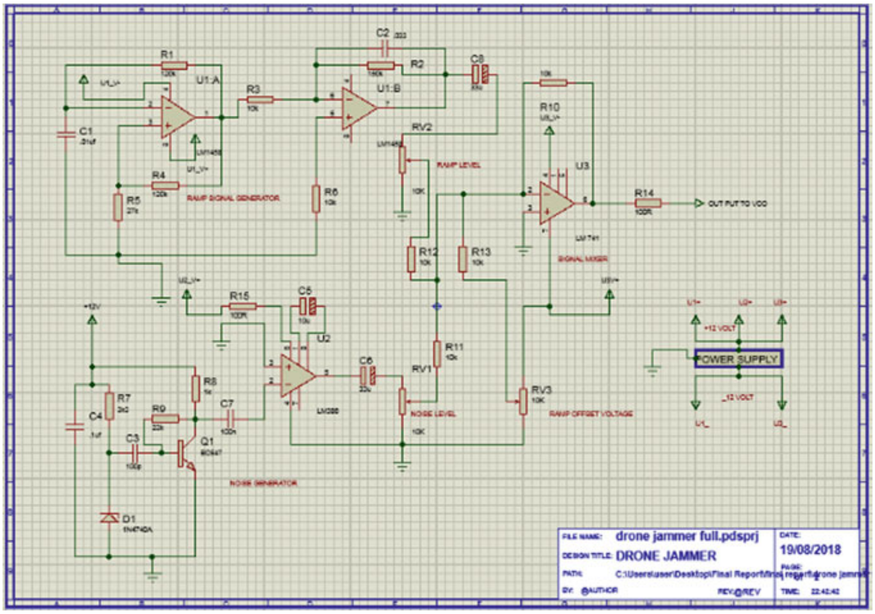


Fig. 6 Drone jammer system schematic (Proteus 8 professional)

10 Results

- **Intermediate Frequency Component:**

A triangle wave generator, noise generator, and mixer are included in the intermediate frequency segment. As a result, the VCO output is swept through the appropriate frequency spectrum, beginning at the lowest frequency and increasing to the desired highest frequency. As a result, the final output signal of the intermediate frequency section is a combination/gathering of the triangle wave generator's output signal and the noise generator's output signal, as seen in this segment. Since the main part of the mixer is responsible for the summation of these two signals, this role is connected to it. The triangle generator component is shown in the diagram below before it is combined with the noise generator signal. The signal is rendered using the output of the triangle generator. The system's output is a digital oscilloscope included in the Proteus 8 professional that generates triangle waves. Instead of relying on the spectrum analyzer all of the time, the FLUKE system was utilized to demonstrate the output in a real-world setting. Figure 7 shows the final output for the triangle circuit.

To generate tuning voltage for the Voltage Regulated Oscillator, a noise signal must be combined with the triangle signal. This noise would help to mask the core shape of the jamming transmission, appearing to outside observers as random/normal noise. So, without a noise wave generator, the jamming signal is just standard sweeping that carries the signal. The noise wave signal is depicted in Fig. 8. Figure 9 depicts the final output for the noise circuit in real-world conditions.

The mixer is the final step of the IF segment, and it is here that the signal from the triangle and the noise generator is combined using an Op-Amp to generate the

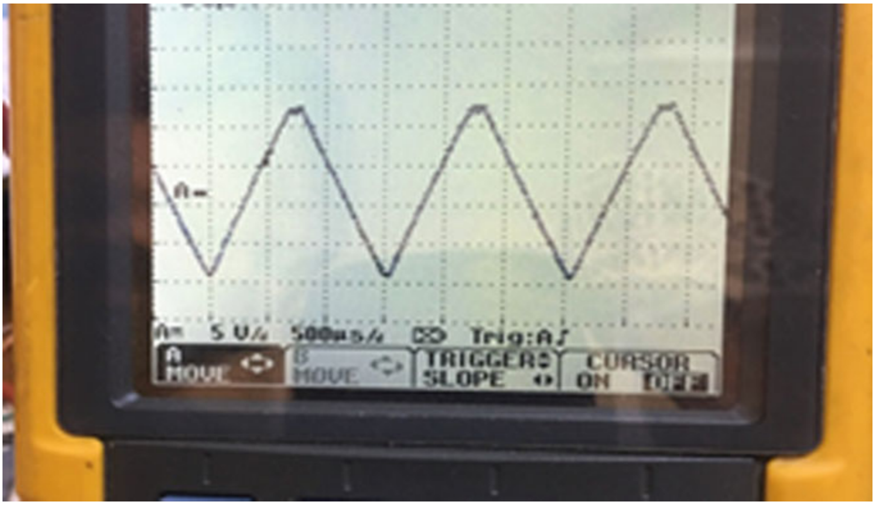


Fig. 7 Triangular wave signal (FLUKE Device)

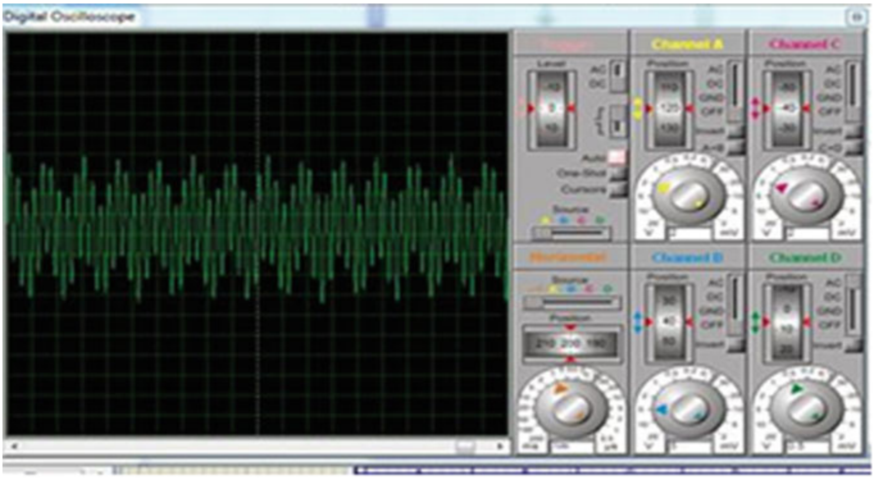


Fig. 8 Noise wave signal (Proteus 8 professional)

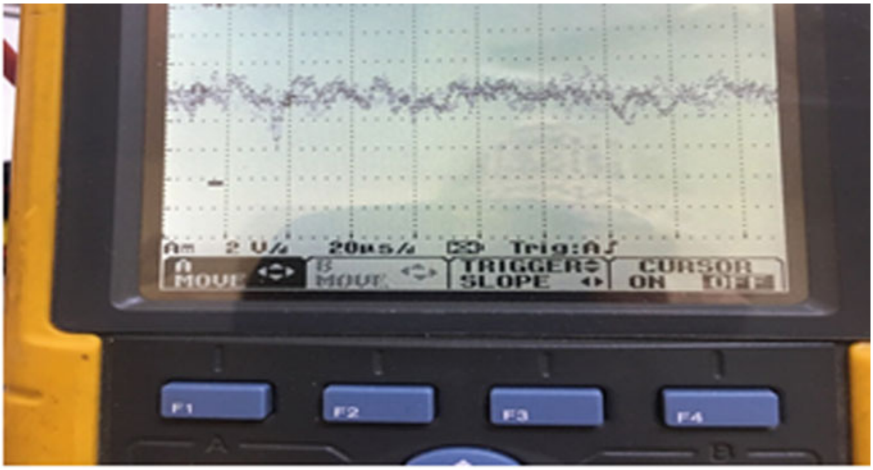


Fig. 9 Noise wave signal (FLUKE Device)

required signal for the jammer unit. The device meets the goal and generates the appropriate signal for the first step, as shown in Fig. 10, where the triangle and noise wave signals are combined to produce the desired signal (Fig. 11).

- **Radio Frequency Component:**

The radio frequency component of a signal jammer is crucial because its output can interface with the signal of the test drone frequency. The VCO, power amplifier, and antenna are all part of the radio frequency segment. The Voltage Regulated Oscillator

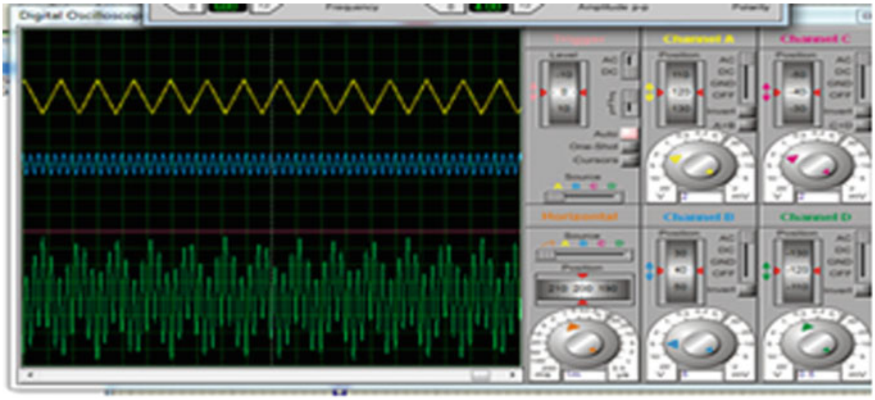


Fig. 10 Mixer wave signal (Proteus 8 professional)

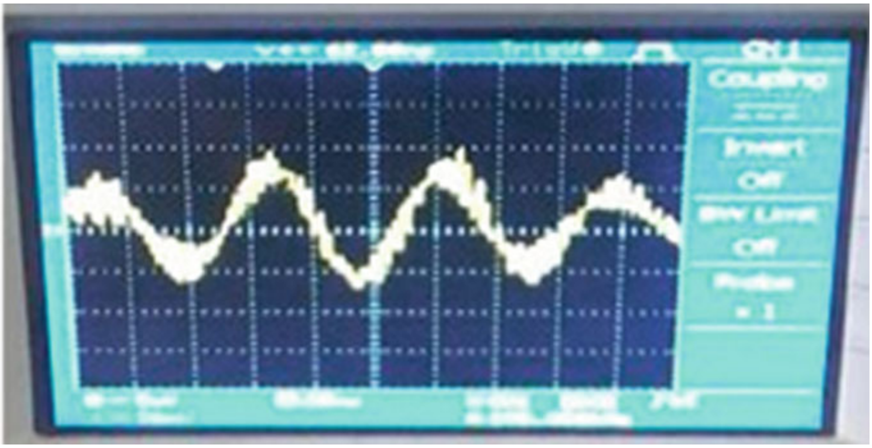


Fig. 11 Mixer wave signal (Spectrum Analyzer)

is solely responsible for generating the radio frequency signal that overpowers the signal of the test drone, causing contact between the test drone and the remote control to be disrupted (Operator). This segment's output provides a signal with the same frequency range as the test drone unit, but at a higher power level. It is necessary to have a directional antenna with a gain value of 24 db for transferring the signal from the jammer to the test drone in order to achieve the maximum degree of transmission (Fig. 12).

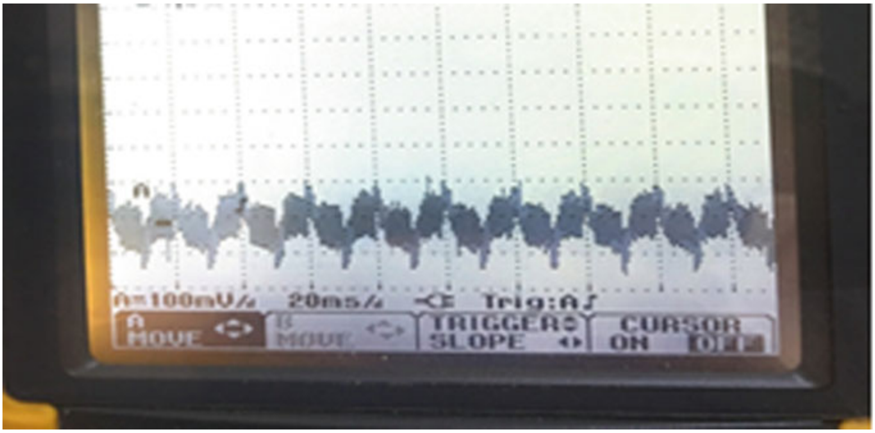


Fig. 12 Performance of the jammer device as a simulated waveform (FLUKE Device)

11 Conclusion

The simulation and analysis findings of the various components and parts in the jammer system circuit, as shown in the previous figures, indicate that the power signal can be changed to satisfy the desired requirement for interfacing with the test drone signal. When the final signal of the jammer system is compared to the IF segment, it is clear that the output power is far higher than when it has passed through more amplification. The project met its goals of producing a 2.4 GHz radio frequency signal in order to jam a 2.4 GHz test drone. When the prototype was tested within a single room, the effective jamming distance was 500 m. As seen in this project, power must be carefully monitored and calculated in order to prevent signal jamming weaknesses. The developers should pay more attention to the temperature aspect because it is crucial in building a good drone jammer.

References

1. S. Shahdadpuri, J. Patel, GSM mobile phone jammer. IJSRD**2**(08), 154 (2014)
2. K. Aryan, R. Karthikeyan, M. Aditya, K. Nikhil, Signal jammer. Degree Bachelor of Engineering, Dept. Electronic & Communication Engineering, BMS Institute of Technology and Management (2015–2016)
3. A. Pardhasaradhi, R.R. Kumar, Signal jamming and its modern applications. Int. J. Sci. Res. **2**
4. A. Sârbu, D. Neagoie, Wi-Fi jamming using software defined radio, in *International conference Knowledge-Based Organization.* (Sciendo, 2020), pp. 162–166
5. K. Pärlin, Jamming of spread spectrum communications used in UAV remote control systems. Tallinn University of Technology, School of Information Technologies, Thomas Johann Seebeck Department of Electronics, (2017)
6. S. Mili, et al., Jamming detection methods to protect railway radio communication. Methods **4**(7) (2015)

7. D. Looze, et al., Current counter-drone technology solutions to shield airports and approach and departure corridors (2016)
8. P.L. Lineswala, S.N. Shah, R. Shah, Different categorization for jammer: the enemy of satellite navigation, in *2017 2nd International Conference for Convergence in Technology (I2CT).* (IEEE, 2017), pp. 282–287
9. P. Hell, M. Mezei, P.J. Varga, Drone communications analysis, in *15th International Symposium on Applied Machine Intelligence and Informatics [online].* Held 26–28 January 2017 (2017)
10. I. Güvenç, et al., Detection, localization, and tracking of unauthorized UAS and jammers, in *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC).* (IEEE, 2017), pp. 1–10.
11. S.K. Bhatia, K. Sharma, K. Chaudhary, D. Singh, Signal jammer and its applications. Int. J. Electr. Electron. Res. [online] **3**(2), 463–467 (2015)
12. https://www.mwrf.com/technologies/test-measurement/article/21848712/introduction-to-rf-wireless-communications-systems

Performance Comparison of Machine Learning Algorithms in Symbol Detection Using OFDM

Siva Satya Sri Ganesh Seeram, Avuthu Yuvaraja Reddy, N. J. Basil, Akella Viswa Sai Suman, K. Anuraj, and S. S. Poorna

Abstract Nowadays, vast amounts of data transmission and data retrieval are crucial and are done using many ways. Orthogonal Frequency-Division Multiplexing (OFDM) is one of the efficient ways to transmit data with the help of orthogonal subcarriers, which is used in applications such as WiFi, WiMax, and cellular communication. In this paper, instead of conventional detection techniques, machine learning (ML)-based methods are adopted to detect the symbols after data is being received through the Additive White Gaussian channel (AWGN). Detection is one of the areas in which the bit error rate (BER) performance of the OFDM system can be improved. Machine learning algorithms only depend on the training data to predict the outputs; hence, we can detect the symbol even without the use of cyclic prefix or channel estimation which can save a lot of time and data if the input data is large. A comparative study on the performance of receivers based on K-means clustering, k-nearest neighbors classifier, support vector machine, linear discriminant analysis, and quadratic discriminant analysis is done. The modulation techniques such as BPSK and QAM with various modulation orders ranging from 4 to 64 are used in this analysis. Performance comparison of aforementioned detection techniques using employing machine learning is done using BER vs signal-to-noise ratio (SNR) in the range of 0–30 dB.

Keywords OFDM · Monte Carlo simulation · Elbow method · KNN · K-means · SVM · LDA · QDA

1 Introduction

As the world is becoming more and more digital, the number of wireless devices being used increases enormously, which brings more challenges for reliable transmission of data. The obsolete technologies, viz. GPRS, radio, etc., are not, at this point, utilized. In order to address the increasing demands from client side such as reduced inter-symbol interference (ISI) and high data rate, OFDM was introduced. The high data

S. S. S. G. Seeram · A. Y. Reddy · N. J. Basil · A. V. S. Suman · K. Anuraj · S. S. Poorna (✉)
Department of Electronics and Communication Engineering, Amrita Viswa Vidyapeetham, Amritapuri, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_36

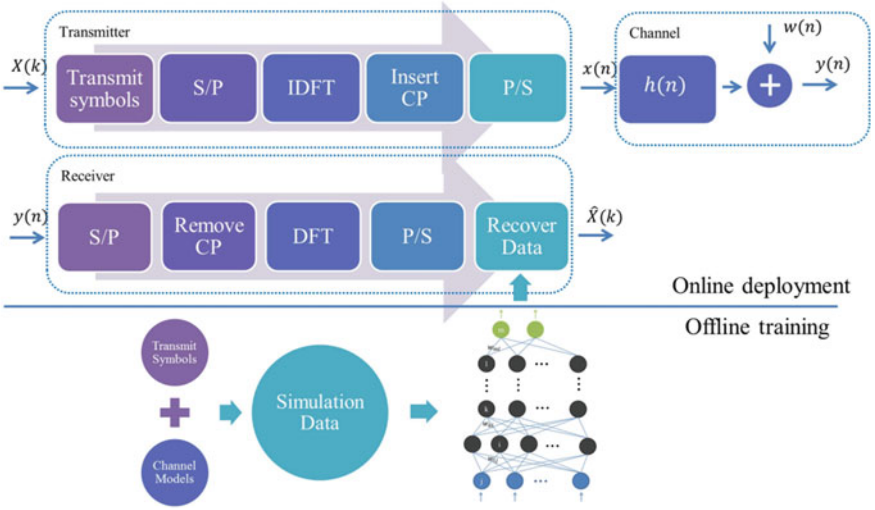


Fig. 1 Visualization of Conventional deployment versus Artificial Intelligence

rate is due to a single enormous data stream transmitted over a scope of low-speed sub-carriers [1]. Since these sub-carriers are orthogonal, it can mitigate ISI without the use of guard bands, hence reducing the required bandwidth for transmission. This OFDM methodology can even be applied to trending technologies like the Internet of Things (IoT) and 5G [2].

From the beginning itself, various new methods have been introduced, in order to improve different aspects of conventional OFDM system [3]. From Fig. 1, the process involved in the detection part of a receiver for both conventional and the artificial intelligence techniques is shown. For an AI-based detector, there will be less task for a detector to process and detect the data. This work focuses on comparing an OFDM system implemented using ML with different algorithms. ML is developed for processing and analyzing large amounts of data [4]. The rapid growth processing powers and data handling capabilities of various computation devices enable the usage of large amounts of data, enabling ML systems to make more intelligent decisions than ever [5]. Many of the applications in the modern world utilize ML algorithms to analyze and process this data. These algorithms adapt according to the data and are able to recognize patterns in the data much faster with greater accuracy. Similarly, the progressions in wireless innovations can give a better solution to the mission for sending huge volumes of information, inside the limited range accessible [6].

ML has proved its efficiency in the underwater acoustic (UWA) communications which seems to perform better than conventional schemes like least square and deep neural network (DNN) in estimating the parameters and scheme selection. The algorithms like convolutional neural networks (CNN) and random forest (RF) seem to perform well in selecting the scheme between CDMA and OFDM in the UWA channel [7]. DL-based OFDM system offers a high spectral efficiency than traditional OFDM system in UWA channel [8].

In [9], DNN was used to estimate the channel state information (CSI) for detecting the transmitted symbols directly. This method is compared with traditional equalization techniques such as minimum mean square error (MMSE), where both use quadrature phase-shift keying (QPSK) modulation. The results showed that the deep learning method provided better performance in symbol detection and channel estimation, rather than the conventional schemes.

In [10], a deep complex-valued convolutional network (DCCN) can be used to replace the process of DFT/IDFT in OFDM system. Using the LMMSE graphs to compare the performance with traditional receivers, the DCCN algorithm provides better SNR performance in Rayleigh fading channels and reduces the complexity of the process along with mitigating ISI.

Signal detection in an OFDM system directly affects the performance [11] and complexity of the system, hence improving its efficiency is important [12]. This work examines the use of ML algorithms, viz. K-means clustering, k-nearest neighbor (KNN), support vector machine (SVM), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) as a symbol detector and analyzes which scheme provides better performance for image transmission. The modulation techniques used are binary phase-shift keying (BPSK) and quadrature amplitude modulation (QAM) of orders 4 to 64. The performance is analyzed with the use of a BER vs SNR graph in the presence of additive white Gaussian noise (AWGN) [13]. This analysis determines which modulation along with which ML algorithm provides better performance for data transmission. This paper gives a brief description about conventional ML algorithms in the next session followed by methodology, results, and analysis, and finally the conclusion.

2 Machine Learning Algorithms

ML is becoming a booming trend day by day, and it takes a major role in our daily activities. In this paper, supervised learning algorithms such as KNN, SVM, LDA, QDA, and unsupervised learning algorithm, viz. K-means clustering, are used. KNN algorithm classifies the data based on the Euclidean distance from each test data to 'k' nearest train data samples. The majority of the points out of 'k' belonging to a particular class is defined as the predicted output [14]. The accuracy of the algorithm is based on the train data and the parameter 'k'. If 'k' varies, then the classification may differ. An optimum value of $k = 15$ is set so that the accuracy is maximum. The optimum value is found for corresponding data set using the elbow method.

SVM is a trained classifier which classifies the test data by drawing the boundaries between the classes, i.e., a hyperplane in N-dimension, where N is the number of features. For the case $N = 2$, the hyperplane is a line. The accuracy is based on the data set. If data has more noise, then the accuracy is not good. Compared to other algorithms, for large data sets the time taken to classify is very high [15]. If the classes of the algorithm increases, the accuracy seems to be degraded and this may be due to the boundary clash with other classes.

LDA computes a linear combination of features that characterizes or distinguishes two or more entity or event classes. LDA is a linear classifier. QDA is a variant of LDA which allows nonlinearity for separation of classes. There will be a separate co-variance matrix for each class. Also, the accuracy of both of these algorithms is affected by the size of training data.

K-means clustering classifies the entire data into K clusters based on the data distribution and pattern. That is, it determines an optimum positioning for K centers by iterative regression, so that the sum of all distances is minimal. The selection of initial centroids and the calculation of the distances play a major role in the accuracy of the output classification [14]. The amount of noise in the data set also plays role in affecting the accuracy.

3 Methodology

The entire implementation of this research work is done in three phases, which is shown as a flowchart in Fig. 2. The complete work is done using MATLAB simulation software.

3.1 Data Preprocessing

To transmit the data like image, audio, text through a transmitter, first the transmitter should have an input data as shown in Fig. 4 which should be preprocessed. This process is done in this phase take whatever data, for instance consider an image. The input 2-D grayscale image should be first converted into 1-D vector array which consists of quantized values from [0–255]. These values are now converted to binary bit strings. This 2-D matrix of bit strings are converted into vector again. Now the obtained series of bits of data can be used as input data and is further passed on to the next processing blocks of the implementation model shown in Fig. 4.

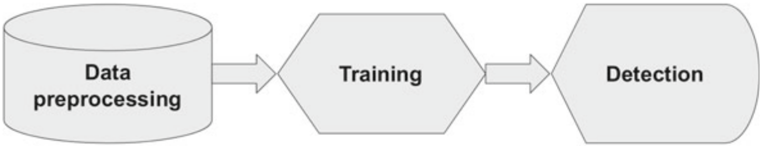


Fig. 2 Flowchart of the implementation

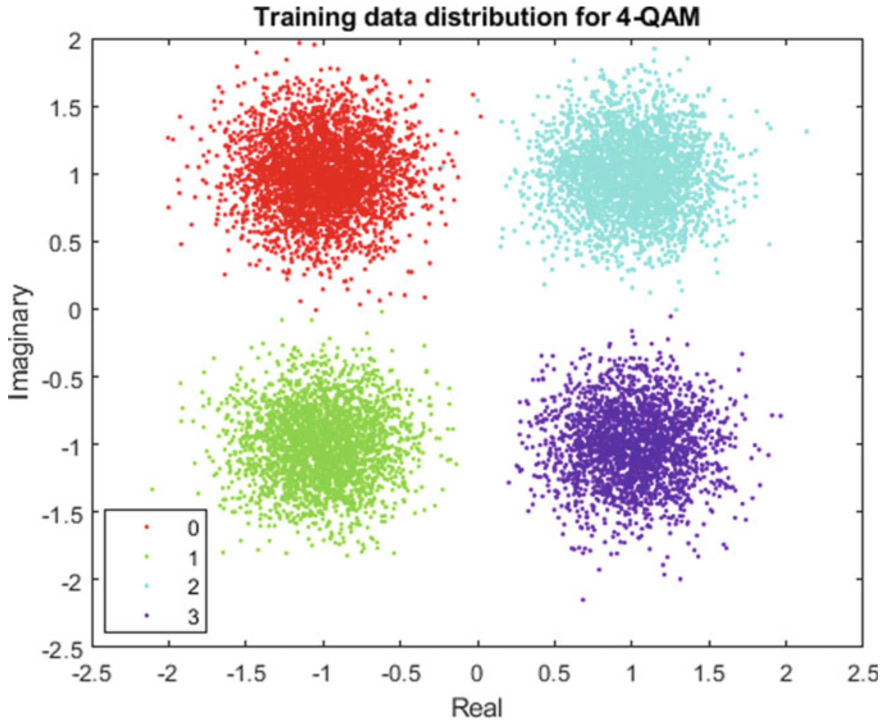


Fig. 3 Train data distribution with classification

3.2 *Training*

For the proposed algorithms to detect the received symbols, they need to be trained. In order to collect the training data to train the models, the implementation model in Fig. 4 is partially executed. For instance, here the generation of training data set for 4-QAM modulation is discussed. The image is loaded, and data pre-processing is carried out; the modulation scheme used is BPSK, and then it is passed through OFDM and then transmitted into the AWGN channel; all this process will be explained briefly in the next section. At the receiver, the data is received and at the output of OFDM the data is recorded with coordinates (x, y) in the constellation plane as the features and labels as the input of modulator block at transmitter as shown in Fig. 4. The noise in the channel [16] should be such that while training, it shouldn't be an issue for the performance of the detector. In Fig. 3, the data is scattered from the ideal position. Scattering should not be more and not be less. The data distribution similar to this works like a training data set for ML-based detectors.

The sample training set of few samples for 4-QAM is shown in Table 1. The number of samples in the training set taken plays an important role. In Table 2, the effect of number of sample with increasing the modulation can be observed. As the

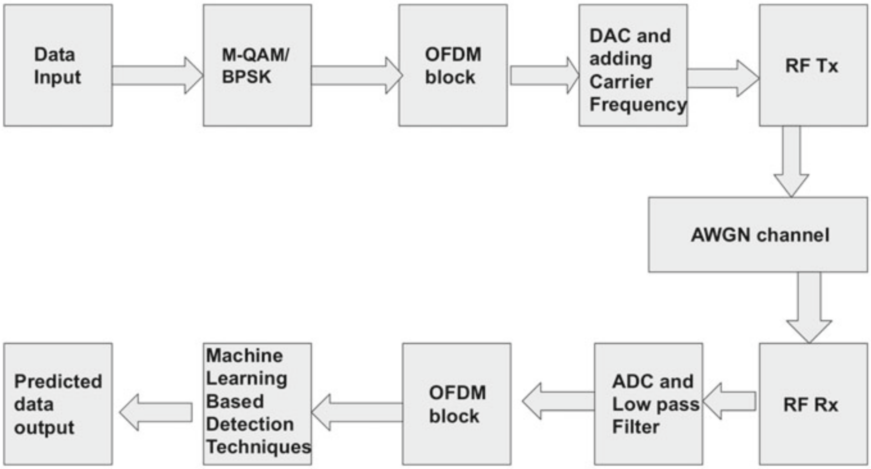


Fig. 4 Block diagram of the implementation model

Table 1 Sample training set points for 4-QAM

X	Y	Label
−1.1311	−1.3469	1
0.68447	−0.9313	3
1.7292	0.76227	2
0.66271	−1.022	3
−1.1426	1.0005	0

modulation order increases, the training points for each symbol decrease. Therefore, a large size image is given as input in order to increase the number of total bits. It can be seen that for 64-QAM, initially there are approximately 352 points for each symbol in training data set. When a large image is given as input, then there will be approximately 1420 points for each symbol in the training set. Hence, better training of data and hence better performance of the model.

In similar way, the training data sets for all the modulation schemes are generated and the models are trained to detect the symbols. For KNN-based detector, the performance of the detector varies with parameter 'k' and also modulation scheme. With optimal value chosen from elbow method, the parameter 'k' is set for better performance.

Table 2 Training samples with increasing order of modulation

Sample data bits	Order of modulation	Training samples for each symbol
22528	2	11,264
22528	4	5632
22528	8	2816
22528	16	1408
22528	32	704
22528	64	352
90912	2	45,456
90912	4	22,728
90912	8	11,364
90912	16	5682
90912	32	2841
90912	64	1420

3.3 Detection

The data obtained, after pre-processing of the image, is fed into modulation block where the bits are mapped to symbols. This can be visualized by constellation points in a 1-D or 2-D plane for BPSK and M-QAM, respectively. If the scheme is BPSK, then in a 1-D plane, bit 1 is mapped to +1 and bit 0 is mapped to -1. Similarly, if 4-QAM is used, then bit 00 is mapped to (1,1) and bit 01 is mapped to (-1,1), and so on.

An OFDM system is developed for the transmission [17] of input data (like image, audio & text) as shown in Fig. 4 and is chosen as the transmission system in this work because of its better spectral efficiency and higher data rates [18] than the non OFDM technology [19]. The inverse fast Fourier transform (IFFT) is used to modulate the sub-carriers with the data, and the cyclic prefix helps in maintaining the orthogonality between them, hence providing ISI-free transmission.

The channel designed is AWGN. This data is now received at the receiver end. First the signal is down-converted using a low-pass filter and passed through ADC, and the data is passed through OFDM block in order extract the data from the sub-carriers. The resultant output of it is given as input to the detector. The data obtained may be noisy. First the data is passed through one of the trained ML detectors. The output from the detector is observed and plotted on BER graph with varying SNR. Since the channel noise is random, the result obtained is not so reliable. In order to rely on the results, the simulation needs to be repeated several times and the average of them is considered to be reliable. This is called Monte Carlo simulation. The same process is followed every time for each detection technique and the modulation scheme when varied.

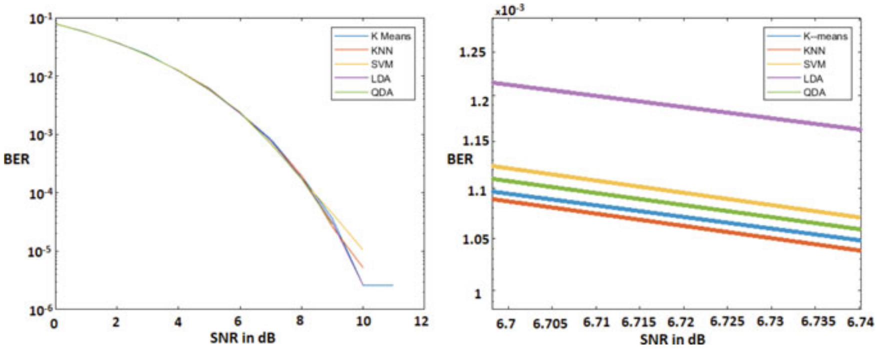
In this way, the results are simulated and are explained briefly in the next chapter.

4 Results & Analysis

Evaluation of the performance of the system is done using the BER curve. Monte Carlo simulation is carried out to obtain these curves.

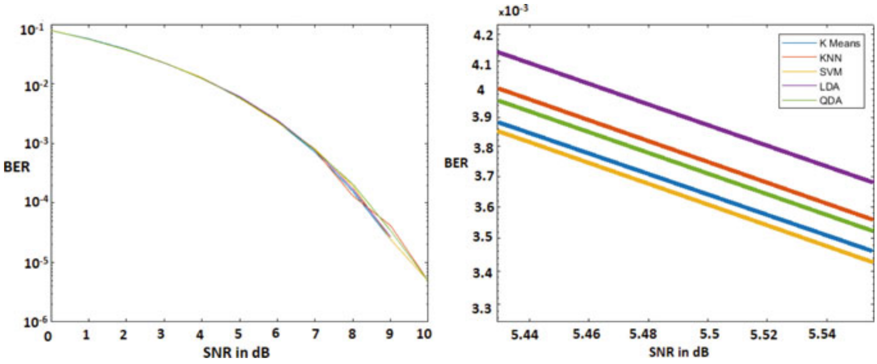
The resultant BER graph is plotted for each modulation scheme when passed through the five detection techniques and through AWGN channel are discussed. So for each modulation scheme, there will be five plots, each passed through AWGN channel.

The results with BER versus SNR for the system implemented using BPSK and 4-64 QAM are shown in Figs. 5, 6, 7, 8, 9 and 10, respectively. From Fig. 5a, it can be seen that for BPSK modulation, BER curve of algorithms seems to be overlapping. So the semilog graph of BER versus SNR is enlarged in Fig. 5b. From the enlarged figure, it can be seen that KNN, K-means & QDA outperforms other algorithms



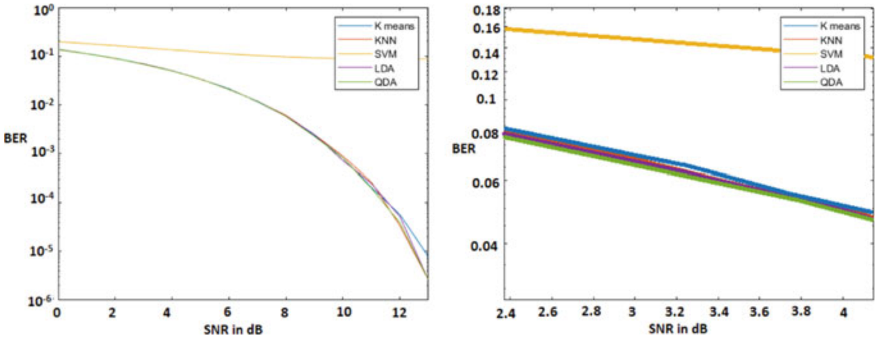
(a) Generated Semilog graph (b) Enlarged graph for analysis

Fig. 5 BER versus SNR graph for BPSK



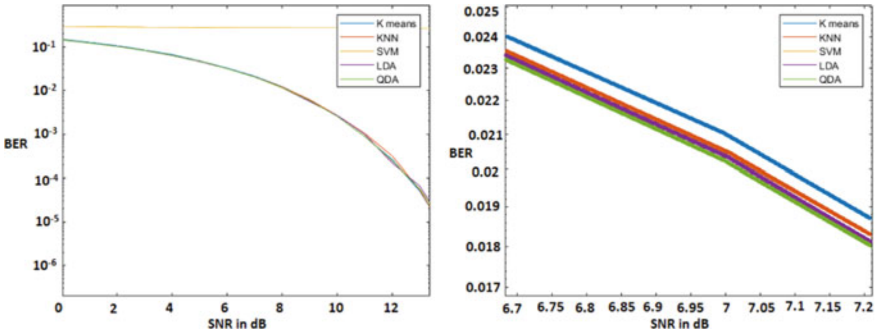
(a) Generated Semilog graph (b) Enlarged graph for analysis

Fig. 6 BER versus SNR graph for 4-QAM



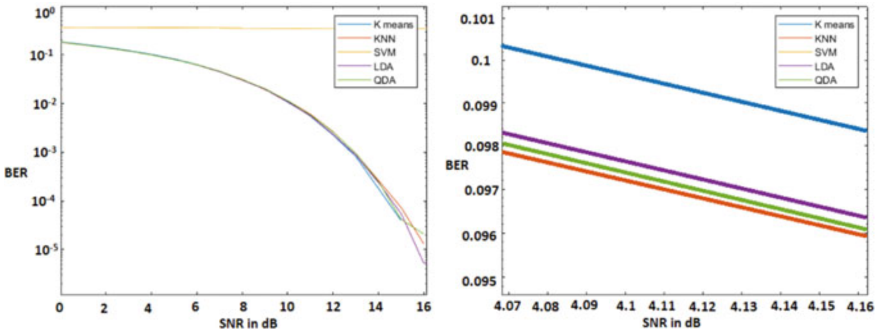
(a) Generated Semilog graph (b) Enlarged graph for analysis

Fig. 7 BER versus SNR graph for 8-QAM



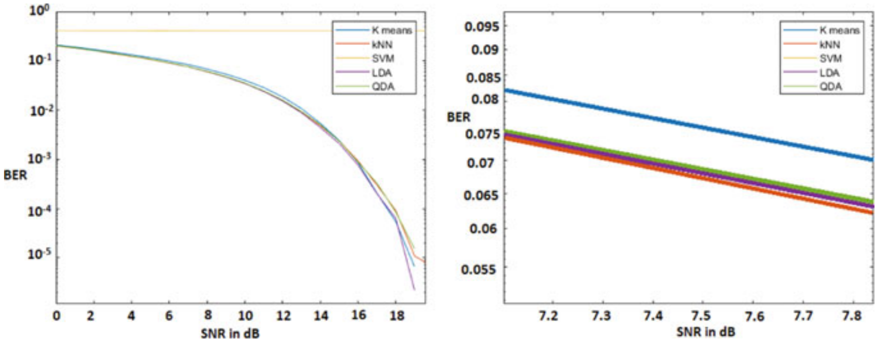
(a) Generated Semilog graph (b) Enlarged graph for analysis

Fig. 8 BER versus SNR graph for 16-QAM



(a) Generated Semilog graph (b) Enlarged graph for analysis

Fig. 9 BER versus SNR graph for 32-QAM



(a) Generated Semilog graph (b) Enlarged graph for analysis

Fig. 10 BER versus SNR graph for 64-QAM

Table 3 Performance ranking of ML detectors for each modulation scheme

Performance	1	2	3	4	5
BPSK	KNN	K-Means	QDA	SVM	LDA
4-QAM	SVM	K-Means	QDA	QDA	LDA
8-QAM	QDA	LDA	KNN	K-Means	SVM
16-QAM	QDA	LDA	KNN	K-Means	SVM
32-QAM	KNN	QDA	LDA	K-Means	SVM
64-QAM	KNN	LDA	QDA	K-Means	SVM

with a lower BER as compared. The graphs for 4-QAM are shown in Fig. 6a, b, respectively. The same for 8-QAM is shown in Fig. 7a, b. For 4-QAM, it can be seen that SVM, K-means & QDA outperform other algorithms with a less BER, while using 8-QAM, QDA outperforms other algorithms. In 16-QAM, shown in Fig. 8a, b, QDA, LDA & KNN give better performance. The results repeat for 32-QAM, as shown in Fig. 9a, b, respectively, as well as for 64-QAM shown Fig. 10a, b. In both the cases, all other algorithms outperform SVM.

In Table 3, the whole analysis of the results is summarized as the performance metrics of each detection technique with varying modulation schemes. This table gives quick analysis and better idea on understanding the plots and also concludes the discussion.

5 Conclusion and Future Work

In this work, communication system using OFDM as transmission technique and detector based on ML is implemented. The goal of this research is to compare which of ML algorithms is able to provide better BER performance, for different modulation schemes used.

The detector part of the receiver is replaced with ML-based detectors using algorithms like KNN, SVM, LDA, QDA, and K-means.

This paper concludes that OFDM data transmissions in the range of SNR 0–30 dB, for low-order modulations like BPSK and 4-QAM, the ML algorithms like SVM, K-means & QDA perform with a good BER. For higher modulation orders such as 8, 16, 32, 64-QAM, the ML algorithms like QDA, LDA & KNN seem to have a better performance. SVM performs worst in the higher modulation schemes. This concludes that ML algorithms can be implemented in place of usual detectors like maximum likelihood, zero forcing, huge data handling capacity.

Future work aims at replacing the ML algorithms with deep learning techniques, and subsequently compared.

References

1. A.Y. Reddy, B.L. Reddy, A. S. Naga Veera Sai, MSE and BER analysis of text, audio and image transmission using ML based OFDM, *IEEE International Conference for Innovation in Technology (INOCON)*. BANGLURU, 1–3 (2020). https://doi.org/10.1109/INOCON50539.2020.9298204
2. D. Chen, Y. Tian, D. Qu, T. Jiang, OQAM-OFDM for wireless communications in future Internet of Things: a survey on key technologies and challenges. IEEE Internet Things J. **5**(5), 3788–3809 (2018). https://doi.org/10.1109/JIOT.2018.2869677
3. C.A. Origanti, C. Naveen, M.S. Sai Bhargav, MSE and BER analysis of OFDM in Rician channel, *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)* (48184) (2020), pp. 877–881. https://doi.org/10.1109/ICOEI48184.2020.9142941.
4. J.S. Raj, Machine learning implementation in cognitive radio networks with game-theory technique. J. IRO J. Sustain. Wireless Syst. (2), 68–75 (2020)
5. B. Tegin, T.M. Duman, *Machine Learning at Wireless Edge with OFDM and Low Resolution ADC and DAC* (2020) . arXiv preprint arXiv:2010.00350
6. V.K.R. Devana, A.M. Rao (2020) A compact 3.1-18.8 GHz triple band notched UWB antenna for mobileUWB applications. IRO J. Sustain. Wireless Syst. **2**(1), 1–12
7. Y. Kim, H. Lee, J. Ahn, J. Chung, Selection of CDMA and OFDM using machine learning in underwater wireless networks. ICT Express. **5** (2019). https://doi.org/10.1016/j.icte.2019.09.002
8. Y. Zhang, J. Li, Y. Zakharov, X. Li, J. Li, Deep learning based underwater acoustic OFDM communications. Appl. Acoustics **154**, 53–58 (2019)
9. H. Ye, G. Y. Li, B. Juang, Power of deep learning for channel estimation and signal detection in OFDM systems. IEEE Wireless Commun. Lett. **7**(1), 114–117 (2018). https://doi.org/10.1109/LWC.2017.2757490
10. Z. Zhao, M.C. Vuran, F. Guo, S. Scott, *Deep-Waveform: A Learned OFDM Receiver Based on Deep Complex Convolutional Networks* (2018). arXiv preprint arXiv:1810.07181
11. K. Anuraj, S.S. Poorna, S. Jeyasree, V. Sreekumar, C.A. Origanti, Comparative study of spatial modulation and OFDM using QAM symbol mapping, *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, Madurai, India (2019), pp. 1246-1249. https://doi.org/10.1109/ICCS45141.2019.9065516
12. X. Zhang, Y. Su, G. Tao, Signal detection technology research of MIMO-OFDM system, in *2010 3rd International Congress on Image and Signal Processing, Yantai, China* (2010), pp. 3031–3034, https://doi.org/10.1109/CISP.2010.5648241

13. C.A. origanti, K. Anuraj, S. S. Poorna, M. Gokul Krishnan, K. Greeshmanth, Performance comparison of orthogonal frequency division multiplexing in White Gaussian and Rayleigh channels, in *2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India* (2020), pp. 712–715. https://doi.org/10.1109/ICISC47916.2020.9171194
14. I. Mesecan, I.ö. Bucak, Searching the effects of image scaling for underground object detection using KMeans and KNN, in *2014 European Modelling Symposium, Pisa* (2014), pp. 180–184, https://doi.org/10.1109/EMS.2014.64
15. J.P. Zhang, Z.W. Li, J. Yang, A parallel SVM training algorithm on large-scale classification problems, in *2005 International Conference on Machine Learning and Cybernetics*, vol. 3 (2005) pp. 1637–1641. IEEE
16. M.M. Amiri D. Gündüz, Federated learning over wireless fading channels, in *IEEE Transactions on Wireless Communications*, vol. 19, no. 5 (2020), pp. 3546–3557. https://doi.org/10.1109/TWC.2020.2974748
17. X. Zhang, Y. Su, G. Tao, Signal detection technology research of MIMO-OFDM system, in *2010 3rd International Congress on Image and Signal Processing*, IEEE, vol. 7 (2010), pp. 3031–3034
18. K. Puntsri, E. Khansalee, Experimental comparison of OFDM SC-FDM and PAM for low speed optical wireless communication systems, in *2019 7th International Electrical Engineering Congress (iEECON)* (2019), pp. 1-4. https://doi.org/10.1109/iEECON45304.2019.8938969
19. A. Agarwal, B.S. Kumar, K. Agarwal, BER performance analysis of image transmission using OFDM technique in different channel conditions using various modulation techniques, in *Computational Intelligence in Data Mining*. Springer, Singapore, pp. 1–8

Security Schemes for Integrity Protection and Availability of Service in Cloud Environment: A Review

Amrutha Muralidharan Nair and R. Santhosh

Abstract Cloud computing is a blend of utility computing plus different form of service. This intense feature of cloud computing helps users to get many computational and cost-effective benefits. As a wild set of user groups approaching the cloud computing environment, the use of services and resources for storing the data has increased. Such a dynamic growth in the cloud faces many security challenges. The two main challenge is privacy protection of the data, that is, maintain the data integrity and other is the data availability. So, solving these security affecting challenges, we need proper knowledge related to the issues and its solution scope. Hence, this paper suggests a comprehensive study to identify diverse challenges and also the different solution to evade the challenges.

Keywords Cloud Environment · Security Issues · Privacy Preservation · Verification · Availability

1 Introduction

Cloud computing is a computing forum growing at a very fast pace which offers shared resources such as network, storage services, and applications based on user demands. The main focus of these trending technology is to implement, a provisioned and quick release of resources with minimum effort. Some researchers specify that "cloud is a style of computing with massive scalable IT capabilities." The cloud environment provides five significant feature, three service models, and three deployment models.

1.1 Significant Features

- On-demand services: Users can access the cloud as per their needs.

A. M. Nair (✉) · R. Santhosh
Department of CSE, Karpagam Academy of Higher Education, Coimbatore, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_37

- Broad access: Service provided by the cloud can be accessed from anywhere via a network.
- Resource Pooling: Multiple users are served the resource as their needs based on the multi-tenancy concepts. Dynamic allocation and removal are based on user demand.
- Location independent: User is irrespective of the knowledge of the exact location where its data is stored.
- Elasticity: Users have the right to scale up and down their demand at any time.

1.2 Service Model

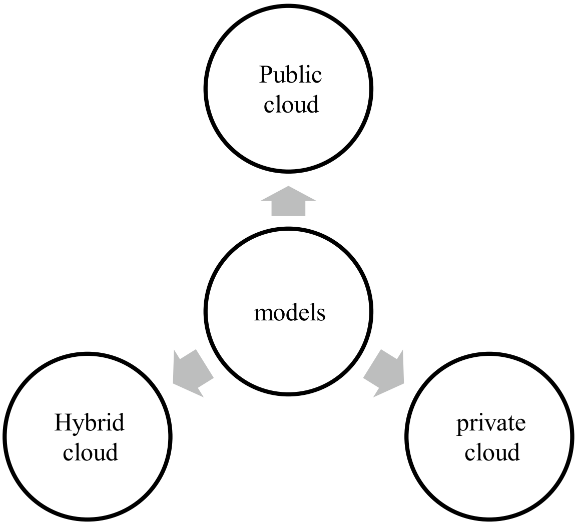
There are three types of service provide by the cloud:

- **Infrastructure as a Service (IaaS)**: This facility offers virtual infrastructure (like extra storage, virtual system, virtual networking devices, etc.).
- **Platform as a Service (PaaS)**: This service provides a provision that the user can work on a different platform using a single system.
- **Software as Services (SaaS)**: This provides different software to the user as their needs. It reduces the overhead related to installation and space required to establish the software.

1.3 Deployment Models

Cloud provides three different forms of the models shown in Fig. 1.

Fig. 1 Deployment model



- **Public cloud**: It provisioned for the public that can be operated by third parties or a single owner.
- **Private cloud**: To a particular organization or industry and operated by an organizational member.
- **Hybrid cloud**: It is a combination of public features and private features of cloud.

The promote section is divided into Sect. 2 security services and its challenges, Sect. 3 review of different data availability and integrity providing schemes, Sect. 4 conclusion about the comparatives study.

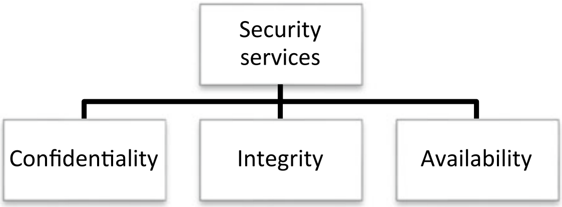
2 Security Services and its Challenges

Cloud computing provides a good range of advantages associated with scalability, elasticity, and self-provisioning. There are some security issues with the instability in guarantee to access the services (availability and confidentiality of the recorded data) [1]. In the cloud environment, the service provider helps to protect information stored in the server. This information or data stored in the cloud faces different vulnerabilities [2]. These vulnerabilities that occur in the network and server-side have to be identified and a proper dynamic solution needed to be explored [3].

During this section, we are browsing different vulnerability and security challenges within cloud computing. The foremost importance that the user particularly specializes in is the assurance of the data and services that the user is accessing [4]. Some security factors are shown in Fig. 2.

- Confidentiality: The data or the sensitive information transferred from the client-side to the server-side should not be disclosed to any unauthorized use within the network [5].
- Integrity: Data integrity means the data or valuable information transferred to the server-side should not be modified the data or information content by unauthorized users. Consistency of the information should be kept unbreakable by the external agent that is not in the cloud or not an authorized user.
- Availability: This is one of the most important features related to a cloud. This service of the cloud guarantees the accessibility of data but in certain situations, if the rate of traffic is high, the genuine user may not be able to access the services.

Fig. 2 Different security services



- Authentication: It is a process to assure that the user identity is authentic, and that the owner of specific data is verified.
- Authorization: It defines the rights of a particular user who will be able to access the data that is stored in the cloud. The owner of data can grant the accessing rights to every user who wishes to access the data.

2.1 Threats Affecting Cloud Environments

The utmost purpose of the threats is to effect the cloud delivery models where the information assets are residing. Table 1 shows cloud threats categorized based on the CIA triangle.

3 Review of Different Data Availability and Integrity Providing Schemes

3.1 Data Integrity

Cloud computing is an immense system that provides data storage facilities for data servicing, data processing, and remote data backup. Therefore, the user can access the data from anywhere [6]. Individual users or organizational associations, outsource integrity test processes to the third-party auditor to minimize the checking overhead on their server-side.

The data present inside the cloud server will be monitor continuously for maintaining the security of the data. Verification of data integrity against untrusted servers is that the main factor within the cloud data storage system. In case any modifications are made by the untrusted server, the client is in an exceedingly position to spot the particular changes within the cloud. The cloud contains multiple servers in which the data is stored. Therefore, technical knowledge of storage is avoided.

3.2 Verifying the Correctness of Data and Preserving the User Identity in the Cloud

Thomas et al. [7] proposed a basic, simple idea related to data integrity checking which is stored in a remote location. Here, the parity bits and data are blinded XoR using pseudo random sequence.

Ateniese et al. [8] provide a mechanism for public auditing scheme “provable data possession” in untrusted storage. It uses the “Homomorphic method” to generate tags

Table 1 Main threats that affect the CIA features

	Threats	Description
Confidentiality	Insider attack or internal threats	Multiple user will access the customer data in an organization so chance can occur in some malicious manner in: • Cloud provider side • Application side cloud customer • Third-party user side
	External threats	Here, all types of delivery models of cloud is affected. It can happen by providing remote software or hardware attack that affect the cloud infrastructure, cloud application
	Data leakage	This type of threats occur due to the human error or faulty hardware that will lead to compromise the information
Integrity	Data segregation	The threat in integrity occur when computing resource which is shared among the user and the resource is not properly segregated. mainly occur due to • Security parameters is defined incorrect • Configuration of virtual machine is not proper
	User access	Implementation of poor access control and user identity procedures
	Loss in quality of Data	Due to introducing of faulty application in cloud computing and infrastructure components
Availability	Denial of service	The attack target to deny the user or client access to a specify system service or resources. It mainly target the CPU utilization, bandwidth of a network, etc
	Session Hijacking	In which the session is hijack with lot of false packets toward the server
	Malware injection attack	Here unauthorized person will intrude malicious code to the cloud service, so that it will affect the network

and a procedure based on the RSA algorithm. The author introduces the concepts of independent third-party auditors that minimize the overload at the user side.

Juels et al. [9] explored the new methodology of “proofs of retrievability (POR).” This outline enables the verifier to generate a concise proof report of a targeted file that the user is retrieving. It uses the kind concepts of “cryptographic proof of knowledge” in which the user can check the integrity without downloading the data.

Zheng et al. [10] the author has introduced the modified concepts of “provable data possession scheme” so that it works on dynamic data. It also incorporates the outsourcing concepts of the third-party auditor.

Shah et al. [11] proposed a protocol to verify the data periodically and help the user to retrieve the data in the correct formatted manner.

Cong et al. [12] proposed public suitability to the data stored in cloud services by entrusted an expertise entity so that the user can reduce the computational costs, while incurring the data.

Yu et al. [13] suggested a scheme of data integrity. The adversary will record a fraction of data and record the maximum shared data between different users.

Bowers et al. [14] presented a technique "HAIL—High availability and integrity layer for cloud storage," which uses the concept of a distributed cryptographic system with IP error-correcting code.

Malina et al. [15] examined the verification concepts for guaranteeing some security arrangements of data protection in cloud services. They perform using the "bilinear method" and give anonymous authentication benefits to the enlisted client. The user imperative personal attributes were shown without uncovering the attributes of clients.

Boyang et al. [16] proposed a technique "PANDA" in which the user can share the data with a group of users. This mechanism utilizes the concepts of proxy re-signature with user invocation.

Holod et al. [17] projected a method for providing authentication, data management services outside the cloud, and eliminate illegal retrieval. This method is based on the Openstack, and the service is dispersed in two different servers.

Chen et al. [18] suggested a methodology "to verify data," mutual authentication is performed between the client and server by using the concepts of generating the public key (PU) and private key (PR) pair using a pseudo random generator. Here, a metadata checklist is generated that is used at the time of the verification process.

A summary of the comparative study about the different schemes of data integrity and privacy of the above literature is provided in Table 2.

3.3 Data Availability

The DDoS attacks ["Distributed Denial of Service"] is a severe hazard in which the invader sets an immense flow of request toward the server-side, intending to deny the service for the genuine user. Assault consumes a large quantity of network bandwidth so that authoritative needs is discarded [19]. To detect the assault in the network, it requires a multi deploy detection system that provides immense accuracy toward the awful traffic. The DDoS attacks can be classified as h-Rate and l-Rate attacks. The h-Rate attack the targeted server by over flooding the traffic with a non-relevant request. The l-Rate attack differs from the h-Rate attack in which it affects the network by over flooding the path with a non-relevant request in a periodical manner (Fig. 3).

Table 2 Schemes and privacy method of data integrity

References	Year	Methodology	Description
Thomas et al	2006	• Blinded XoR • Pseudo random sequences	Using the pseudo random generator a sequence of parity bits is generated and XoR with the data so that it can stored securely in remote area
Ateniese et al	2007	• Merkle Hash Tree • RSA based Hash • Function • Diffie–Hellman • Homomorphic hash function is used	Probabilistic proof of possession is generated by using sampling random sets and sever use Single modular exponentiation
Zheng et al	2011	• Homomorphic computational • RSA • Modified MHT	Modified the concepts of PDP that works in dynamic data
A. Juels et al	2007	• Homomorphic Property • Computational Diffie-Hellman • Bilinear mapping	Explored a method "proofs of retrievability (PORs)". The scheme enables the prover to generate a concise proof report of a targeted file that the user is retrieving
Shah et al	2008	• Periodical verification • BLS based instantiation	Protocol support the auditing process periodical and retrieve the data of the in proper way
Yang et al	2012	• Data fragmentation technique • Bilinear mapping	It works on the multi owner scenario. It combines the cryptography method with the bilinearity property and also batch auditing is possible
Chen et al	2012	• Pseudo random • Permutations • Cauchy RS Encoding	Here the reed Solomon code based matrices is used so that we can modify the lower bandwidth by reducing the overhead in insertion and deletion operations
Yu et al	2015	• Fractional of data block • Homomorphic authenticator is used • Random masking	Adversary will record a fraction of data and computed the maximum shared data from by using the recorded data
Bowers et al	2009	• HAIL • Distributed cryptographic • IP-ECC	"*Hail: A High-Availability And Integrity Layer For Cloud Storage"* Data integrity is verified with the cryptographic P-P key pair and ECC concepts

(continued)

Table 2 (continued)

References	Year	Methodology	Description
Malina et al	2015	• Bilinear mapping • Anonymous authentication	The verification concepts for guaranteeing some security arrangement of data protection in cloud services
Wang et al	2015	• PANDA	PANDA is a Privacy preserving security solution that help us to access the cloud services privately from a group of user
Holod et al	2017	• OpenStack • Separate server	Data authentication, data management services outside the cloud, it also eliminate illegal retrieval
Y. Chen et al	2017	• Mutual Exclusion • Pseudo random sequence	Client and server build up mutual authentication is performed between the client and server by generating the public and private key pair using pseudo random generator

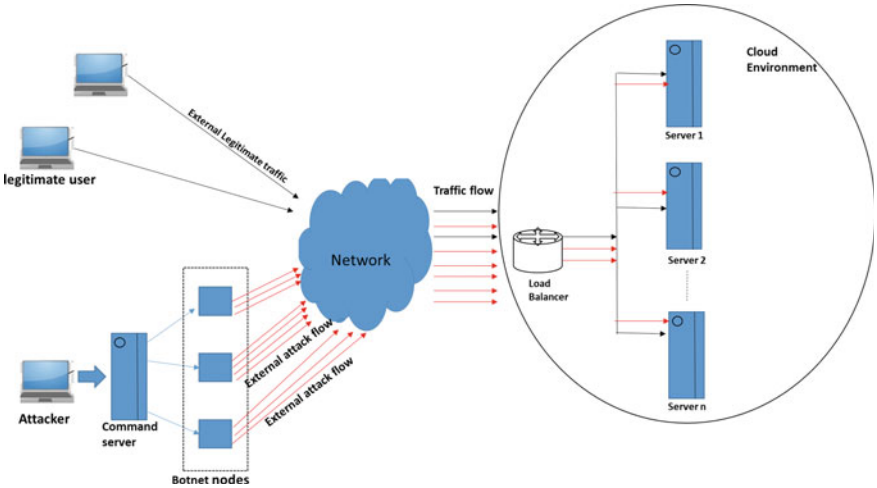
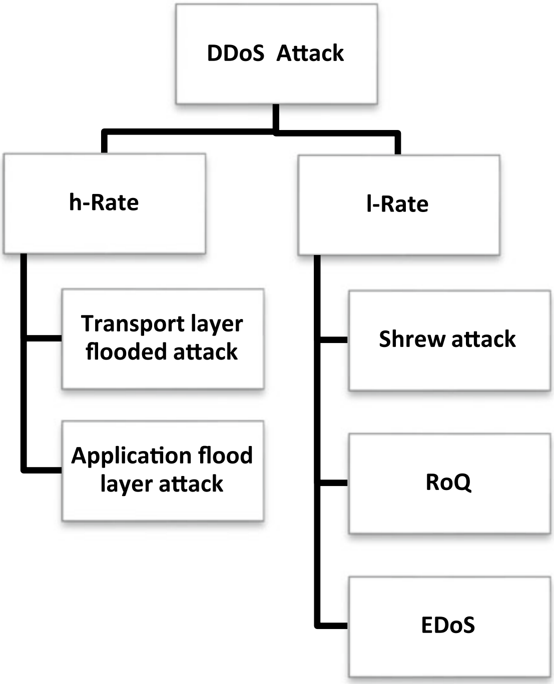


Fig. 3 DDoS scenario in cloud structure

3.4 Different Type of DDoS Attack

In the h-Rate DDoS attack, the assault flood the network traffic with a non-relevant request by either disrupting the service or using the user connection, dissipate the router capacity, bandwidth usage, etc. Such a typical problem in user connectivity is called transport layer flooding [20]. Example: TCP-SYN flood, ICMP flood,

Fig. 4 Different types of DDoS attack


DDoS Attack
h-Rate
l-Rate
Transport layer flooded attack
Application flood layer attack
Shrew attack
RoQ
EDoS

UDP flood [21, 22]. Another type of attack in h-Rate is application layer flooding performed by dissipating the server resources. Examples are HTTP flood, DNS flood, SMTP flood.

The l-Rate DDoS attack is sophisticated and shows the difficulty to detect because of its low-rate flow and protected behavior. This attack affects the QoS of the cloud services experienced by the user rather than stopping the services. The l-Rate attack is classified based on the attack length (AL), attack rate (AR), attack duration (AD) as shrew attack, EDoS attack, RoQ attack.

Osanaiye [23] proposed technique to detect the IP spoofing in the cloud by implementing the host-based fingerprinting. Here, all packet that comes to the server checks with the previous packet information stored in the database. 1

Jiao et al. [24] proposed a scheme of detecting DDoS attacks by considering the Baidu cloud computing platform dataset. This scheme includes the false alarm, extract the feature of the packet also uses a tree classifier to label between the normal and assault traffic flow.

Deka et al. [25] presented a new outline “self—similarity using Hurst parameter.” Its self-similarity features help in differentiating the traffic and provide low-cost analysis in the network. This statistical similarity is used to generate a low-rate detection model more efficiently. The model provides coupling concepts between the parameter and estimates the Hurst parameter.

Zhi et al. [26] presented a new approach against the "low-rate DDoS attack," in which the traffic is loaded with unwanted but similar pattern packets. So a mathematical model based on the concepts of queuing theory is proposed to find the strength and weakness of the session from which the data flow reaches the server.

Nazrul et al. [27] proposed a methodology for the statistical measurement "Feature score FFSc," in this multi-variable are considered for the analysis process. The result of the analysis helps to differentiate the traffic.

Shi et al. [28] the proposed method is an "Improved KNN" it used the Degree of gain and KNN to detect the attack in the SDN environment.

Mais et al. [29] recommended a "Data Mining Approach" to control and analyze the "dynamic priority assignment," the system allows free monitoring and early detection process.

Conti et al. [30] presented a scheme "Scale Inside-Out" which reduces the RUF to a minimal value. This scheme ensures resource availability by using collocated services.

Yang et al. [31] proposed an "entropy-based metrics" to detect the low-rate (LR) DDoS attack. Here, the popular "Kullback Leibler" approach is used as an IP trace back pattern to find the assault and owner. The method uses two metrics that is entropy and information distance from which the false rate of occurrence can be determined and reduces the low-rate attack.

Lee et al. [32] in the proposed method "Proactive Detection of DDoS attack," the intrusion detection system used in the concepts cluster analysis for detecting the assault traffic pattern from the normal traffic pattern.

Zhou et al. [33] proposed a "Detection mechanism using the exception of packet size" This system set a sample time, tolerance factor, and also the earliest time. After that the input packet will be compared with the previously calculated values, which shows the rise in the traffic and detects the attack.

Myo et al. [34] proposed a technique to detect the DDoS attack based on SDN. It uses the advanced support vector machine (ASVM) in which a multiclass classification method is used to detect the attack. It is evaluated by calculating false alarm rate and accuracy of each traffic.

Tan et al. [35] proposed an "EMD detection System," which is based on the calculation of dissimilarity using Earth Mover's Distance [EMD]. It uses cross-bin matching to evaluate the dissimilarity measurements and perform ten-fold cross validation to classify the legitimate and assault user traffic.

Summary on the comparative study about the different proposed technique of data availability and detection of DDoS attack of the above works of literature shown in Table 3.

4 Conclusion

Cloud computing support different types of service that can be used by the user to work faster and efficiently since it data is stored in a secure domain. The services

Table 3 Different method and detection strategies

References	Year	Methodology	Description
O. Osanaiye	2015	• Operating system based fingerprint • Nmap technique	It is a knowledge based approach in which fingerprint is used to detect the attack
Jiao et al	2017	• Feature extraction • Tree classifier	Data mining approach is applied to detect the TCP based attack
Deka et al	2016	• Self-similarity using Hurst parameter	Knowledge based approach which provide a low-rate analysis to detect the attack from the normal flow
Nazural et al	2016	• FFSc	Data mining approach in which the feature are collected based on some score to perform the multivariate analysis in the network
Mais et al	2017	• Data mining engine • Runtime behaviors analysis	Detect the DDoS attack using the concepts of data mining with the behavioral analysis
Conti et al	2017	• Scale Inside out	It is a statistical method with structure signature analysis. it minimize the resource utilization factor and increase the attack detection time
Yang et al	2011	• Entropy metric • Distance metric • Kullback leibler approach	Pattern based approach with IP trace back mechanism
Lu Zhou et al	2017	• Detection mechanism using the exception of packet size	Detecting the attack by comparing with the sample time and tolerance rate
Myo et al	2019	• Advanced SVM	Enhance the Support vector machine by incorporating multiclass classification method
Tan et al	2015	• Earth Mover's Distance • Ten-foldcross-validations	It is based on dissimilarity calculation with cross-bin matching of parameter and also validation of traffic is done by ten-fold techniques

and the storage provided by the cloud are affected by many security challenges. Cloud computing provides a scheme for proper authentication, encryption techniques, protecting the data from illegal modification. The defender tries to break some of these aspects and steal out the data or sensitive information about the user also tries to deny the service that the cloud provides to the user by false request. The paper elaborates on the main two security issues related to data integrity and data availability. It also provides a comparative study of existing solutions for handling the problem related to integrity and availability. Even though there are existing solutions in the environment, the cloud is still facing the problem related to the above two issues. So, it is necessary to improve and mitigate a new approach to prevent or detect the vulnerabilities in the cloud environment.

References

1. M. Madiajagan, M. Jog, Cloud computing: exploring security design approaches in infrastructure as a service, in *International Conference on Cloud Computing Technologies, Applications and Management (ICCCTAM)* (2012), pp. 156–159
2. V. Nirmala, R. Sivanandhan, R. Lakshmi, Data confidentiality and integrity verification using user authenticator scheme in cloud, in *Proceedings of the IEEE International Conference on Green High Performance Computing (ICGHPC)* (2013), pp. 1–5
3. H. Tianfield, Security issues in cloud computing. Systems, Man, and Cybernetics (SMC), in *Processing of the EEE International Conference on cyber security* (2012), pp. 1082–1089
4. V. Anne, J. Rao, R. Kurra, Enforcing the security within mobile devices using clouds and its infrastructure, in *Proceedings of the CSI Sixth International Conference on Software Engineering (CONSEG)* (2012), pp. 1–4
5. R. PatilMadhubala, Information assurance technology analysis center(iatac), in *Data and Analysis Center for Software (dacs)*, software security assurance, state-of-the art report(soar) (2007)
6. S. Ramgovind, M. Eloff, E. Smith, The management of security in cloud computing, in *Proceedings of Association for Information Security for South Africa (ISSA)* (2010), pp. 1–7
7. T. Schwarz, S.J. Ethan Miller, L. Store, Forget and Check: using algebraic signatures to check remotely administered storage, in *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06)*, page 12 (2006)
8. G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, D. Song, Provable data possession at untrusted stores, in *Proceedings of 14th ACM Conference Computer and Communications Security (CCS'07)* (2007), pp. 598–609
9. A. Juels, B.S. Kaliski, Proofs of retrievability for large files, in *Proceedings of the 14th ACM Conference on Computer and Communications Security* (2007), pp. 584–597
10. Q. Zheng, Fair, S. Xu, Fair and dynamic proofs of retrievability, in *Proceedings of 1st ACMM Conference on Data and Application Security and Privacy (CODASPY)* (2011), pp.237–248
11. M. Shah, R. Swaminathan, M. Baker, Privacy-preserving audit and extraction of digital contents, in *Cryptology ePrint Archive*, Report 2008/186 (2008), pp.1–21
12. C. Wang, K. Ren, W. Lou, J. Li, Toward publicly auditable secure cloud data storage services. IEEE Netw.: Mag. Global Internetworking **24**(4), 19–24 (2010)
13. Y. Jia, K. Ren, V. Varadharajan, Enabling cloud storage auditing with key-exposure resistance. IEEE Trans. Information Forensics Security **10**(6), 1167–1179 (2015)
14. K.D. Bowers, A. Juels, A. Opera, Hail: a high-availability and integrity layer for cloud storage, in *ACM Conference Computer and Communications Security* (2009), pp. 187–198
15. L. Malina, J. Hajny, P. Dzurenda, V. Zeman, Privacy preserving security solution for cloud services. J. Appl. Res. Technol. **13**(1), 20–31 (2015)
16. B. Wang, B. Li, H. Li, PANDA: public auditing for shared data with efficient user revocation in the cloud. IEEE Trans. Services Comput. **8**(1), 92–106 (2015)
17. A.N. Rukavitsyn, K.A. Borisenko, I.I. Holod, A.V. Shorov, The method of ensuring confidentiality and integrity data in cloud computing, in *2017 IEEE International Conference on Soft Computing and Measurements (SCM)* (2017), pp. 272–274
18. Y. Chen, L. Li, Z. Chen, An approach to verifying data integrity for cloud storage, in *13th International Conference on Computational Intelligence and Security (CIS)* (2015), pp. 582–585

19. J. Quan, A cloud enabled DDoS, in dependable system and network, in *Proceedings of IEEE International Conference on Cloud Computing* (2014), pp. 264–275
20. J. Mirkovic, P. Reiher, A taxonomy of DDoS attack and DDoS defense mechanisms. ACM SIGCOMM Comput. Commun. Rev. **34**(2), 39–53 (2004)
21. B.B. Gupta, O.P. Badve, Taxonomy of DoS and DDoS attacks and desirable defense mechanism in a cloud computing environment. Neural Comput. Appl. **28**(12), 3655–3682 (2017)
22. K. Kalkan, G. Gur, F. Alagoz, Filtering-based defense mechanisms against DDoS attacks: A survey. IEEE Syst. J. **11**(4), 2761–2773 (2017)
23. O.A. Osanaiye, Short Paper: IP spoofing detection for preventing DDoS attack in Cloud Computing, in *18th Int. Conf. Intell. Next Gener. Networks* (2015), pp. 139–141
24. J. Jiao et al., Detecting TCP-based DDoS attacks in Baidu cloud computing data centers, in *Proc. IEEE Symp. Reliab. Distrib. Syst.*, vol. 2017-Septe (2017), pp. 256–258
25. R.K. Deka, D.K. Bhattacharyya, Self-similarity based DDoS attack detection using Hurst parameter. Secur. Commun. Netw. **9**(17), 4468–4481 (2016)
26. Z. Li, H. Jin, D. Zou, B. Yuan, Exploring new opportunities to defeat low-rate DDoS attack in container-based cloud environment. IEEE Trans. Parallel Distrib. Syst. **31**(3), 695–706 (2020)
27. N. Hoque, D.K. Bhattacharyya, J.K. Kalita, A novel measure for low-rate and high-rate DDoS attack detection using multivariate data analysis, in *8th Int. Conf. Commun. Syst. Networks, COMSNETS*, no. 1 (2016), pp. 1–2
28. S. Dong, M. Sarem, DDoS attack detection method based on improved KNN with the degree of DDoS attack in software-defined networks. IEEE Access **8**, 5039–5048 (2020)
29. M. Nijim, H. Albataineh, M. Khan, D. Rao, FastDetict: a data mining engine for predecting and preventing DDoS attacks, in *IEEE Int. Symp. Technol. Homel. Secur. HST* (2017), pp. 1–5
30. G. Somani, M.S. Gaur, D. Sanghi, M. Conti, M. Rajarajan, Scale inside-out: rapid mitigation of cloud DDoS attacks. IEEE Trans. Dependable Secur. Comput. **15**(6), 959–973 (2018)
31. Y. Xiang, K. Li, W. Zhou, Low-rate DDoS attacks detection and traceback by using new information metrics. IEEE Trans. Inf. Forensics Secur. **6**(2), 426–437 (2011)
32. K. Lee, J. Kim, K.H. Kwon, Y. Han, S. Kim, DDoS attack detection method using cluster analysis. Expert Syst. Appl. **34**(3), 1659–1665 (2008)
33. Z. Lu, M. Liao, C. Yuan, H. Zhang, Low-rate DDoS attack detection using expectation of packet size. Hindawi Secur. Commun. Netw. **10**, 14 (2017)
34. M.O. Myo , S. Kamolphiwong, T. Kamolphiwong , S. Vasupongayya, Advanced Support Vector Machine- (ASVM-) based detection for distributed Denial of Service (DDoS) attack on Software Defined Networking (SDN). J. Comput. Netw. Commun. 2019, Article ID 8012568, 12 pages (2019)
35. Z. Tan, A. Jdagni, X. Hez, P. Nanda, R.P. Liu, J. Hu, Detection of denial-of-service attacks based on computer vision techniques. IEEE Trans. Comput. **64**(9), 2519–2533 (2015)

Amrutha Muralidharan Nair is currently pursuing Ph.D. in Computer Science and Engineering at Karpagam Academy of Higher Education, Coimbatore. Her research area is cloud computing. Her work focuses on the security and availability of cloud services in the network.

Dr.R. Santhosh received his M.E degree in Software Engineering from Sri Ramakrishna Engineering College in 2009, M.B.A in Education Management from Alagappa University in 2011 and Ph.D. in Computer Science and Engineering from Karpagam Academy of Higher Education in 2016. He is currently working as an Associate Professor in the department of Computer Science and Engineering, Faculty of Engineering at Karpagam Academy of Higher Education. His current research interests include Cloud Computing, Distributed and Parallel Computing and Computer Networks.

Analysing the Impact of Security Attributes in Fog-IoT Environment Using AHP Approach

Richa Verma and Shalini Chandra

Abstract The concept of cloud computing has made the sharing and processing of bulk of data generated over Internet an easy job. Its major drawback lies with the fact that it suffers from high network latency as Internet-enabled devices have flooded the web enormously. Hence, for latency-sensitive applications, a middle-layer called fog layer has been introduced that has minimal processing capabilities. On one hand, this layer helps in overcoming the latency issue; but, on the other hand, it presents its unique issues because of it being closer to the end-devices. Fog scenario poses a serious security concern about the sensitive data being processed by it. Hence, the authors have unveiled a hierarchy of various Fog-IoT security factors and sub-factors and have ranked them using analytical hierarchy process (AHP).

Keywords Internet of Things (IoT) · Cloud computing · Fog computing · IoT security · Fog–IoT security · MCDM · AHP

1 Introduction

The Internet of Thing (IoT) paradigm is basically the collection of self-configuring nodes that are interconnected to make a dynamic network that greases the transmission of data [1]. IoT has improved the life of masses by improving interaction between people and smart infrastructures and services [2]. With this enormous increase in demand of IoT devices, the need for timely response has also elevated. Also, the IoT devices are vulnerable to cyberattacks due to their close proximity with the end users. Further, for time-sensitive application, the response time for the generated requests is the key. For such applications, cloud on its own could not provide the response in stipulated amount of time. This brings in the need for some early processing layer which is named as Fog Layer/ Fog Computing Layer by CISCO [3]. Basically, fog computing utilizes the capabilities of all the resources that are available at the networks' edge.

R. Verma (✉) · S. Chandra
Department of Computer Science, BBA University, (A Central University), Lucknow, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_38

Fog computing paradigm delivers the processing and computation capabilities proximate to the data originator, i.e., ground devices. The fog nodes process the large volume of data hence, the have a surfiet amount of data. Due to its architectural placement and availability of the copious amount of data, fog layer is vulnerable to various attacks and possess some major security concerns [4]. Different sub-factors/ attributes combine to ensure security of scenario. To manage security in an efficient manner all these sub-attributes must be considered. The order in which these should be handled also plays an integral role [5]. In the light of above discussion, the author has proposed a hierarchy of attributes and their sub attributes. Further, the authors have employed analytical hierarchy process (AHP) methodology which is used to deal with the problems that fall in the category of multi-criteria decision-making (MCDM). AHP because of its simple implementation, by far is the most prominent and widely used methodology [30].

Further, the article organization is as follows: Sect. 2 is about the related work. In Sect. 3, the detailed description of identified security factors and sub-factors of hierarchy is given. Later, Sect. 4 completes AHP methodology is explained. Section 5 deals with the implementation of the given methodology in Fog-IoT scenario. Finally, the paper is concluded in Sect. 6 with conclusion and future directions.

2 Related Work

Substantial amount of work can be seen while defining the need of security in Fog-IoT environment. The introduction of the layer as a mid-layer between cloud and ground devices has made it even more vulnerable towards attacks. A lot of efforts from various researchers can be seen in this regard, but still very less has been achieved. Ahmed et al. [6] have given various aspects of security in the fog environment. The authors have brought up the role of service-level agreement (SLA) while considering security and have also presented a comparative study of authentication privacy measures and authorization. Kumar et al. [7] have given the study on the security threats that are prevalent in the said environment. The authors have proposed a theoretical technique by considering the issues related to location privacy and confidentiality. Rauf et al. [8] have addressed the necessity of security and privacy in IoT. The authors have taken the architecture in to account and have explained different attacks that persist at each level. They have mentioned that this activity will help in providing the appropriate counter for the presented attacks and will, thus, increase the acceptability among the masses.

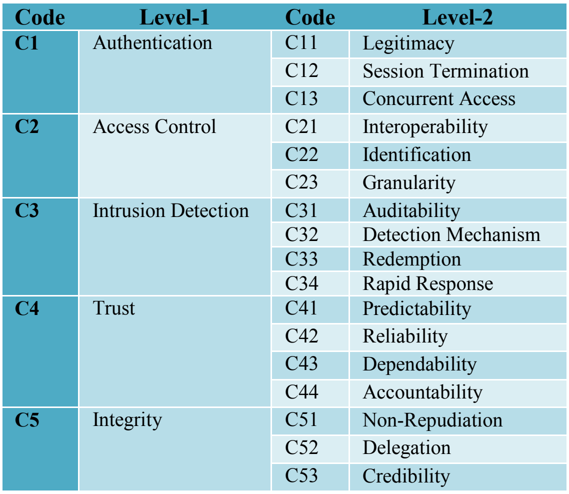
Various researchers have also employed MCDM techniques in this environment. Besset et al. [9] have stressed upon the way of selection of selection of security solution with respect to the need and demand of fog and mobile edge services. In this regard, the authors have employed Neutrosophic PROMETHEE (NPROMTHEE); a hybrid MCDM approach and security service selection was done accordingly. Geerish et al. [10] have also given that the cloud selection can be solved with MCDM approach and have employed IIVIFS-WASPAS as a hybrid approach towards the selection of trustworthy cloud service provider. Suma [31] has also used a hybrid deep fuzzy hashing algorithm for retrieving information in distributed cloud environment.

3 Fog–Internet of Thing Security

With the migration of sheer volume of Internet users towards IoT and connected smart devices, the urge of defining security in the said scenario has also elevated. Security is something that goes hand-in-hand along with the development of the application. It cannot be handled efficiently and effectively once the product is ready for implementation [11]. Therefore, a thorough study right from the beginning and true understanding of security helps better in catering it. In Fog-IoT paradigm, the collaborative interaction between fog and heterogeneous smart devices makes security handling very difficult. Keeping this fact in mind, there is an exigent need to address security at this level.

Security is composed of different sub-attributes; therefore, dealing these security attributes will help in providing a better grip over it. In the light of above fact, the researcher has defined the security along with its attributes and has made a hierarchy out of them. Figure 1 shows a two-level hierarchy in which the parent security attributes are at level-1, whereas their attributes are at level-2.

Fig. 1 Hierarchy of Fog-IoT security factors



Code	Level-1	Code	Level-2
C1	Authentication	C11	Legitimacy
		C12	Session Termination
		C13	Concurrent Access
C2	Access Control	C21	Interoperability
		C22	Identification
		C23	Granularity
C3	Intrusion Detection	C31	Auditability
		C32	Detection Mechanism
		C33	Redemption
		C34	Rapid Response
C4	Trust	C41	Predictability
		C42	Reliability
		C43	Dependability
		C44	Accountability
C5	Integrity	C51	Non-Repudiation
		C52	Delegation
		C53	Credibility

3.1 Authentication (C1)

With the establishment of fog layer at the middle of conventional cloud-IoT architecture, the need for authentication has also come into the picture. Different researchers [11, 12] have also mentioned the need of considering authentication in fog environment. Different sub-factors of authentication are mentioned below.

Legitimacy (C11)—Legitimacy with respect to authentication is defined as the virtue by which the user is given certain privileges in which the actions performed under that prerogative should have acceptability towards the assigned user [13].

Session Termination (C12)—Session termination plays a major role in ensuring authentication as improper session termination can provide entry doors to someone with malicious intent [14].

Concurrent Access (C13)—It has its vital role in terms of authentication as two concurrent users working with different access duties can exploit the system if their boundaries are not managed effectively [15].

3.2 Access Control (C2)

In Fog-IoT environment, due to the mobility of the nodes, there is less level of reliance among them, thus, keeping this in mind defining access bounds of the users are of high importance [16]. The author has defined the sub-factor of access control as-

Interoperability (C21)—In terms of access control, it is defined as the ability by which the user adjusts in the said environment even in the case of migration and upgradation [17].

Identification (C22)—The access footprints of the user have also a significant impact in terms of security. The monitoring and recognition of uses access help in user tracking in case of any kind of attack [16].

Granularity (C23)—The access to the smallest porting of data is governed under this. Maintenance of track and ensuring that only specific uses have access to them [18].

3.3 Intrusion Detection (C3)

Intrusion detection mechanism is basically adopted to trace the obstructive behaviour of the user. In terms of the said environment, this mechanism is of higher importance as the fog nodes can emit such behaviour by keeping track to login information and other log files [19]. The sub-factors are as follows-

Auditability (C31)—With regards to intrusion detection, auditability means keeping the track of the traces of the user activity so as to reveal the malicious intent of the user [20].

Detection Mechanism (C32)—This factor is very important in terms of intrusion detection. The real-time detection can help in better and quick response towards the attack [21].

Redemption (C33)—In this factor, it is said that if a system gets under attack scenario it should recover back and come back to its actual state in certain time frame [19].

Rapid Response (C34)—By rapid response, it is meant that the system should generate a quick response as an answer to the attack as this can lead to comparatively less data loss [22].

3.4 Trust (C4)

Trust plays a major role while dealing with the security of a system. In distributed environment maintain trust between the two communicating parties is of prime importance. The approaches that were used in cloud scenario cannot be directly deployed in fog scenario due the heterogeneous behaviour of the entities involved in communication [23]. Different sub-factors of trust are defined as follows-

Predictability (C41)—With regard to trust, predictability means estimating the behaviour of the node from the past communication and then anticipating any kind of work in future [24].

Reliability (C42)—In terms of trust, reliability means the probability to which the behaviour of node will be in consonance with the expectation [25].

Dependability (C43)—Dependability actually measures that behaviour of the node to which the communication process can rely on [25].

Accountability (C44)—It basically deals with the individual nature of the node. Under this sub-factor, every node is responsible for its behaviour and should stick to it in any circumstances [25].

3.5 Integrity (C5)

In Fog–IoT environment, integrity points to the prevention of any kind of tampering to the data that is at fog servers or the data that is movement at the time of communication. In IoT environment, the enormous amount of data is created and transmitted at very high speed, thus, handling the correctness of the data emerges as an important aspect [26]. The various sub-factors of integrity are as follows-

Non- Repudiation (C51)—Under this factor, it is assured that the node should not deny about its behaviour in due course of communication process [27].

Delegation (C52)—It is act of transfer of certain privileges to the one who is not obligated for them. This factor should be taken into strict consideration while dealing with the integrity factor [28].

Credibility (C53)—This basically deals with the truthfulness of the node in regard to the data that it is carrying and transmitting [29].

4 Methodology

For handling security in better way, both qualitative and quantitative measurement of the impact of the security factors must be considered. In Fog-IoT environment, only the qualitative estimation of the attributes is done. However, the quantitative is also required for handling security well. For such issue where quantitative assessment is required, MCDM proves to be the best fit. The ranking of the attributes can be done with the help of MCDM approach. Among the different MCDM techniques, the author has employed AHP. Due to simplicity in execution and efficiency in terms of result, AHP is most widely and frequently used techniques for ranking of the factors. Considering this, the author has employed AHP for ranking of the attributes in this research endeavour. In this section, the step-by-step methodology for AHP is depicted-

Step 1 The responses that have been collected from the industry personnel's and academicians with the help of questionnaire were linguistic in nature. So, they are converted into numerical values with the help of the linguistic scale presented in Table 1 The converted numerical responses are collected, and pair-wise comparison matrix is formed.

Step 2 In this step, the aggregated pair-wise comparison matrix taken, and normalization is performed by taking summation along the row and then dividing each element with the summed value (xij). The formulas are depicted below-

$$\sum_{i=1}^{n} xij = r1j + r2j + r3j \ldots rnj \tag{1}$$

Table 1 Linguistic scale values for AHP

Linguistic scale	Numerical value
Absolutely high (AH)	9
Very high (VH)	7
High (H)	5
Medium high (MH)	3
Exactly equal(EE)	1
Medium low (ML)	1/3
Low (L)	1/5
Very low (VL)	1/7
Absolutely low (AL)	1/9

Table 2 Table of random index (RI)

N	1	2	3	4	5	6	7	8	9	10
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

$$\left(r_{ij}\right) = \sum\nolimits_{i,j=1}^{n} r_{\mathrm{ij}}^{k}/xij \tag{2}$$

(n = number of experts).

Step 3 For calculating criteria weights summation of the elements is taken in row-wise manner and then divided by the "n"

$$\sum_{j=1}^{n} rij = r1i + r2i + r3i \ldots rni/n \tag{3}$$

Step 4 This step deals with the computation of consistency of pair-wise matrix. This is done by multiplying the computed criteria weights with each element of non-normalized pair-wise comparison matrix with respect to column.

Step 5 Then, the summation of all the elements is taken in row-wise manner to compute the weighted sum.

Step 6 The ratio of weighted sum and criteria weights is computed.

Step 7 The value of $\lambda_{\max}$ is calculated by taking the average of the ratio values computed in step 6.

Step 8 In this step, the consistency index (CI) value is calculated with the help of given formula

$$\mathrm{CI} = \frac{(\lambda_{\max} - n)}{n - 1} \tag{4}$$

Step 9 Finally, the value of consistency ratio (CR) is computed by dividing the CI value calculated in step 8 by the RI value taken from Table 2.

$$\mathrm{CR} = \frac{CI}{RI} \tag{5}$$

According to Saaty [30], if CR value is less than 0.10, the pair-wise comparison matrix is consistent.

5 Implementation of the Methodology

In this section, the researchers have implemented the methodology that is given in Sect. 4. The questioner was filled by thirty-two experts (Academicians, Researchers, Industry experts, etc.), and the data collected from them was analysed and ranked with the help of AHP methodology. Table 3 shows the linguistic response of Expert "X.". These linguistic responses are then changed into numerical values by using Table 1. Further, Table 4 shows the normalized pairwise comparison matrix for level -1 criteria (as in Fig. 1). Finally, the final weights and ranking of the criteria of both the levels are shown from Tables 5, 6, 7, 8, 9 and 10. The consistency of the obtained matrices was also computed, and they are found to be consistent.

After implementation of the methodology, it was found that at level-1 the "Trust" is the factor that has obtained the first rank and is highest in priority amongst different sub-factors. For Level-2, Legitimacy being on top for main factor Authentication,

Table 3 Linguistic pair-wise comparison matrix of expert "X"

Expert "X"	C1	C2	C3	C4	C5
Authentication (C1)	EE	MH	H	L	H
Access Control (C2)	ML	EE	VL	EE	ML
Intrusion Detection (C3)	L	VH	EE	VL	MH
Trust (C4)	H	EE	VH	EE	VH
Integrity (C5)	L	MH	ML	ML	EE

Table 4 Normalized pair-wise comparison matrix for level-1

	C1	C2	C3	C4	C5
C1	1	1.011	3.124	0.377	2.510
C2	0.989	1	3.284	0.606	2.393
C3	0.320	0.305	1	0.472	2.194
C4	2.650	1.649	2.119	1	4.572
C5	0.398	0.418	0.456	0.219	1

Table 5 Final weights of level 1 criteria

	Weights	Rank
Authentication(C1)	0.213	3
Access control(C2)	0.231	2
Intrusion detection(C3)	0.115	4
Trust(C4)	0.363	1
Integrity(C5)	0.075	5

Table 6 Final weights of authentication(C1)

	Weights	Rank
Legitimacy(C11)	0.412	1
Session termination(C12)	0.209	3
Concurrent access(C13)	0.379	2

Table 7 Final weights of access control (C2)

	Weights	Rank
Interoperability(C21)	0.396	2
Identification(C22)	0.436	1
Granularity(C23)	0.168	3

Table 8 Final weights of intrusion detection (C3)

	Weights	Rank
Auditability(C31)	0.141	4
Detection mechanism(C32)	0.395	1
Redemption(C33)	0.228	3
Rapid response(C34)	0.236	2

Table 9 Final weights of trust (C4)

	Weights	Rank
Predictability(C41)	0.148	4
Reliability(C42)	0.332	1
Dependability(C43)	0.204	3
Accountability(C44)	0.316	2

Table 10 Final weights of integrity (C5)

	Weights	Rank
Non-repudiation(C51)	0.462	1
Delegation(C52)	0.149	3
Credibility(C53)	0.389	2

Identification for Access Control, Detection mechanism for Intrusion Detection, Reliability for Trust and Non-Repudiation for Integrity.

6 Conclusion

In this research endeavour, the major focus of authors was to emphasise the need for security in Fog-IoT environment. In this regard, the researchers have also determined various security factors and their sub-factors by performing exhaustive survey (with the help of questionnaire) and consultation with experts. A hierarchy of such security factors, and their sub-factor is also presented. Further, for better management and qualitative and quantitative assessment of security of the said scenario, the prioritization of the attributes is done using AHP methodology. The respective rank of the factors at both the levels is also presented in the paper. The ranks generated at both the levels will help the researchers working in this area for defining the required effort to a particular factor for creating an overall impact on the scenario. In the near future, various hybrid techniques can also be employed for ranking the security factors for managing security in an efficient manner.

References

1. J. Kaur, A. Agrawal, R.A. Khan, Security issues in fog environment: a systematic literature review. Int. J. Wireless Inf. Networks **27**, 467–483 (2020)
2. R. Verma, S. Chandra, A systematic survey on fog steered IoT: architecture, prevalent threats and trust models. Int. J. Wireless Inform. Netw. 1–18 (2020)
3. F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things. in *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing* (ACM, 2012), pp. 13–16
4. R. Verma, S. Chandra, Security and privacy issues in fog driven IoT environment. Int. J. Comput. Sci. Eng. **7**(5), 367–370 (2019)
5. C. Kahraman, S.C. Onar, B. Oztaysi, Fuzzy multicriteria decision-making: a literature review. Int. J. Computat. Intell Syst. **8**(4), 637–666 (2015)
6. A. Ali, M. Ahmed, M. Imran, H.A. Khattak, Security and privacy issues in fog computing. Fog Comput.: Theory Practice 105–137 (2020)
7. K. Praveen, N. Zaidi, T. Choudhury, Fog computing: Common security issues and proposed countermeasures. in *IEEE International Conference System Modeling and Advancement in Research Trends* (SMART) (2016), pp. 123–129
8. A. Rauf, R.A. Shaikh, A. Shah, Security and privacy for IoT and fog computing paradigm. in *2018 15th Learning and Technology Conference* (L&T), February. (IEEE, 2018), pp. 96–101
9. A.-B. Mohamed, G. Manogaran, M. Mohamed, A neutrosophic theory based security approach for fog and mobile-edge computing. Comput. Netw. **157**, 122–132 (2019)
10. G. Obulaporam, N. Somu, K. Krithivasan, V.S. Shankar Sriram, IIVIFS-WASPAS: an integrated multi-criteria decision-making perspective for cloud service provider selection. Future Gener. Comput. Syst. **103**, 91–110
11. P. Hu, S. Dhelim, H. Ning, T. Qiu, Survey on fog com puting: architecture, key technologies, applications and open issues. J. Netw. Comput. Appl. **98**, 27–42 (2017)
12. J. Kaur, R. Verma, N.R. Alharbe, A. Agrawal, R.A. Khan, Importance of fog computing in healthcare 4.0. in *Fog Computing for Healthcare 4.0 Environments.* (Springer, Cham, 2021), pp. 79–101
13. G. Prosanta, LAAP: lightweight anonymous authentication protocol for D2D-Aided fog computing paradigm. Comput. Secur. **86**, 223–237 (2019)
14. S. Patonico, A. Braeken, K. Steenhaut, Identity-based and anonymous key agreement protocol for fog computing resistant in the CanettiKrawczyk security model. Wireless Netw. **219**, 1–13 (2019)
15. L. Loffi, C.M. Westphall, L.D.Grdtner, C.B. Westphall, Mutual authentication for IoT in the context of fog computing. in *International Conference on Communication Systems and Networks* (COMSNETS). (2019), pp. 367–374
16. P. Zhang, J.K. Liu, F.R. Yu, M. Sookhak, M.H. Au, X. Luo, A survey on access control in fog computing. IEEE Commun. Mag. **56**(2), 144–149 (2018)
17. M. Redowan, F.L. Koch, R. Buyya, Cloud-fog interoperability in IoT-enabled healthcare solutions. in *Proceedings of the 19th International Conference on Distributed Computing and Networking* (2018)
18. G. Rahman, C.W. Chuah, Fog computing, applications, security and challenges, review. Int. J. Eng. Technol. **7**(3), 1615–1621 (2018)

19. F. Hosseinpour, P. Vahdani Amoli, J. Plosila, T. Hmlinen, H. Tenhunen, An intrusion detection system for fog computing and IoT based logistic systems using a smart data approach. Int. J. Digital Content Technol. Appl. **10** (2016)
20. X. An, J. Su, L. X., F. Lin, Hypergraph clustering model-based association analysis of DDOS attacks in fog computing intrusion detection system. EURASIP J. Wireless Commun. Netw. **1**, 1—9 (2018)
21. X. Zhang, Y. Yuan, Z. Zhou, S. Li, L. Qi, D. Puthal, Intrusion detection and prevention in cloud, fog, and internet of things. Secur. Commun. Netw. (2019)
22. S. Prabavathy, K. Sundarakantham, S. Mercy Shalinie, Design of cognitive fog computing for intrusion detection in internet of things. J. Commun. Netw. **20.3**, 291–298 (2018)
23. Y. Ashkan et al, All one needs to know about fog computing and related edge computing paradigms (2018)
24. A. Alrawais, A. Alhothaily, C. Hu, X. Cheng, Fog computing for the internet of things: security and privacy issues. IEEE Internet Comput. **21**(2), 34–42 (2017)
25. Ybedokken, Tuva Selstad, Trust Management in Fog Computing. MS thesis. NTNU (2017)
26. A.A. Mutlag, M.K.A. Ghani, N.A. Arunkumar, M.A. Mohammed, O. Mohd, Enabling technologies for fog computing in healthcare IoT systems. Futur. Gener. Comput. Syst. **90**, 62–78 (2019)
27. N. Abbas, M. Asim, N. Tariq, T. Baker, S. Abbas, A mechanism for securing IoT-enabled applications at the fog layer. J. Sens. Actuator Netw. **8**(1), 16 (2019)
28. A.A. Alsaar, H.P. Pham, C.S. Hong, E.N. Huh, M. Aazam, An architecture of IoT service delegation and resource allocation based on collaboration between fog and cloud computing. Mobile Information Systems (2016)
29. S. Pei, M. Radovanovi, M. Ivanovi, C. Badica, M. Toi, O. Ikovi D. Bokovi, CAAVI-RICS model for analyzing the security of fog computing systems. in *International Symposium on Intelligent and Distributed Computing* (2019), pp. 23–34
30. T.L. Saaty, *The Analytic Hierarchy Process* (McGraw-Hill, New York, USA, 1980)
31. V. Suma, A novel Information retrieval system for distributed cloud using hybrid deep fuzzy hashing algorithm. JITDW **2**(03), 151–160 (2020)

A Deep Learning Approach to Invisible Watermarking for Copyright Protection

Tejas Jambhale and H. Abdul Gaffar

Abstract Watermarking has been an age-old process serving a multitude of purposes from copyright protection to maintaining confidentiality. Invisible watermarking is a subcategory of watermarking which involves the use of a digital signature embedded in the image to be protection in such a way that keeps the watermark imperceptible to the naked eye. The proposed technique embeds an invisible and ineffaceable watermark to any image, without any distortion or loss in image quality. This is done using autoencoder neural network which is trained for the process of embedding the watermark onto the image to be protected. This technique provides a lower loss after extraction and makes sure the watermark is invisible with minimal distortion to original image. Easy extraction of watermark is also ensured using the model specifically designed for extraction.

Keywords Deep learning · Encoder-decoder · Autoencoder · Neural network · Image processing · Watermarking

1 Introduction

Stock image archives like Shutterstock and Pexels, display obvious logos in images, leaving them unusable without their respective licensing rights. This is an example of visible watermarking. The obvious drawback is that this watermark reduces focus on the actual photo forcing the content provider to buy rights for it. Once the rights are acquired, there is no way for the image provider to tag it as its own. Though this is desirable when content needs to be shared by official sources only, it provides no mechanism for writing a permanent signature to an image that does not disturb content whilst simultaneously provide source information.

T. Jambhale (✉) · H. A. Gaffar
Vellore Institute of Technology, Vellore, India

H. A. Gaffar
e-mail: abdulgaffar@vit.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_39

Invisible watermarking is a technique used to fight against high-tech copyright and pirated methods [14, 15]. The proposed technique embeds an invisible and ineffaceable watermark to any image, without any distortion or loss in image quality. This enables photographers and social media influencers to keep track of their work and provide proof of ownership when images are illegally distributed online or in print. This model was specifically developed for this purpose and provides such creators with an opportunity to tag their work.

Instead of using a blatantly visible signature, an invisible mark is embedded into the image, such that the image quality is not reduced and proof of ownership can still be checked through easy extraction of that watermark [16].

Here, when you watermark a photo, the embedding process modifies individual pixels all over the image in such a way that the difference with the original image is unnoticeable. If a photo is modified after the watermark is embedded, some pixels will be altered, but not all. This makes sure that the watermark can always be extracted.

Digital images are prone to easily copying, modifying, and cropping [1]. Our proposed technique adds a digital signature to any digital image, making it easy to keep track and identify the source of the image without any loss to image quality. The watermark will be added such that to remove the invisible watermark, image modification will need to be performed at such a scale, that image will be rendered unusable by any entity.

In recent times, image watermarking using different deep learning techniques has started to see receive some attention [13]. The solutions presented till now are unable to achieve, blindness, robustness, and easy embedding and extraction simultaneously with desirable results. Actions like compressing, adding noise, rotating, skewing, and even screen capture have should have no effect on any technique. This makes it robust against attackers looking to alter the watermark.

The image degradation in invisible watermarking can be measured using Peak Signal to Noise Ratio (PSNR). Usually, a PSNR of 35db and above is considered acceptable but this level often depends on the context where it is being applied.

Another advantage of this approach is that this model can be directly extended to videos using the same autoencoder, providing good results. Video watermarking is different in that the watermark video is embedded in the original video in such a way that each frame of the original video is watermarked. This is done by processing video frames in batches and encoding them individually. Using autoencoders and training it on a video dataset, we can obtain high PSNR whilst also maintaining high video embedding capacity.

1.1 Problems with Existing Methods

Invisible watermarking is an example of active verification technology. Digital image watermarking methods have become a crucial approach to copyright protection schemes of visual data. It provides a non-intrusive way of protecting the image rights of creators. Nevertheless, vulnerability to geometric attacks, lack of robustness and inability to find a balance between high security and invisibility are prevalent problems in the domain of image watermarking [12].

Existing novel image watermarking techniques are often implemented using DCT and DFT transforms which perform block operations in the frequency domain of an image [2, 17]. These techniques use corresponding coefficients from these domains and embed the watermark. Finally, these embedded coefficients are extracted and transformed back to the original state during the extraction process [3–5]. These techniques provide good enough results with respect to invisibility and distortion—PSNR of ~ 39.47 dB [20]—but lag when compared to deep learning techniques in terms of both factors.

Modifications of DCT involve using the SVD algorithm as found in [14]. Here, the watermark and original image are transformed using SVD, and then the watermark is embedded into the blue channel of the original image.

With the recent advancements in neural networks and in general, the field of machine learning, deep learning and specifically generative adversarial networks have been implemented in the field of watermarking [6, 7, 10].

Steganographic GANs have been used with good results. In a GAN model, the generator model and the discriminator model are trained simultaneously. The generative network uses the information present in the embedding module to create images that appear natural despite being generated by the neural net and embeds the image in the mid-frequency region [8]. A discriminator works to differentiate between the original image and the watermarked image [9]. This technique is found to give high robustness and resistance to noise attacks. But, due to the fact that image is generated by a network, it results in lower image quality and higher distortion. The PSNR was found to be ~ 39.93 dB [8].

Even a custom network was proposed to specifically target invisible watermarking and improve robustness against image manipulation techniques. This was done by training the model on data containing geometric attacks and noisy images [11]; but again, this custom CNN sacrificed image distortion in favour of robustness which may be better but certainly not ideal.

Therefore, in this paper, we propose using auto encoder-decoder as an alternative to GANs. We look to solve the disadvantages faced by GANs by creating a more specialised network to solve the problem of invisible watermarking. Autoencoders learn abstractions in images and convert the watermark into lower-dimensional data [19]. Autoencoders compress the image input in a linear fashion, differentiating it from other encoder formats, which creates a bottleneck-type structure such that only important features are passed on. The hiding network looks to find suitable locations in the original image to hide the watermark without causing any distortions.

The proposed model increases complexity but brings a balance between robustness and efficiency, which is required for any real-life applications. We propose this autoencoder model to translate spatial images and complete watermark encoding.

2 Architecture

The algorithm involves two neural networks: A Hiding Net and Reveal Net. They follow an auto encoder-decoder format. The encoder reduces the image to be embedded, in this case, the watermark, into smaller dimensional data. This is done in such a way that loss is minimal. Subsequently, the reduced image is then hidden in the other image. As seen below, the input layers and output layers are equal in number for an autoencoder. In the proposed implementation, the model has two inputs: for original image and watermark. Therefore, there are two outputs corresponding to each input. Both the hide and Reveal Net models composed of 3 × 3, 4 × 4 and 5 × 5 layers concatenated in parallel. ReLU activation function is used with a kernel size of 3 on each layer (Fig. 1).

An autoencoder works such that it forces a compressed representation of the original input data by creating a bottleneck. It relies on the correlation between input features. The bottleneck is what differentiates it from other networks. The trained encoder embeds the watermark at parts of the original image which are least visible. The autoencoder network learns how to reconstruct the encoded data but makes sure not to overfit. The decoder works by learning the most important attributes required to create a low distortion image and generates a new image from such attributes. Noise is added to the output of watermarked image to avoid overfitting. The model learns its representation from the training set, so it is important to choose the right dataset and stop it from overfitting.

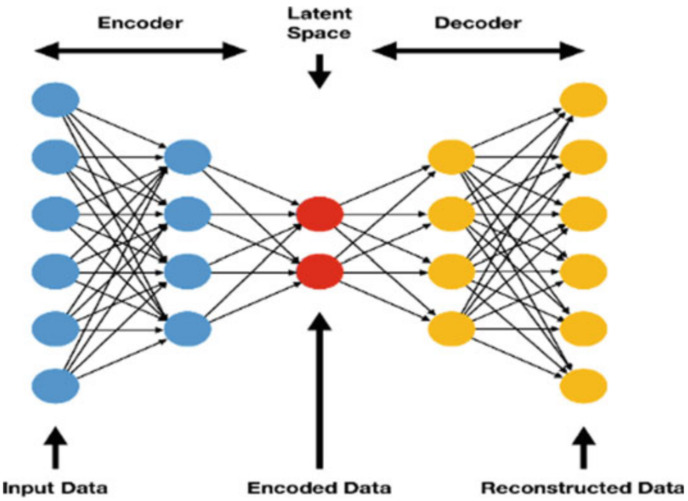


Fig. 1 Autoencoder structure [18]

2.1 Metrics

The following are the metrics which need to be minimised to define a good model. Both these metrics calculate the distortion and the amount of noise in the image.

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{1}$$

where MSE is mean squared error, *m* and *n* are image dimensions, and *I* and *K* are the images being compared with *I* (*i*, *j*) representing a pixel in image *I*. The MSE gives us the difference between two images by comparing pixel values. A high MSE which indicates two images have large deviations in individual pixel values. Although MSE is an adequate metric to compare two images and find distortion, the final value we arrive at is relative, and there is no “good value” to compare against. Another problem with MSE is that it depends on the image bit size as well. PSNR solves this problem by using the image bit to scale the final result from MSE. Therefore, we use the MSE value to calculate PSNR. PSNR is given as:

$$\text{PSNR} = 20.\log_{10}(\text{MAX}_I) - 10.\log_{10}(\text{MSE}) \tag{2}$$

where PSNR is the peak signal-to-noise ratio. MAX is the maximum pixel value of the image and MSE represents the mean squared error calculated in the previous equation. PSNR is calculated in decibels (dB). PSNR provides the ratio of the original image to noise in the images. PSNR is often used to measure distortion between two images with help from MSE. PSNR can be used to compare results of the proposed technique with other prevalent approaches to digital watermarking as it is widely used in the industry for comparing both images and videos.

Any watermarking technique must ensure a low MSE and, thus, a high PSNR to be a viable solution. Low PSNR can mean low image quality which would be counterproductive to the cause.

The MSE and the subsequent PSNR may be calculated for two sets of images: First, between the original image and the embedded watermarked image; second, between the original image and the final image after watermark removal.

3 Methodology

The encoder-decoder is both trained together for 44,000 images. The Hide net (encoder) is used to embed the watermark into the secret image. This is done by distributing the pixels value of the watermark over the RGB channels of the secret image. The autoencoder creates a bottleneck on the embedding process to encode data in lower dimensions. This watermarked image can now be freely distributed with

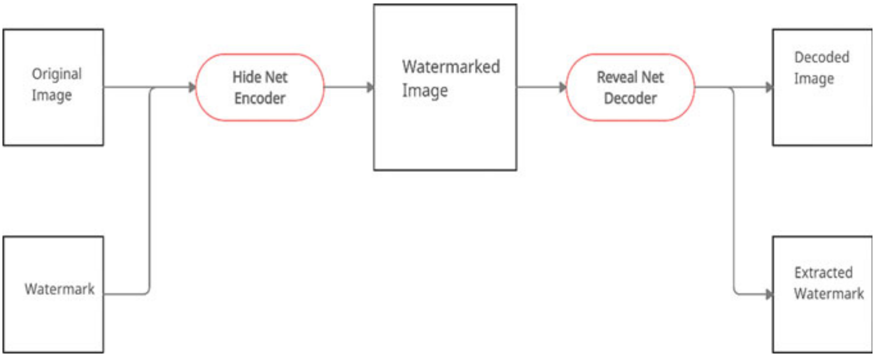


Fig. 2 System flow

this proof of ownership. When validation needs to be done, this watermarked image passes through the decoder (Reveal Net) to remove the watermark. Same procedure is followed for video watermarking as well.

All images are resized to 256*256 for testing purposes. Other image dimensions can be used, but the model needs to be trained accordingly.

The models are trained such that only the corresponding decoder which was trained simultaneously with the encoder can extract the watermark. Any other model or technique will not be able to give extract without substantial damage to image. This makes the system robust and resistant to attempts of tampering. The process for video watermarking is also the same (Fig. 2).

3.1 Proposed Steps

For watermarking.

1. Select the watermark and the secret image to be embedded with the watermark
2. Resize both to 256*256 dimensions
3. Using the trained encoder model, the watermark will be embedded into the image to create the container image
4. Compute the MSE between container and original image to calculate PSNR. This gives a measure of distortion after watermark is embedded
5. At receiver end to get the original image, we use the decoder neural network and remove the watermark
6. Additionally, we can compute PSNR between resultant image and original image, to check if the overall process of watermarking caused any loss to the original image

4 Results

Here we first embed a watermark onto an image and see the distortion caused between the original image and the embedded watermarked image. The Hide Net Encoder is used for this purpose. As seen above, the invisibility when watermark is embedded is very high. The difference can only be seen when we zoom in only. Otherwise, the difference in result image is nearly zero. The fourth column provides a diff of original and watermarked image. As seen, it is very minimal, and watermark embedding has not caused any visible deviation in the images. Given below are the metrics of the above process (Table 1) and (Fig. 3).

Similarly, below are the results of the decoding procedure, where we compare the original image with the final de-watermarked image (Fig. 4)

Above we have compared the original image with the image obtained after the removal of watermark. This is just to show that in case watermark is no longer required, it can be removed easily without any loss to image quality. The image diff seen in the third column for the decoding process has slightly become more prominent as compared to the hiding procedure. This can be especially seen in the first row. It can be inferred that the PSNR has slightly decreased but still is high enough as seen in the table below. Visibly, there is no difference to be found in the images being compared (Table 2).

The overall verification loss (MSE) was found to be ~ 1.84 during training process. After testing 120 images randomly taken which were not included in the training dataset, the average PSNR was found to be 45.57 dB which is well above the recommended values. More importantly, in all samples, shown the embedded watermark, are completely invisible to the naked eye. Both these factors make the proposed technique suitable for this application. As seen above, textual data can also be hidden. This allows the freedom to store metadata in the image such as name, location, date and time. This additional information can be helpful to identify image origins easier. Additionally, this algorithm successfully could encode videos as well trained on a different dataset. The average PSNR was ~ 31.43 dB tested on 30 videos which although is lower than that of image watermarking is still high when considering video size. The embedded video watermark was invisible in all tested samples, and though the distortion is visibly higher, the PSNR is still high. Videos take more time to execute due to frame division but follow the same Hide-Reveal Procedure.

This proposed technique was also tested on noise attacked images. The encoder and decoder continue to be able to encode and remove watermark even after a watermarked image suffers a noise attack. Although the MSE does increase considerably

Table 1 Results of encoding procedure

Row	MSE	PSNR (in dB)
Row 1	1.68	45.87
Row 2	1.84	45.48
Row 3	2.08	44.94

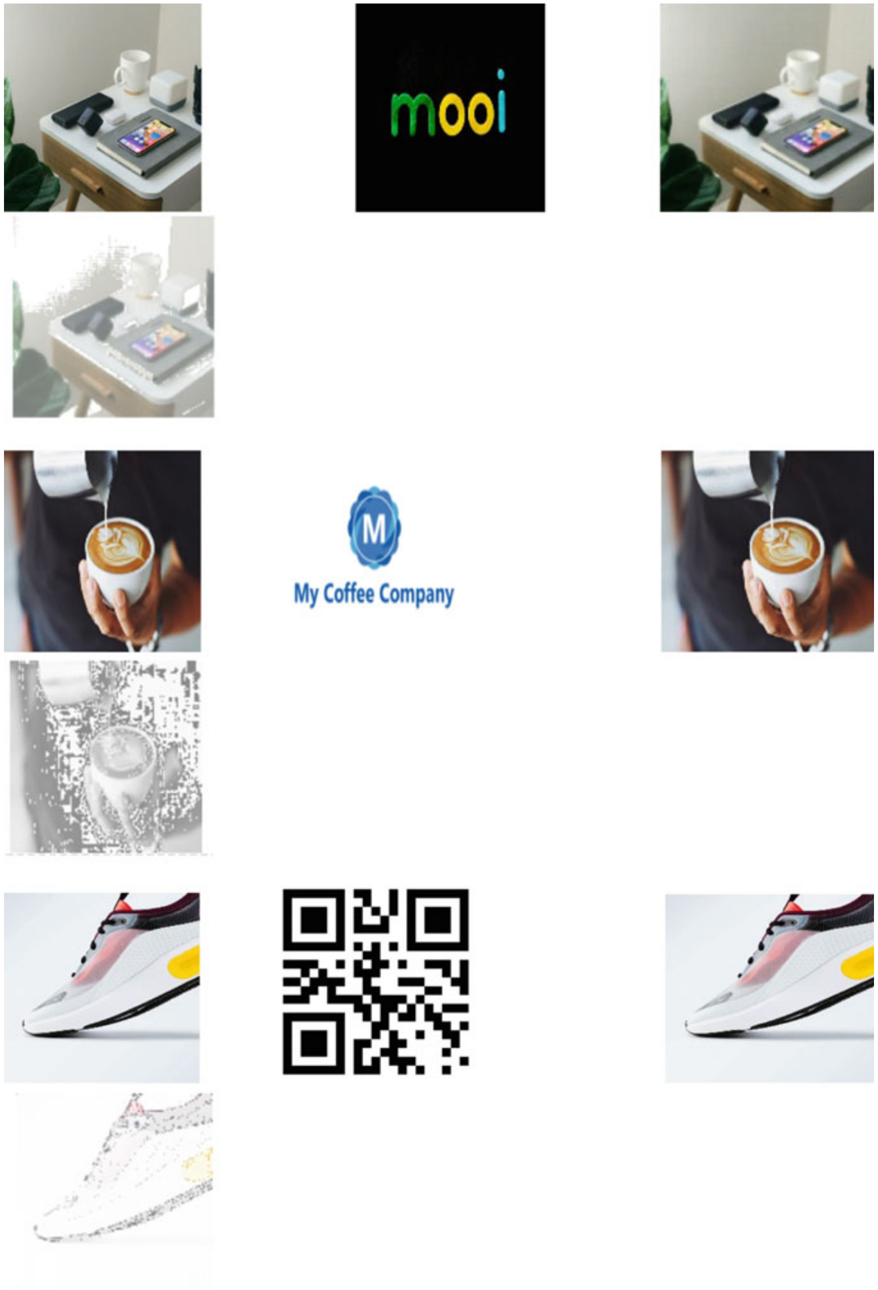


Fig. 3 Results of encoding process, first column: the original image, second column: watermark to be embedded, third column: the watermarked image (after embedding), fourth column: the image difference between original image and watermarked image

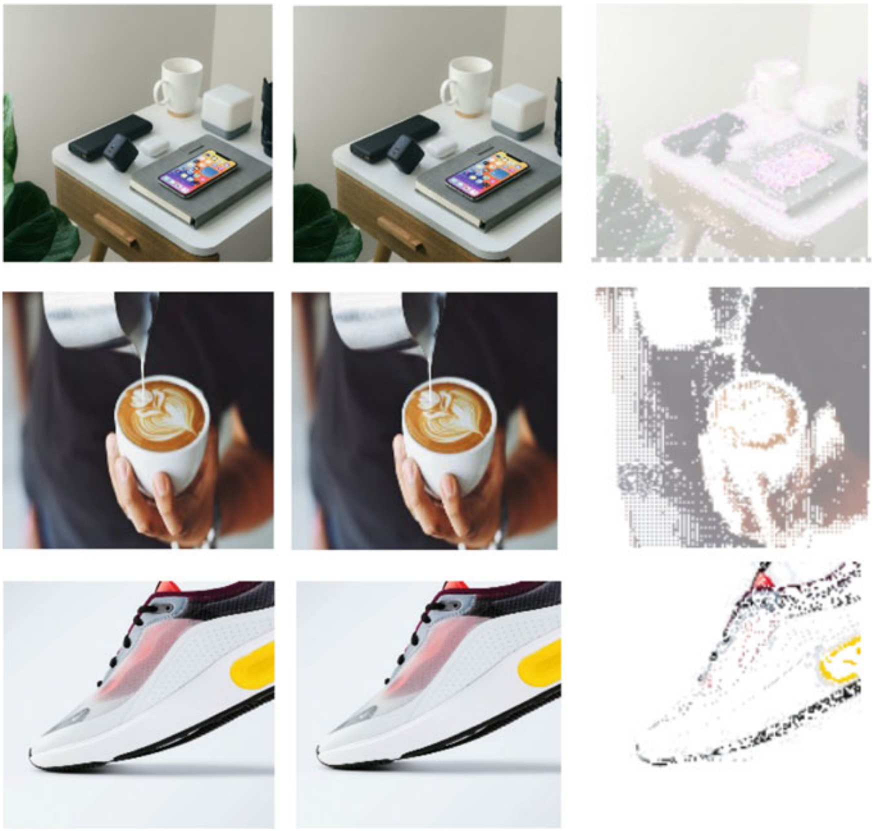


Fig. 4 Results of decoding, first column: the original image, second column: the image obtained after remove watermark from encoded image, third column: the image difference between original image and final image

Table 2 Results of decoding procedure

Row	MSE	PSNR (in dB)
Row 1	1.92	45.29
Row 2	2.09	44.92
Row 3	3.01	43.34

due to the noise attack, the entire process is not affected, and image quality is still high enough. Additionally, any attacker wishing to remove a watermarked image through noise attacks will be unable to do so without destroying the image itself.

5 Conclusion

In this paper, we proposed using autoencoder architecture to embed and extract watermarks for any digital image. The proposed technique embeds an invisible and ineffaceable watermark to the image, without any distortion or loss in image quality to maintain image copyright. This provides a high PSNR ~ 45.57 dB approach to invisible watermarking as compared to ~ 39 dB seen in GAN implementations [8]. Output images maintain high imperceptibility, low distortion and low noise. The results seen are better than other commonly found algorithms whilst also maintaining the imperceptibility of the encoding. The high PSNR and low MSE values make sure that this is a viable approach to real-life invisible watermarking. The autoencoder approach also allows video watermarking as well as providing good results. The implementation can also be modified to be applied in other domains providing similar results.

References

1. Z. Xin et al., An automated and robust image watermarking scheme based on deep ral networks (2020). arXiv preprint arXiv:2007.02460
2. W. Lu, J. Zhang, X. Zhao, W. Zhang, J. Huang, Secure robust JPEG steganography based on autoencoder with adaptive BCH encoding. IEEE Trans. Circuits Syst. Video Technol. 1–1 (2020). https://doi.org/10.1109/tcsvt.2020.3027843
3. S.D. Lin, C.-F. Chen, A robust DCT-based watermarking for copyright protection. IEEE Trans. Consum. Electron. **46**(3), 415–421 (2000)
4. N. Kashyap, G. Sinha, Image watermarking using 3-level discrete wavelet transform (DWT). Int. J. Modern Educ. Comput. Sci. **4**, 1–7 (2012)
5. X. Hu, S. Peng, W. Hwang, EMD revisited: a new understanding of the envelope and resolving the mode-mixing problem in am-fm signals. IEEE Trans. Signal Process. **60**(3), 1075–1086 (2012)
6. D. Volkhonskiy, I. Nazarov, B. Borisenko, E. Burnaev, Steganographic generative adversarial networks. in *Proceedings of the NIPS 2017 Workshop on Adversarial Training*, Long Beach, CA, USA, November (2017), pp. 201–208
7. D.P. Kingma, M. Welling, Auto-encoding variational bayes (2013). http://arxiv.org/abs/1312.6114
8. K. Hao, G. Feng, X. Zhang, Robust image watermarking based on generative adversarial network. China Commun. **17**(11), 131–140 (2020). https://doi.org/10.23919/JCC.2020.11.012
9. L. Xiang et al., Towards photo-realistic visible watermark removal with conditional generative adversarial networks. in *International Conference on Image and Graphics*. (Springer, Cham, 2019)
10. P. Wu, Y. Yang, X. Li, Stegnet: mega image steganography capacity with deep convolutional network. Future Internet **10**(6), 54 (2018)
11. Z. Chaoning et al., A brief survey on deep learning based data hiding. Steganograph. Watermark. (2021). arXiv preprint arXiv:2103.01607
12. J.-E. Lee, Y.-H. Seo, D.-W. Kim, Convolutional neural network-based digital image watermarking adaptive to the resolution of image and watermark. Appl. Sci. **10**(19), 6854 (2020)
13. Z. Jiren et al., Hidden: hiding data with deep networks. in *Proceedings of the European conference on computer vision (ECCV)* (2018)

14. D. Vaishnavi, T.S. Subashini, Robust and invisible image watermarking in RGB color space using SVD. Proc. Comput. Sci. **46**, 1770–1777 (2015)
15. N. Nikolaidis, I. Pitas, Robust image watermarking in the spatial domain. Signal Process. **66**(3), 385–403 (1998)
16. S. Pratibha, S. Swami, Digital image watermarking using 3 level discrete wavelet transform. in *Proceedings of the Conference on Advances in Communication and Control Systems-2013*. (Atlantis Press, 2013)
17. P. Pallavi, D.S. Bormane, DWT based invisible watermarking technique for digital images. Int. J. Eng. Adv. Technol. (IJEAT) (2013) ISSN 2249
18. S. Flores, Variational encoders are beautiful (2019). https://www.compthree.com/blog/autoencoder/
19. Z. Chong, R.C. Paffenroth, Anomaly detection with robust deep autoencoders. in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2017)
20. Ü. Arda, U. Guzin, U. Mustafa, DCT based image watermarking method with dynamic gain. (2015) pp. 550–554. https://doi.org/10.1109/TSP.2015.7296323

Web-Based Patient Centric Drive Health Care Monitoring System

K. Stella, S. Nithya Dharshni, M. Nivedha, and D. Nivedha

Abstract Recent increase in the demand for health care services is leading to a shift towards a more technology-centric strategy to collect and maintain a large scale patient data. Furthermore, the increase in dynamic population results in a scarcity of physical assets and techniques to store and scan the large-scale data. Moreover, the contemporary impacts of SARS COV have vested the demand for monitoring the outdoor patients. The shift in the needs have gained the research attention towards WBAN (wireless body area network) and wearable sensors. Nevertheless, there are no optimal research accomplishments related to this technology. Therefore, this paper proposes a health care monitoring system by utilizing the WBAN technology and then the collected data (from sensor nodes) will be sent to the public web page on real-time. Here, a web page resource framework has also been proposed for collecting and queuing the data in order to aid the research objectives. The adopted methodology has achieved an increased response time and compact physical properties with an increased efficiency.

Keywords Health care monitoring · Wireless body area network (WBAN) · Sensors · Sensor nodes · Web-framework

1 Introduction

Since the advent of digital technologies, several significant research ideas are available to collect and analyze the large-scale data. Especially in the health care sector, there is a constant need for monitoring the patient data in order to averse the repercussions in the rest of the mankind. There have been lots of research in the technological domain related to health care sector causing high reliability for data mining and extraction. As by the Mc Kinsey's Insights [1], the SARS COV has pushed the

K. Stella (✉) · S. N. Dharshni · M. Nivedha · D. Nivedha
Vel tech High tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India
e-mail: drkstella@velhightech.com

S. N. Dharshni
e-mail: vh10497.ece17@velhightech.com

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_40

telehealth sector forward leading to 76% consumer interest towards adopting the telehealth practise. The sudden spike shall remain constant as the industry is advancing with its innovation objectives. It is notable that approximately 250 billion dollars or 20% of all Medicare, Medicaid or commercial OP would be potentially virtualized therefore accelerating the surges in telehealth market (Fig. 1).

The statutory norms imposed by the government in regulating the telehealth sector has also surged after the impact of SARS COV. Therefore, a reliable system for monitoring the outpatients has become a major need for the contemporary generation. The increased need had accelerated the deployment of various health monitoring systems. However, they lack the archetypes of efficiency and real-time monitoring. Monitoring a large scale of outpatient data possess adverse technological challenges which has promoted the concept of WBAN (wireless body area network). The gain of WBAN's popularity has resulted in wearable sensors and monitoring systems. But there is less advancement in publishing the large-scale data on a web page to conduct analysis, research and for delicate outpatient monitoring purpose. Another

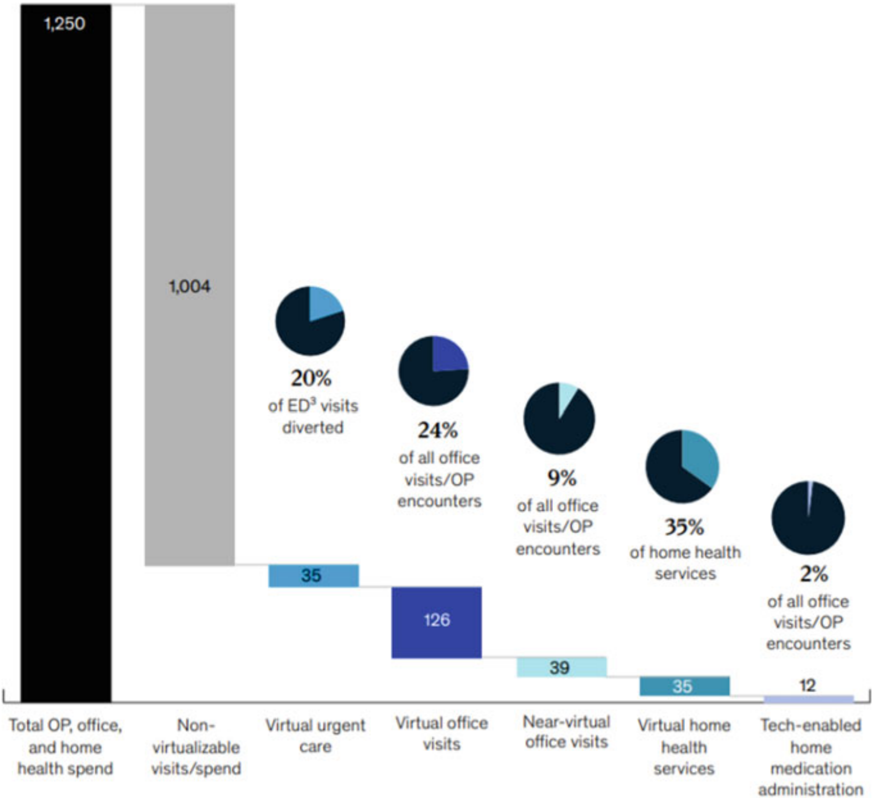


Fig. 1 Outpatient health spend at 2018 in US (*Source* Mc Kinsey & Company— Telehealth & Services)

major concern in developing such system is the vulnerability issues. Handling and transfer of such outpatient health data must be assured with advanced cryptographic algorithms. The other concerns include minimal response time, efficiency, confusion metrics and compactness of the device. These concerns are well-defined and objectified in this paper, and an efficient health monitoring system is proposed. The paper proposes a web-based health monitoring system by employing wearables that are integrated with the biosensors. The system adopts the WBAN technology to send the data to the website over the internet. The website is built with the page-enabled resource framework for aversion of the confusion metrics of the outpatient providing additional information.

2 Research Background

The recent proposal [2] "Secure IoT communications for smart health care monitoring system" proposes a novel approach of designing the health care system using the artificial neural network which is trained by fuzzy based interference system (FBIS). It also depends on the trust environment for patient data collection and utilizes the GSM to send the data to the server. It is implemented through Azure IoT platform. However, the system failed to provide possibilities of monitoring the outpatient, and moreover it is notable that it doesn't adopt any strong encryption codec which poses high risk for vulnerabilities. Moreover, the proposal is complex in design aspects. The paper [3] "Health care and Patient Monitoring Using IoT" proposes the employment of wearable biosensors, its architecture and implementational aspects for assessing the specific health data such as pulse ratio, plethysmogram and relative oxygen ratio of the patients. The collected data is standardized against correctness, network stability and range effectiveness. The major disadvantage of this system is, it is event-driven. So, it consumes more energy depleting the battery power quickly. Moreover, in this paper, the access control measures are not discussed. The paper [4] "IoT based health monitoring system" advocates the usage of portable network framework topology that is used to screen the patient's heart beat and temperature continuously. It is enabled with remote accessibility, therefore providing the patient health data on real time with all the medical professionals, so that medical help can be deployed well. Nevertheless, the paper disregarded the security patches that is needed to be encrypted for protecting the health data from all sort of vulnerabilities. As the system is unprotected, it is vulnerable to all sorts of attacks. The paper [5] proposes an improvisation model which analyses the heart rate from the ECG graph and substantially reduces the memory space required for analysing such. The analysis is done based on the improved PQA. But the paper didn't produce enough information on its simplified application. The paper [6] proposes a new methodology for monitoring the heart rate by using the portable device (photoplethysmography). The device is compact, and the data is provided on the smartphone on real-time. But however, the paper lacked showing the effective approach for displaying the data in an obvious manner. The paper [7] shows the study of Alzheimer's disease within the genomes of the family. It identifies the age-at-onset (AAO) of those genomes and presents a clear picture of those AAO outliers which also foresees a mechanistic discernment in the field of etiopathology. The paper [8] discusses the importance of the wearable technologies for assessing the Parkinson disease, heart rate monitoring and for other physiological parameters assessments. It focuses on various wearable technologies and provides a consolidated study therefore claiming that wearable technologies are the next generation technology. The paper [9] shows the study results of 18 persons tested using the plantar impedance measurement methodology. It also discusses the heat related impedance measurements in the legs. It is also done along with the ECG. It uses differential amplifier and CMRR for measuring the impedance. However, it lags the concepts of data extraction, analysis and real-time monitoring. The paper [10] proposes a novel technique of adopting optoplethysmograms which monitors the arterial pulsation of the fingers in a non-invasive manner, and it also monitors continuously without any intervals. The paper lacked adopting the networking and data harvesting methods for remote monitoring of the patient's health. Wearable devices [11] are used to manage mobility, bandwidth and usage of storage [12]. The health conditions are monitored and processed by IOT sensor network [13].

3 Proposed System

Figure 2 shows the proposed block diagram of health care monitoring system. It consists of biosensors, microcontroller, raspberry PI system, interrupt generator and remote monitoring system. The biosensors are integrated in wearables and designed ergonomically to remain in contact with the person's body. The biosensors are of many types and kinds.

Based on the outpatient's vital data for monitoring the respective biosensor is employed for distinct diseases. The threshold is set for detecting the signal abnormalities of the body using biosensor through time-driven optical lightning-based signal measurements. The biosensor is activated using the control signals initiated by the microcontroller. The interrupt generator is used to generate the interrupts for raspberry PI system. The controller can thus be configured anytime by this interrupt generator module. The raspberry PI acts as the main processor here, and it sends the collected data to the remote monitoring system. The raspberry PI employs the USB ACM protocol (Abstract Control Model) for effective transmission and reception of data over standard USB. It is enabled for raspberry controller interface. As mentioned, the ACM is preferred to USB because it can support duplex communication which is well supported by virtually different operating system (Fig. 3).

The remote monitoring system consists of the page enabled resource framework that reduces the confusion metrics of the patients allow the frequent changes in the data displayed. The confusion metrics includes the patient height, weight, age, name and other credentials; therefore, the aversion of errors in displaying the data is achieved. The LAMP-based dynamic web page is built to collect and display the data received from raspberry PI. The web page is designed to be dynamic to allow

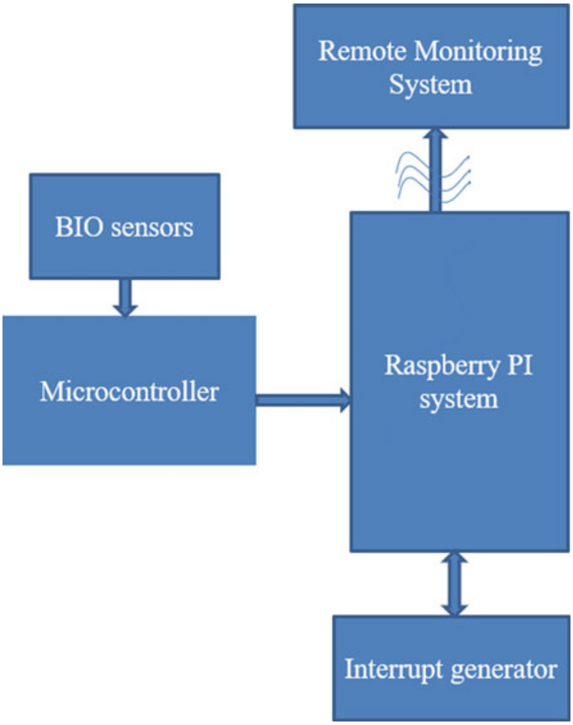


Fig. 2 Block diagram of proposed health care monitoring system

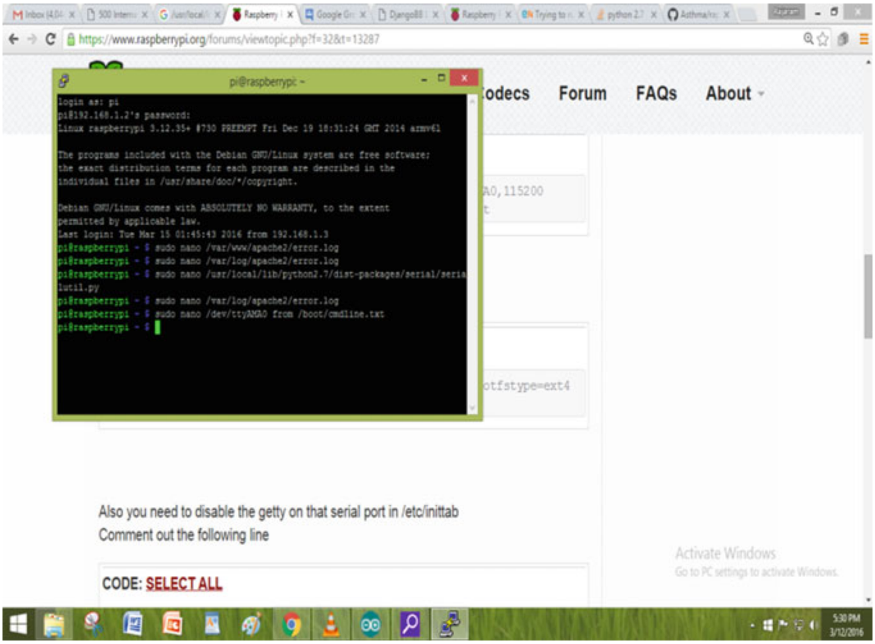
the frequent changes in the data displayed. For the security purpose, the TINY key based encryption codec is adopted for cypher stream generation. The adoption of TINY key reduces the latency rate highly.

4 Design and Implementation

The proposed system possesses various components such as biosensors, microcontroller, raspberry PI, interrupt generator and remote monitoring system. The design and implementational aspects of the above components are described as below.

4.1 Biosensors

Biosensors are transducers that is used to assess a biological analyte's concentration or presence including a biomolecule, a biological structure or a microorganism. It is made up of three parts: an analyte detection and signal generation portion, a signal transducer and a reader interface. There are many kinds of biosensors including


https://www.raspberrypi.org/forums/viewtopic.php?f=32&t=13287
pi@raspberrypi: ~
Codecs
Forum
FAQs
About
A0,115200
otfstype=ext4
Activate Windows
Go to PC settings to activate Windows.
5:30 PM
3/12/2016

```
login as: pi
pi@192.168.1.2's password:
Linux raspberrypi 3.12.35+ #730 PREEMPT Fri Dec 19 18:31:24 GMT 2014 armv6l

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Tue Mar 15 01:45:43 2016 from 192.168.1.3
pi@raspberrypi ~ $ sudo nano /var/www/apache2/error.log
pi@raspberrypi ~ $ sudo nano /var/log/apache2/error.log
pi@raspberrypi ~ $ sudo nano /usr/local/lib/python2.7/dist-packages/serial/seria
lutil.py
pi@raspberrypi ~ $ sudo nano /var/log/apache2/error.log
pi@raspberrypi ~ $ sudo nano /dev/ttyAMA0 from /boot/cmdline.txt
pi@raspberrypi ~ $
```

Also you need to disable the getty on that serial port in /etc/inittab
Comment out the following line

CODE: SELECT ALL

Fig. 3 Employment of USB ACM protocol

SPR based sensors, AuNP based sensors, FET based sensors, etc. The evolution of biosensors has now settled at nanochip integration in which the device is integrated by adopting very large-scale integration (VLSI) principles, fabricated using photolithography. The proposed system suggests using the nanoscale integration of such biosensors which can be fabricated on the wearables. The sensor should be in contact with the skin at all times. As different biosensors are used for different purposes and diagnosis, the biosensors are specified generally in this paper. These are connected with the microcontroller for deducing the data and processing the data.

4.2 *Microcontroller*

The microcontroller used in this system is Arduino nano. There is a total of 22 GPIO pins on the board. It has 14 optical pins and 8 analogue pins. There are also 6 PWM pins amongst the optical pins. It has a 16 MHz crystal oscillator and operates between 5 and 12 V. Serial Protocol, I2C Protocol and SPI Protocol are amongst the networking methods it supports. There is also a mini-USB pin for uploading the software. There is even a reset button.

It has different memory space integrated within it, and the followings are the memory component of Arduino nano.

- Flash memory–32 Kb.
- Boot loader flash memory of 2 kb.
- SRAM–8 kb.
- EEPROM memory–1 kb.

The microcontroller is used to activate the control signals and receive the analogue signals from the biosensors. The received signal is converted to digital data which is then fed to the raspberry PI controller to send the data to the web page.

4.3 Raspberry PI

The raspberry Pi 3 Model B is used in this system which is the earliest model of the third-generation raspberry Pi, and the specification of the model can be defined as Quad Core 1.2 GHz Broadcom BCM2837 64bit CPU. It possesses 1 GB RAM and has BCM43438 wireless LAN and Bluetooth Low Energy (BLE) on board. It also consists of 100 Base Ethernet, 40 GPIO pins, 4 USB and 2 ports, 4 Pole stereo output and composite video port. Along with that it also has full size HDMI, CSI camera port, DSI display port, Micro SD port and finally it has upgraded switched Micro USB power source range of 2.5A. The raspberry PI is established with LAN; therefore, it would send the data to the LAMP based web page.

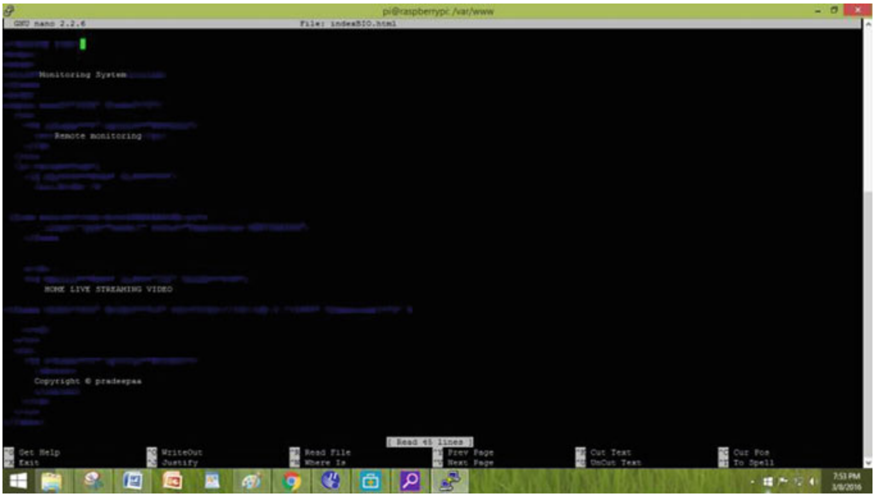
4.4 LAMP Based Web Page

The LAMP based web page is built for the easy configuration from the client-side browser. It is built by integrating the confusion metrics so that avoiding the confusion of patient's health care data.

Figures 4 and 5 show the screenshots of patient centric web page and CGI bin directory, respectively.

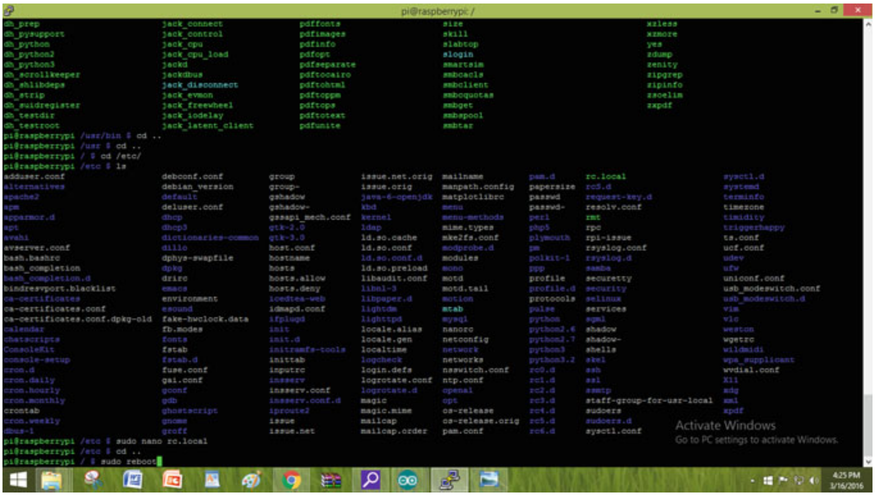
5 Future Enhancement

The patient centric model web page can be enhanced with good UI and UX designs. It can be made as a private web page; therefore, the clinic or a medical institution can possess the private accessibility of their own patient health as well as public accessibility of the patient health with authoritative access. It can be improved to be developed as a SaaS product of telehealth category. The features such as sharing the patient's portfolio and medical records can be implemented using the blockchain technology as it is more secure and reliable for such use cases. Along with that a crypto currency can be built for exchanging the right medical values in order to promote



```
  GNU nano 2.2.6                    File: indexBIO.html

      Monitoring System

         Remote monitoring

       HOME LIVE STREAMING VIDEO

     Copyright © pradeepaa

                                          [ Read 45 lines ]
^G Get Help      ^O WriteOut      ^R Read File     ^Y Prev Page     ^K Cut Text      ^C Cur Pos
^X Exit          ^J Justify       ^W Where Is      ^V Next Page     ^U UnCut Text    ^T To Spell
```

Fig. 4 Screenshot of HTML webpage



```
dh_prep                        jack_connect               pdffonts                        size                                xzless
dh_pysupport                   jack_control               pdfimages                       skill                               xzmore
dh_python                      jack_cpu                   pdfinfo                         slabtop                             yes
dh_python2                     jack_cpu_load              pdfopt                          slogin                              zdump
dh_python3                     jackd                      pdfseparate                     smartsim                            zenity
dh_scrollkeeper                jackdbus                   pdftocairo                      smbcacls                            zipgrep
dh_shlibdeps                   jack_disconnect            pdftohtml                       smbclient                           zipinfo
dh_strip                       jack_evmon                 pdftoppm                        smbcquotas                          zsoelim
dh_suidregister                jack_freewheel             pdftops                         smbget                              zxpdf
dh_testdir                     jack_iodelay               pdftotext                       smbspool
dh_testroot                    jack_latent_client         pdfunite                        smbtar
pi@raspberrypi /usr/bin $ cd ..
pi@raspberrypi /usr $ cd ..
pi@raspberrypi / $ cd /etc/
pi@raspberrypi /etc $ ls
adduser.conf                   debconf.conf         group             issue.net.orig mailname         pam.d      rc.local                   sysctl.d
alternatives                   debian_version       group-            issue.orig     manpath.config   papersize  rcS.d                      systemd
apache2                        default              gshadow           java-6-openjdk matplotlibrc     passwd     request-key.d              terminfo
apm                            deluser.conf         gshadow-          kbd            menu             passwd-    resolv.conf                timezone
apparmor.d                     dhcp                 gssapi_mech.conf  kernel         menu-methods     perl       rmt                        timidity
apt                            dhcp3                gtk-2.0           ldap           mime.types       php5       rpc                        triggerhappy
avahi                          dictionaries-common  gtk-3.0           ld.so.cache    mke2fs.conf      plymouth   rpi-issue                  ts.conf
avserver.conf                  dillo                host.conf         ld.so.conf     modprobe.d       pm         rsyslog.conf               ucf.conf
bash.bashrc                    dphys-swapfile       hostname          ld.so.conf.d   modules          polkit-1   rsyslog.d                  udev
bash_completion                dpkg                 hosts             ld.so.preload  mono             ppp        samba                      ufw
bash_completion.d              drirc                hosts.allow       libaudit.conf  motd             profile    securetty                  uniconf.conf
bindresvport.blacklist         emacs                hosts.deny        libnl-3        motd.tail        profile.d  security                   usb_modeswitch.conf
ca-certificates                environment          icedtea-web       libpaper.d     motion           protocols  selinux                    usb_modeswitch.d
ca-certificates.conf           esound               idmapd.conf       lightdm        mtab             pulse      services                   vim
ca-certificates.conf.dpkg-old  fake-hwclock.data    ifplugd           lighttpd       mysql            python     sgml                       vlc
calendar                       fb.modes             init              locale.alias   nanorc           python2.6  shadow                     weston
chatscripts                    fonts                init.d            locale.gen     netconfig        python2.7  shadow-                    wgetrc
ConsoleKit                     fstab                initramfs-tools   localtime      network          python3    shells                     wildmidi
console-setup                  fstab.d              inittab           logcheck       networks         python3.2  skel                       wpa_supplicant
cron.d                         fuse.conf            inputrc           login.defs     nsswitch.conf    rc0.d      ssh                        wvdial.conf
cron.daily                     gai.conf             insserv           logrotate.conf ntp.conf         rc1.d      ssl                        X11
cron.hourly                    gconf                insserv.conf      logrotate.d    openal           rc2.d      ssmtp                      xdg
cron.monthly                   gdb                  insserv.conf.d    magic          opt              rc3.d      staff-group-for-usr-local  xml
crontab                        ghostscript          iproute2          magic.mime     os-release       rc4.d      sudoers                    xpdf
cron.weekly                    gnome                issue             mailcap        os-release.orig  rc5.d      sudoers.d
dbus-1                         groff                issue.net         mailcap.order  pam.conf         rc6.d      sysctl.conf
pi@raspberrypi /etc $ sudo nano rc.local
pi@raspberrypi /etc $ cd ..
pi@raspberrypi / $ sudo reboot
```

Fig. 5 Screenshot of CGI bin directory

good diagnosis and unbiased prescriptions by doctors and physicians. The location accessibility can also be traced and store when the abnormality in the patient's health is detected.

6 Conclusions

The patient centric health care monitoring of data is forecasted with hockey-stick traction. As the pandemic has impacted in this contemporary world heavily in all the sectors in a down-slide pattern, except the pharmaceutical sector in which its impact has dramatically increased the consumer base. The surge has raised the demand for the telehealth services; therefore, this patient centric health monitoring system is largely useful for the health ministry and health related sectors for prescribing good medicines and for other purposes. In the existing system, the proponents weighed the concepts of measuring the heart rate and diagnosing the heart related ailments. It didn't rely heavily on wireless transfer of the sensed data. Some of them though adopted few WBAN technology for transferring the monitored data, it didn't provide good security for a reliable data transmission. Moreover, the access control features of most existing systems are not well-defined, and most of them are large in design aspects which makes them unsuitable for ease of operations. As the proposed system has biosensors and the WBAN as part of the system, it is integrated in nanochip level; therefore, the system is simple and compact and easy to be adapted by large set of demography making it a reliable system for the future. Thus, it possesses the access control features. The use of LAMP based web page with patient centric approach increases the credibility and accessibility amongst the users as the system is highly encrypted with the help of TINY key based codec.

References

1. O. Bestsennyy, G. Gilbert, A. Harris, J. Rost, in *Telehealth: A Quarter-Trilliondollar Post-COVID-19 Reality*?" (Mc Kinsey & Company, 2020)
2. H.A. El Zouka, M.M. Hosni, in *Secure IoT Communications for Smart Healthcare Monitoring System*. (Elsevier, 2019)
3. M.A. Akkaş, R. Sokkulu, H.E. Çetin, in *Healthcare and Patient Monitoring Using IoT*. (Elsevier, 2020)
4. C.C. Tai, J.R.C. Chien, An improved peak quantification algorithm for automatic heart rate measurements. in *IEEE 27th Annual Conference on Engineering in Medicine and Biology*, China (2005), pp. 6623–6626
5. H. Shim, J.H. Lee, S.O. Hwang, H.R. Yoon, Y.R. Yoon, Development of heart rate monitoring for mobile telemedicine using smartphone. in *13th International Conference on Biomedical Engineering* (ICBME 2008), Singapore (2008), pp. 1116–1119
6. M.A. Lalli, G. Garcia, l. Madrigal, M. Arcos-Burgos, M.L. Arcila, K.S. Kosik, F. Lopera, Exploratory data from, complete genomes of familial alzheimer disease age-at-onset, outliers. Human mutation (2012).https://doi.org/10.1002/humu.22167
7. P.F. Binkley, Predicting the potential of wearable technology. IEEE Eng. Med. Biol. Mag. **22**, 23–27. 7 (2003)
8. R.G. Landaeta, O. Casas, R.P. Areny, Heart rate detection from plantar bioimpedance measurements. in *28th IEEE EMBS Annual International Conference*, USA (2006), pp. 5113–5116
9. S. Rhee, B.-H.Yang, H.H. Asada, Modeling of finger photoplethysmography for wearable sensors. in *21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Soc. BMES/EMBS Conference* (1999)

10. C.R. Ibáñez, Revista ingeniería biomédica, escuela de ingeniería de antioquia-universidad ces, medellín, colombia, casos de innovación en salud en Colombia, vol 6 11 (2012), pp. 10–21. ISSN 19099762
11. K. Stella, E.N. Ganesh, Manikandan, Experimental analysis of Fault tolerant authentication in non-invasive epidermal glucose sensor using SQEZLMRP based information transmission in wireless sensor networks. Sensor Lett. **16**(3), 224–233 (2018)
12. S. Shakya, L. Nepal, Computational enhancements of wearable healthcare devices on pervasive computing system. J. Ubiquitous Comput. Commun. Technologies (UCCT) **2**(02), 98–108 (2020)
13. J.S. Raj, A novel information processing in IoT based real time health care monitoring system. J. Electron. **2**(03), 188–196 (2020)

Early Fault Detection in Safety Critical Systems Using Complex Morlet Wavelet and Deep Learning

A. Gandhimathinathan and R. Lavanya

Abstract Automated Fault Detection and Diagnosis (FDD) plays an important role in health monitoring of safety critical systems. Typically, critical industrial processes involve voluminous number of sensors that are capable of assessing the system's working condition and health. Time series analysis of sensor measurements can be used to predict potential system errors before the damage is irreparable. Predictive analysis is quintessential to reduce the system downtime and the costs associated. A major challenge in FDD is to detect faults much before the full-length time series is available, such that reliable predictions are achieved early in time. Thus, Early Classification of Time Series (ECTS) has to deal with the trade-off between accuracy and earliness, unlike conventional approaches that handle only accuracy. This paper proposes Complex Morlet Wavelet (CMW)-based time–frequency analysis for ECTS in a Deep Learning (DL) framework that combines Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) Network. The proposed approach is validated using the publicly available Tennessee Eastman Process (TEP) dataset. Results demonstrate that CMW in combination with hybrid CNN-LSTM architecture outperforms the state-of-the-art approaches for ECTS. The scheme is benefited by the DL architecture that combines CNN and LSTM, rather than these architectures considered individually. The proposed approach is able to achieve superior joint accuracy-earliness optimization when compared to time domain and frequency domain analyses considered separately, conventional Short Time Fourier Transform (STFT)-based time–frequency analysis, and other state-of-the-art time–frequency approaches.

Keywords Convolutional neural network · Early classification of time series · Long short-term memory · Complex morlet wavelet · Time–frequency analysis · Stopping rule · Cost function

A. Gandhimathinathan · R. Lavanya (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_lavanya@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_41

1 Introduction

Major industries comprise safety–critical systems and processes whose failure could end up in severe damage. Faults in such systems have to be addressed at an early stage to avoid progression to failures, irrecoverable losses and the associated huge cost-to-repair or replacement. Online monitoring of industrial processes facilitates predictive detection and diagnosis, offering huge benefits when compared to other maintenance strategies that include corrective and preventive diagnosis. While these two strategies are on two extremes of the maintenance spectrum leading to under-maintenance and over-maintenance, respectively, predictive maintenance strikes an optimum balance between the two, leading to condition-based maintenance, i.e., maintenance when need arises. The significance of predictive fault diagnosis is illustrated in Fig. 1.

The key to predictive maintenance is automated Fault Detection and Diagnosis (FDD) that employs mathematical models or approaches to analyze sensor measurements of the system that are indicative of prospective faults. Sensor measurements are typically available as a collection of data values, sequentially ordered in time, termed as time series. Automation in FDD can be realized by formulating it as a classification problem, where the time series is assigned to a predefined class which could be indicative of whether a fault has occurred or not, or could represent further classification of the fault type. Conventional time series classification considers a fixed length of time series which is supposedly the entire input data sequence. On the other hand, the goal of Early Classification of Time series (ECTS) is to correctly predict the class label of the time series with as few data points of the time series as possible [1].

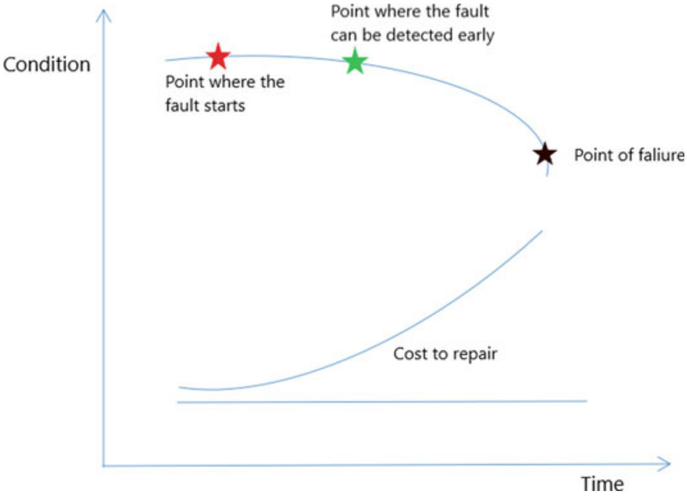

Condition
Point where the fault can be detected early
Point where the fault starts
Point of faliure
Cost to repair
Time

Fig. 1 Predictive fault diagnosis

ECTS is a natural choice when the data acquisition is costly or decision-making is time critical, though the latter aspect is the focus of the present work. The main objective of ECTS is to classify the time series as early as possible with accuracy as high as possible; however, these two objectives are contradictory in nature. Thus, the main challenge in ECTS is the identification of the earliest time instance of the data point at which an unlabeled time series can be classified with reliable prediction to achieve an optimum performance [1]. Early classification is finding increasing demand in many critical real-time applications in diverse industries. Most of the ECTS methods perform predictions at early time stamps and employ different approaches to decide whether the predictions are reliable or not. ECTS methods are broadly classified into three main categories, namely distance-based, shapelet-based and probabilistic classifier approaches.

The most popular distance-based approach for early classification is 1-Nearest Neighbor (1-NN) [2]. 1NN-based time series classification involves estimating the nearest neighbor to measure the linear alignment between sequences, followed by labeling based on majority voting. Minimum prediction length (MPL) was determined to extend classical 1-NN to ECTS, based on the concept of Reverse Nearest Neighbor (RNN). 1-NN has been combined with dynamic time warping (DTW) to measure the nonlinear alignment between sequences. This is achieved by warping the sequences to determine their similarity irrespective of the non-linear variations, if any, in the time domain. Though independent of the feature extraction stage, distance-based methods still pose the risks of large computational cost associated with distance computation and learning the classifier, which becomes prominent with complex measures and large datasets. Proper choice of the distance metric can control the trade-off between complexity associated with the metric and classification accuracy [3].

Shapelet-based techniques involve determination of similarity based on subsequences that are discriminative and maximally representative of the training data in the sense of information gain [4]. The use of local features makes this technique robust to noise. Further, shapelet exhibits phase-invariance characteristics as against approaches like 1-NN. On the downside, though relatively accurate and faster than the distance-based approach, this class of techniques suffers from a slow training time. Guoliang et al. [5] proposed a two-stage shapelet-based ECTS scheme, that included feature extraction and feature selection stages. In the feature extraction stage, local shapelets were extracted. During the selection phase, the optimal feature set was selected using a rule-based classifier.

Probabilistic classifiers aim to improve early classification by optimizing functions dependent on the time at which the decision is made and the probability output of the classifier [6–9]. The major advantage of this approach when compared to other non-probabilistic approaches such as distance-based and shapelet-based techniques is that the prediction result is associated with a confidence measure for the prediction. Schafer et al. [10] addressed the ECTS problem by employing a pair of probabilistic classifiers, instead of a single classifier. A two-stage classification approach was employed, where the classifier in the first stage was used to compute the class probabilities periodically and the classifier in the second stage was used to decide if the labels predicted by the first classifier is reliable.

Deep Learning (DL) can be considered as the most complex machine learning strategy and has recently found a prominent place in ECTS problem, so also a wide range of many other applications [11–16]. This is due to the inherent hierarchical feature representation capability of DL and recent revolution in high performance computing. Kotaro et al. [17] designed an Earliness-Aware Deep Convolutional Neural Network (CNN) for feature learning and classification. A multi-scale CNN architecture was coupled with a dynamic truncation model to allow early classification. A hybrid architecture employing CNN and Long Short-Term Memory (LSTM) was employed in a Multi-Domain Deep Neural Network (MDDNN) framework, to learn the embedded information in the sequence. CNN-based feature extraction was performed in time and frequency domains and the resulting feature maps were processed by LSTM layers to learn long term dependencies. The outputs of LSTM layers corresponding to the two domains were then merged and processed by fully connected layers for final prediction [18].

Ensemble of classifiers and the agreement between them has proved to improve the reliability of early classification [19]. More specifically, the agreement between time and frequency domain analyses has been an interesting research area. This has led to the hypothesis that techniques that implicitly provide joint time–frequency analysis can lead to improved performance in ECTS. This hypothesis has driven the authors to explore the performance of wavelet-based analysis for ECTS. More specifically, the Complex Morlet Wavelet (CMW) transform has been used for the purpose, due to its superior frequency tracking and temporal resolution advantages [20]. Further, the success of hybrid DL architectures such as CNN-LSTM has prompted the use of this architecture for early classification of faults, applied to industrial processes.

The remainder of this paper is organized as follows. Section 2 gives a brief overview on early classification. Proposed methodology is detailed in Sect. 3. In Sect. 4, experimental results and discussion are presented. Section 5 gives an insight to the future work and draws the conclusion.

2 Early Classification of Time Series

A time series is defined as an ordered sequence of data points consisting of time stamps and corresponding data values over a fixed length given by Eq. 1.

$$TS = \left\{\left(t_j, v_j\right), j = 1, \ldots, M\right\} \tag{1}$$

where t_j refers to real values of time instances and v_j corresponds to real parametric values measured at respective time instances.

Let $Y = \{(TS_1, CL_1), (TS_2, CL_2), \ldots, (TS_n, CL_n)\}$ be a labeled time series representing the training set, where TS_i represents the time series used for training, and $CL_i \in \{1, 2, \ldots, p\}$ represents the respective class labels where p is the total

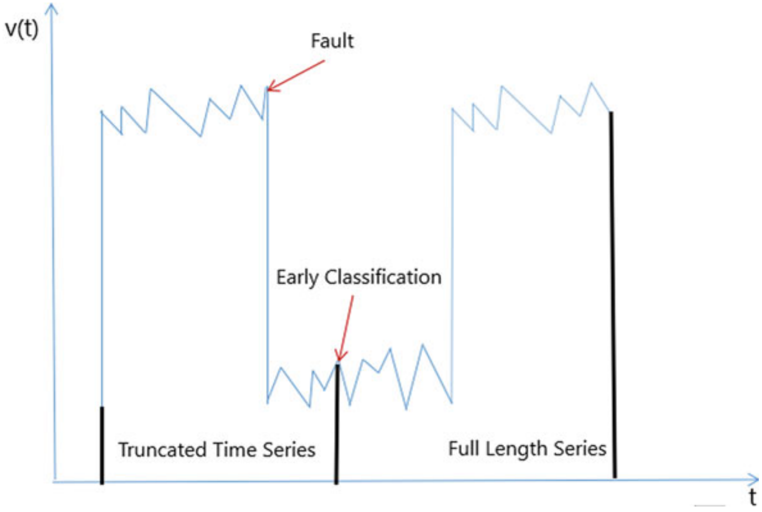


Fig. 2 Early time series classification

number of classes. Time series classification can be realized as a supervised learning problem in which the objective is to build a mapping from the time series to their class labels by using Y so as to predict the classes of new unlabeled time series as accurately as possible.

Let TS^t represent a truncated time series that considers only the first t pairs as specified in Eq. 2.

$$TS^t = \{(t_j, v_j),\ j = 1, \ldots, t\} \tag{2}$$

Suppose $X = \{(TS_1^t, CL_1^t), (TS_2^t, CL_2^t), \ldots, (TS_n^t, CL_n^t)\}$ is the labeled data representing the training set corresponding to the truncated time series TS_i^t and $CL_i^t \in \{1, 2, \ldots, p\}$ represents the respective class labels. Then, ECTS is a supervised learning task which attempts to build a mapping from the truncated time series to their class labels by using X (training set), which will be able to reliably predict the classes of new unlabeled time series as early as possible, using only a part of the series, namely TS^t. The reliability of the predictions is usually determined based on class discrimination. The idea of ECTS is illustrated in Fig. 2.

3 Proposed Methodology

The proposed scheme for ECTS involves the use of a hybrid CNN-LSTM based DL architecture that processes time–frequency features extracted from the raw time series using CMW as shown in Fig. 4. A stopping rule has been employed for determining the optimal earliness at which reliable prediction can be obtained.

This optimization as well as tuning of other hyper parameters is performed during the validation phase. A model is then built during the training phase with optimum truncation length series as the input, followed by testing the model using time series input of the same truncation length.

3.1 Complex Morlet Wavelet (CMW)

In this work, the CMW transform has been employed to perform joint time–frequency analysis of the input data, before it is fed to the DL architecture for fault detection. CMW is a complex sine wave multiplied by a real valued Gaussian window.

The ability to preserve the temporal resolution of the original signal is the major advantage of the CMW. CMW is defined as per Eq. 3.

$$\Psi(\mathrm{t}) = \frac{1}{\sqrt[4]{\pi}}\,\mathrm{e}^{\left(\mathrm{j}\omega_c \mathrm{te}^{--\frac{\mathrm{t}^2}{2}}\right)} \tag{3}$$

where $\omega_c = 2\pi \mathrm{fc}$ represents the central frequency of the wavelet.

3.2 Hybrid CNN-LSTM Architecture

The hybrid DL architecture employed in the proposed scheme is illustrated in Fig. 3. Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are

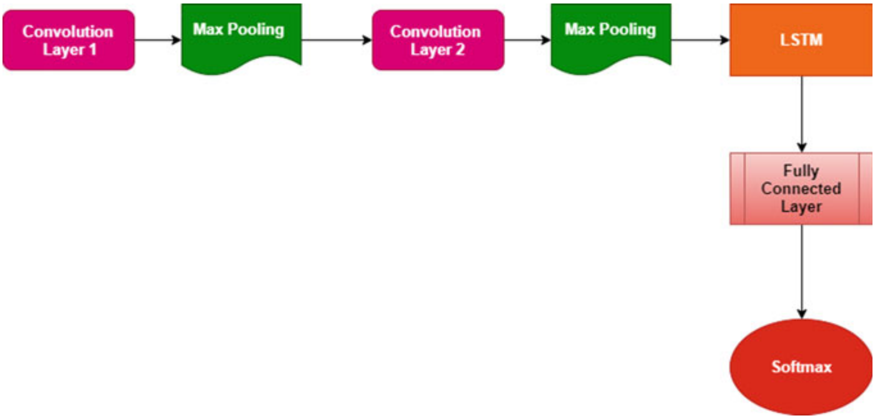


Fig. 3 Hybrid CNN-LSTM architecture

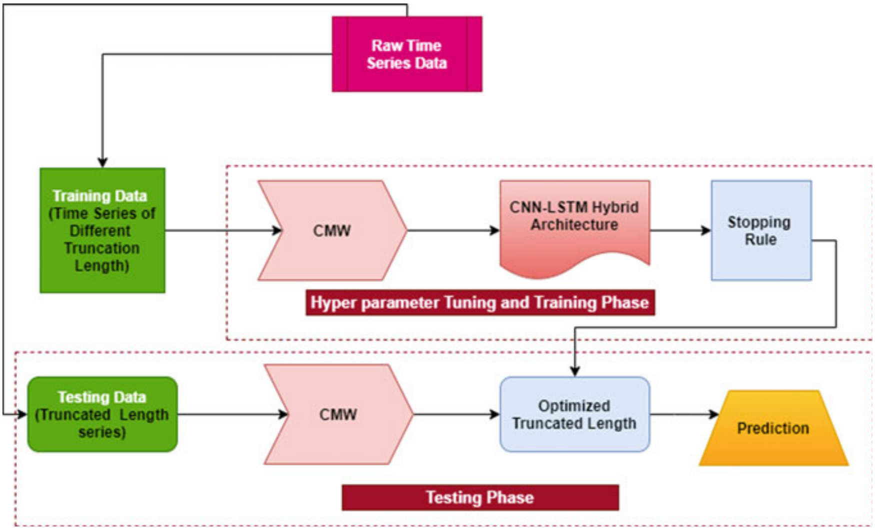

Raw Time Series Data
Training Data (Time Series of Different Truncation Length)
CMW
CNN-LSTM Hybrid Architecture
Stopping Rule
Hyper parameter Tuning and Training Phase
Testing Data (Truncated Length series)
CMW
Optimized Truncated Length
Prediction
Testing Phase

Fig. 4 Proposed model

combined to learn both the feature representation as well as the temporal relationship embedded in the time series data, for efficient time series classification.

The architecture used in this work comprises of two CNN layers for extraction of feature maps, followed by one LSTM layer for extracting temporal relationship from the feature maps. The output of the LSTM layer is processed by a fully connected (FC) layer, followed by a softmax layer that classifies the time series, producing a probability score at its output. Thus, the proposed approach is based on a serial architecture (CNN and LSTM layers connected in series), used to process CMW features extracted from the raw data.

3.3 Model Building

Models with different truncated lengths are built during the validation phase. For each case, the corresponding class probabilities and truncation length are given as the input to a stopping rule, as detailed in Sect. 3.4, to determine the final model with optimum truncation length that yields reliable prediction. The optimal parameters of the stopping rule as well as other hyper parameters for each of the model built is performed in the validation phase. Following this, the final model is built during the training phase. This optimal model is then tested with unseen data of corresponding optimal truncation length to determine reliable early prediction.

3.4 Stopping Rule

The stopping rule is used to find the optimal truncation length of the time series, where the prediction is reliable. This rule considers the posterior probabilities obtained by the classifier during the validation phase and the truncation length, while deciding the optimal length [21, 22]. The stopping rule (SR) is defined in Eq. 4.

$$SR_{\beta}\left(pp^{t}, t\right) = \begin{cases} 0 \text{ if } \beta 1 pp^{t}_{1:m} + \beta 2\left(pp^{t}_{1:m} - pp^{t}_{2:m}\right) + \beta 3 \frac{t}{N} \leq 0 \\ 1 \text{ otherwise} \end{cases} \quad (4)$$

where $pp^{t} = \left(p^{t}_{1}, p^{t}_{2}, p^{t}_{3} \ldots p^{t}_{m}\right)$ represents the posterior probabilities of m possible classes and $pp^{t}_{1:m}$ and $pp^{t}_{2:m}$ denotes the first and the second largest class probabilities obtained at a particular time instant t. $\beta 1, \beta 2\, and\, \beta 3$ denote the parameters of the stopping rule which typically lie in the range [−1,1]. N denotes the length of the full series and $\frac{t}{N}$ is the ratio of the length of the truncated series to the length of the full series and is called earliness.

The values of $\beta 1, \beta 2\, and\, \beta 3$ are determined through an optimization process that involves minimization of the cost function defined in Eq. 5.

$$\text{CF}\left(X,\ \text{SR}_{\beta}\right) = \frac{1}{|X|} \sum_{x \in X} \left(\mu\, C_{\text{miss}}(X, \text{SR}_{\beta}) + (1 - \mu) C_{\text{delay}}\left(X, \text{SR}_{\beta}\right) + \theta \|\beta\|_{2} \quad (5)$$

where μ is the weight parameter associated with two objectives earliness and accuracy and takes the values of 0.6, 0.7, 0.8 and 0.9 [19]. θ is the regularization parameter which assumes values given by 0.003, 0.001, 0.03, 0.01, 0.3 and 0.1 [19]. C_{miss} is the misclassifications cost which is 0 if true class label matches with predicted class label, otherwise 1 and $\frac{t}{N}$ is the cost associated with delay function (C_{delay}), where t be the earliest time at which the stopping rule is optimized.

The stopping rule is iteratively evaluated for progressively increasing truncation lengths in steps of 5%. The main intuition is to find the optimal point where the prediction is reliable. This rule considers the posterior probabilities obtained by the classifier during the training phase to decide the optimal point. The stopping rule outputs a value 1 when the prediction is reliable. On the other hand, if the stopping rule outputs a value 0, it represents the condition that the prediction is not reliable enough, which necessitates waiting for more data.

The highest class probability and the difference between two highest probabilities which are both highly relevant to class discrimination have been used in the stopping rule along with with earliness value as seen in Eq. 4. Under circumstances where the rule is not triggered for the complete time series, predictions obtained at full length series shall be considered.

4 Results and Discussion

In this work, Tennessee Eastman Process (TEP) database involving a large set of sensor measurements, associated with process control of an industrial chemical process, was used for validation. TEP is a benchmark simulation model that can be used for developing and testing fault diagnosis algorithms on time series data and could be extended to ECTS as well. TEP dataset consists of fault-free and faulty data comprising 20 different fault types. A total of 52 parameters corresponding to various sensor measurements have been captured over 500 simulation runs for normal data as well as each of the 20 faults. A particular simulation run comprises of 500 samples for training and 960 samples for testing [23].

This work is restricted to univariate data, i.e., only a single sensor measurement is considered for fault detection problem. Among all fault types, only fault 13 is considered for the analysis in this work, because it exhibits slow drift characteristics and is highly suitable for the early classification problem. Further, as the focus is early classification, the channel in which the sensor data represents gradual variation on fault injection is a natural choice for the analysis considered. On inspection of the database, variable 21 corresponding to reactor cooling water outlet temperature sensor measurement has been found to exhibit such gradual variations with fault type 13, and hence has been considered in this work. A total of 2,50,000 training samples and 4,80,000 testing samples are present for normal category as well as faulty category.

As mentioned earlier, the proposed scheme in this work explores time and frequency domain information embedded in the data for early classification of the time series. For comparison purposes, to start off, time domain ECTS has been attempted with raw time series using CNN, LSTM and hybrid CNN-LSTM architectures. All these three approaches are extended to frequency domain analysis and time–frequency analysis. For mapping the time domain information to frequency domain, Fast Fourier Transform (FFT) was employed. For time–frequency analysis, Short Time Fourier Transform (STFT) and Complex Morlet Wavelet (CMW) have been applied to the time series.

To demonstrate the efficiency of the proposed CMW-based ECTS approach listed below as S12, it is compared with other schemes labeled as S1 through S11, as well as a state-of-the-art approach [18], mentioned as S13 below.

S1: Raw time series + CNN.
S2: FFT + CNN.
S3: STFT + CNN.
S4: Raw time series + LSTM.
S5: FFT + LSTM.
S6: STFT + LSTM.
S7: Raw time series + CNN-LSTM.
S8: FFT + CNN-LSTM.
S9: STFT + CNN-LSTM.
S10: CMW + CNN.

Table 1 Accuracies for time domain ECTS approach

Time-domain approach (%)			
Earliness	CNN	LSTM	CNN-LSTM
1	88.00	76.42	91.78
0.99	87.72	75.80	91.71
0.98	87.40	74.10	91.65
0.97	87.20	73.25	91.59
0.96	86.80	72.50	91.40
0.95	86.00	71.20	91.03
0.90	83.81	67.81	90.63
0.80	80.00	61.18	86.88
0.70	71.20	53.08	81.24
0.60	63.40	52.20	78.10
0.50	55.00	42.45	65.14
0.40	47.90	35.11	58.36
0.30	41.10	31.38	50.81
0.20	34.20	28.64	43.07
0.10	29.00	27.20	36.81

S11: CMW + LSTM.

S12: CMW + CNN-LSTM: Proposed.

S13: Parallel Combination of S7 & S8: State-of-the art [18]

In the hybrid architecture, CNN and LSTM layers are connected serially in that order. The input signal is processed by the CNN layers resulting in feature maps. These CNN features are passed on to LSTM, whose output is then propagated to fully connected and softmax layers.

The proposed CMW-based approach involving hybrid CNN-LSTM architecture has also been compared with a state-of-the-art scheme [18], which is also a time–frequency approach. This scheme involves a parallel combination of time and frequency-based hybrid CNN-LSTM architectures, where features from the LSTM layers of the two hybrid architectures are merged and fed to fully connected and softmax layers.

Unlike normal classification schemes which report only the classifier performance, ECTS should attempt to assess the classifier performance at a given earliness, to understand how early in time reliable predictions are possible. Hence, the accuracies for various schemes listed above have been reported below, for different earliness values. The optimum earliness value for each scheme, derived using the stopping rule, is presented thereafter. Typically, the scheme that co-optimizes accuracy and earliness is considered as the most optimum solution for ECTS.

Tables 1, 2, 3 and 4 summarize the accuracies achieved at different earliness values for the time domain approaches (S1, S4 and S7), FFT-based approaches (S2, S5 and

Table 2 Accuracies for FFT-based frequency domain ECTS approach

FFT-based frequency domain approach (%)			
Earliness	CNN	LSTM	CNN-LSTM
1	91.44	78.02	92.13
0.99	91.10	77.63	92.04
0.98	90.89	77.26	91.86
0.97	90.56	76.0.31	91.66
0.96	90.38	75.08	91.53
0.95	90.10	73.98	91.09
0.90	88.90	71.41	90.71
0.80	87.00	67.08	88.97
0.70	82.00	58.92	85.08
0.60	77.50	55.76	81.64
0.50	63.80	50.26	77.38
0.40	56.78	46.52	69.06
0.30	51.20	39.98	62.31
0.20	47.20	33.71	51.60
0.10	35.60	30.83	40.09

Table 3 Accuracies for STFT-based ECTS approach

STFT-based time–frequency domain approach (%)			
Earliness	CNN	LSTM	CNN-LSTM
1	97.28	86.87	97.98
0.99	97.58	85.61	97.70
0.98	97.35	84.02	97.57
0.97	97.03	82.37	97.32
0.96	96.94	81.48	97.16
0.95	96.83	78.55	97.02
0.90	95.31	75.91	96.47
0.80	88.84	71.02	90.21
0.70	84.34	66.34	87.86
0.60	78.04	60.59	84.45
0.50	66.98	56.25	80.63
0.40	60.56	50.49	72.34
0.30	56.81	44.04	66.11
0.20	47.85	39.22	59.53
0.10	41.95	35.07	49.76

Table 4 Accuracies for CMW-based ECTS approach

CMW- based time–frequency domain approach (%)			
Earliness	CNN	LSTM	CNN-LSTM
1	97.92	90.12	98.83
0.99	97.81	87.91	98.74
0.98	97.73	86.44	98.40
0.97	97.62	84.97	98.35
0.96	97.49	83.96	98.26
0.95	96.94	81.64	98.08
0.90	96.29	78.49	97.84
0.80	92.38	75.28	94.08
0.70	89.29	72.16	90.46
0.60	83.32	67.85	88.19
0.50	76.78	60.40	83.54
0.40	71.56	53.21	78.35
0.30	65.61	47.88	71.29
0.20	58.98	43.98	66.06
0.10	49.80	41.98	59.92

S8), STFT-based approaches (S3, S6 and S9) and CMW-based approaches (S10, S11 and S12), respectively, using CNN, LSTM and hybrid CNN-LSTM architectures.

It can be observed from Tables 1,2,3 and 4 that LSTM results in a low accuracy for a given earliness compared to CNN for all schemes. The combination of CNN and LSTM not only outperforms LSTM, but also CNN. The computational complexity of the proposed hybrid model is higher than the constituent single model trading off training time for better performance.

It can also be observed for all schemes that as earliness decreases, which means the earlier the prediction is performed, the accuracy drops accordingly. This is expected, as earliness and accuracy are conflicting parameters as discussed already. However, the earliness value at which the prediction is reliable has to be determined.

Figure 5 compares the accuracies of CNN approaches (S1, S2, S3 and S10) for all the schemes, viz. time, frequency, STFT and CMW analysis. It is evident that CMW-based CNN approach excels the other CNN-based approaches by an average of 16.05%, across various earliness values. Figure 6 compares the accuracies of LSTM approaches (S4, S5, S6 and S11) for all the schemes. It is observed that CMW based LSTM approach excels the other LSTM based approaches by an average of 14.26%. Figure 7 compares the accuracies of CNN-LSTM approaches (S7, S8, S9 and S12) for all the schemes. It is evident that CMW-based CNN-LSTM approach excels the other CNN-LSTM based approaches by an average of 12.01%. This is also evident by comparing performance of different schemes across Tables 1,2,3 and 4. Thus, across various schemes, CMW yields superior performance, and across various architectures, hybrid CNN-LSTM yields the best performance.

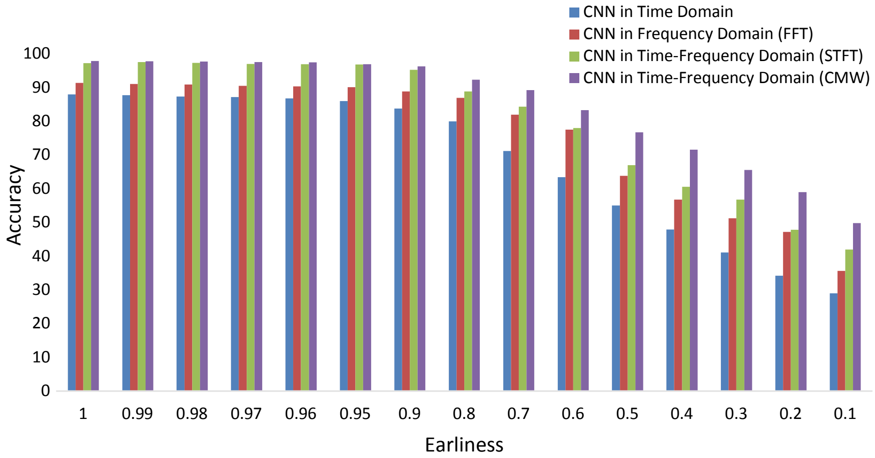


Fig. 5 Comparison of CNN in different domains

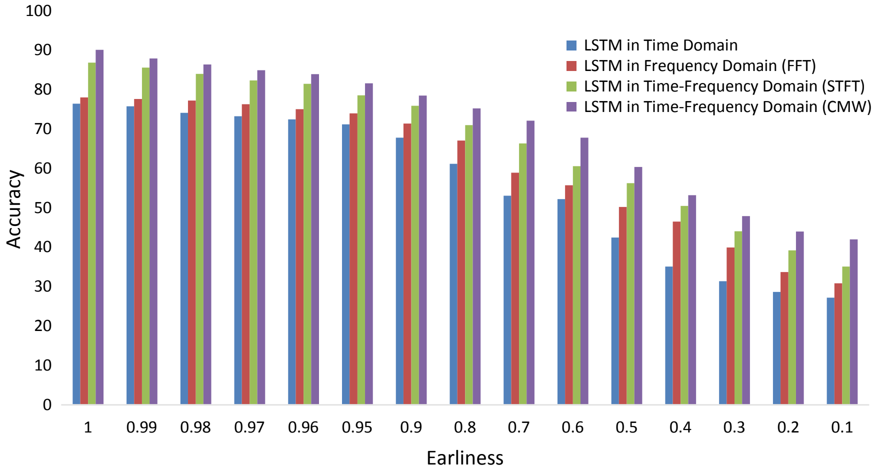


Fig.6 Comparison of LSTM in different domains

The combination of CMW and hybrid CNN-LSTM yields the best classification accuracy across various domains and architectures, justifying the proposed hypothesis. More importantly, the performance improvement in the proposed approach is much higher for early classification when compared to full length classification.

The accuracy of the proposed approach is higher at most by 22.41% when compared to the other methods for full length classification. On the other hand, for ECTS with low earliness values (corresponding to time series truncated earlier in time), the improvement is higher. For the least earliness corresponding to 0.1, the

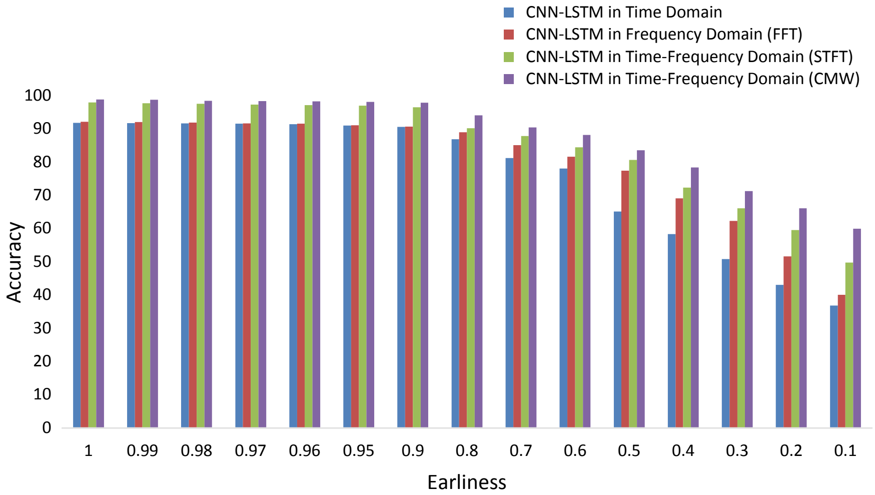


Fig.7 Comparison of CNN-LSTM in different domains

Table 5 Comparison of the proposed approach with state-of-the-art approach

Comparison of state-of-the-art with proposed approach			
Earliness	State-of-the-art	STFT + CNN-LSTM	CMW + CNN-LSTM (Proposed)
1	94.56	97.98	**98.83**
0.99	94.19	97.70	**98.74**
0.98	94.07	97.57	**98.40**
0.97	93.84	97.32	**98.35**
0.96	93.55	97.16	**98.26**
0.95	93.24	97.02	**98.08**
0.90	91.37	96.47	**97.84**
0.80	89.26	90.21	**94.08**
0.70	86.51	87.86	**90.46**
0.60	83.69	84.45	**88.19**
0.50	79.08	80.63	**83.54**
0.40	70.17	72.34	**78.35**
0.30	65.82	66.11	**71.29**
0.20	56.35	59.53	**66.06**
0.10	45.23	49.76	**59.92**

maximum improvement in accuracy achieved by the proposed system is 32.72%. This shows the effectiveness of the proposed approach for the ECTS problem.

In Table 5, the proposed scheme combining CMW and hybrid CNN-LSTM is also compared with that of the state-of the-art approach, i.e., S13. The results of the proposed scheme are reproduced in Table 5 for the sake of comparison.

It can be recalled that scheme S13 is also based on time–frequency analysis and realized by a parallel combination of time and frequency-based CNN-LSTM architectures, with LSTM features from the two parallel branches used in further processing. From Table 5, it can be inferred that the CMW approach performs better performance when compared to the state-of-the-art approach.

So far, the results presented focus only on accuracies achieved at different earliness value for various schemes. The forthcoming results and discussion pertain to joint optimization of accuracy and earliness, using the stopping rule.

The results of evaluating the stopping rule for various schemes are tabulated in Table 6. The table basically summarizes the optimum earliness for the different schemes and the respective accuracies as determined by the stopping rule. From Table 6, it is observed that proposed CMW-based approach achieves a maximum accuracy of 78.91% with an earliness of 0.4 when compared to state-of-the-art approach with 77.38% with an earliness of 0.5. It is also observed that the proposed method yields the best ECTS performance, by providing superior joint optimization of accuracy and earliness, when compared to all the other schemes.

Table 6 Comparison of accuracy (%) versus earliness using stopping rule for different schemes

Schemes		Earliness	Accuracy (%)
Raw time series-based schemes	S1	0.5	54.51
	S2	0.5	41.03
	S3	0.4	56.51
FFT-based schemes	S4	0.4	54.28
	S5	0.5	49.40
	S6	0.4	67.97
STFT-based schemes	S7	0.5	65.01
	S8	0.6	58.49
	S9	0.4	71.27
CMW-based schemes	S10	0.5	76.23
	S11	0.6	65.81
	S12 (Proposed Method)	0.4	78.91
S13 (State-of-the-Art)		0.5	77.38

5 Conclusion

In this work, a time–frequency-based approach using Complex Morlet Wavelet (CMW) transform in a Deep Learning (DL) framework comprising a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM) has been proposed for Early Classification of Time Series (ECTS) applied to predictive fault detection in industrial processes. The proposed approach has been validated using the Tennessee Eastman Process (TEP) dataset, resulting in superior joint accuracy-earliness optimization than state-of-the-art approaches, yielding 78.91% accuracy at an earliness of 0.4 and achieving a better accuracy-earliness trade-off when compared to state-of-the art approaches. Scope of the work can be further extended by considering more sophisticated options for time–frequency analysis. Further, multiple weak classifiers can be constructed in the time–frequency domain and the agreement between the individual classifiers comprising the ensemble can be used for improving the confidence associated with early classification. Better stopping rules, cost function and deep architectures can be employed to improve the performance.

References

1. A. Gupta, H.P. Gupta, B. Biswas, T. Dutta, Approaches and applications of early classification of time series: a review. IEEE Trans. Artif. Intell. (2020)
2. Z. Xing, J. Pei, S.Y. Philip, Early prediction on time series: a nearest neighbor approach. in *Twenty-First International Joint Conference on Artificial Intelligence* (2009)
3. R.J. Kate, Using dynamic time warping distances as features for improved time series classification. Data Min. Knowl. Discov. **30**(2), 283–312 (2016)
4. J. Hills, J. Lines, E. Baranauskas, J. Mapp, A. Bagnall, Classification of time series by shapelet transformation. Data Min. Knowl. Discov. **28**(4), 851–881 (2014)
5. G. He, W. Zhao, X. Xia, R. Peng, X. Wu, An ensemble of shapelet-based classifiers on inter-class and intra-class imbalanced multivariate time series at the early stage. Soft Comput. **23**(15), 6097–6114 (2019)
6. A. Sharma, S.K. Singh, Early classification of multivariate data by learning optimal decision rules. Multimed. Tools Appl. 1–24 (2020)
7. H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.A. Muller, Deep learning for time series classification: a review. Data Min. Knowl. Discov. **33**(4), 917–963 (2019)
8. R. Tavenard, S. Malinowski, Cost-aware early classification of time series. in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (2016), pp. 632–647
9. U. Mori, A. Mendiburu, E. Keogh, J.A. Lozano, Reliable early classification of time series based on discriminating the classes over time. Data Min. Knowl. Discov. **31**, 233–263 (2017)
10. P. Schafer, U. Leser, TEASER: early and accurate time series classification. Data Min. Knowl. Discov. **34**(5), 1336–1362 (2020)
11. A. Balaji, D.S. Jayanth, H. Ram, B.B. Nair, A deep learning approach to electric energy consumption modeling. J. Intell. Fuzzy Syst. **36**(5), 4049–4055 (2019)
12. M. Ganesan, R. Lavanya, M. Nirmala Devi, Fault detection in satellite power system using convolutional neural network. Telecommun. Syst. **2020**, 1–7 (2020)
13. A. Rajkumar, M. Ganesan, R. Lavanya, Arrhythmia classification on ECG using deep learning. in *International Conference on Advanced Computing and Communication Systems* (ICACCS) (2019), pp. 365–369

14. S. Negi, C. Santhosh Kumar, A. Anand Kumar, Feature normalization for enhancing early detection of cardiac disorders. in *IEEE Annual India Conference* (INDICON) (2016), pp. 1–5
15. S. Shakya, Process mining error detection for securing the IoT system. J. ISMAC **2**(3), 147–153 (2020)
16. D. Nirmal, Artificial intelligence based distribution system management and control. J. Electron. **2**(2), 137–147 (2020)
17. K. Nakano, B. Chakraborty, Effect of data representation for time series classification-a comparative study and a new proposal. Machine Learn. Knowl. Extract. **1**(4), 100–1120 (2019)
18. H.-S. Huang, C.-L. Liu, V.S. Tseng, Multivariate time series early classification using multi-domain deep neural network. in *IEEE 5th International Conference on Data Science and Advanced Analytics* (DSAA) (2018), pp. 90–98
19. E.Y. Hsu, C.-L. Liu, V.S. Tseng, Multivariate time series early classification with interpretability using deep learning and attention mechanism. in *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (2019), pp. 541–553
20. M.X. Cohen, A better way to define and describe Morlet wavelets for time-frequency analysis. Neuroimage **199**, 81–86 (2019)
21. A. Sharma, S.K. Singh, A novel approach for early malware detection’. Trans. Emerg. Telecommun. Technol. **32**(2), 3968 (2021)
22. U. Mori, A. Mendiburu, S. Dasgupta, J.A. Lozano, Early classification of time series by simultaneously optimizing the accuracy and earliness. IEEE Trans. Neur. Net. Learn. Sys. **29**(10), 4569–4578 (2017)
23. C.A. Rieth, B.D. Amsel, R. Tran, M.B. Cook, Additional tennessee eastman process simulation data for anomaly detection evaluation. Harvard Dataverse. **2017**, 1 (2017)

Storage Mechanism for Heterogeneous Streamed Sensor Data

J. RubyDinakar and S. Vagdevi

Abstract The evolutions in big data, IoT and cloud native applications like microservices architecture provide a different approach for streaming analytics. The heterogeneous streamed sensor data from multiple sources like sensor, social media and web are generated continuously and that data need to be analysed immediately to gain insights and finding interesting patterns on the data. The enormous volume of data will be generated continuously and retaining only the required information becomes the biggest challenge due to its variety and velocity. Storing of this streamed information directly on to the cloud ends up in incurring additional charges to the end user also the unwanted, redundant information creates a bottleneck for data analysis. It requires large amount of space for storage. While the traditional methods pose a huge challenge for scalability and interoperability, our work suggests a suitable architectural approach to stream and store the information using event-based streaming with microservices architecture model. We store the clean and pre-process data which has the required feature vectors for streaming analytics.

Keywords Apache Flink · Microservices architecture · Event-based streaming · Heterogeneous sensor data · Real-time analytics · Data storage · Data preprocessing · MongoDB

1 Introduction

Every day the world is witnessing the growth of technology and the data generated using those technologies. The volume of data generated is growing like a huge mountain day by day. Now a days, data analytics play a significant role in analysing data from covid-19 to elections. It helps in taking decisions and providing data driven

J. RubyDinakar (✉)
VTU Research Scholar, Faculty of Computer Science and Engineering, PES University, Bangalore, India
e-mail: rubydinakar@pes.edu

S. Vagdevi
Professor, Department of Computer Science and Engineering, City Engineering College, Bangalore, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_42

solutions. In recent years, the application of IoT-based technologies have portrayed a key role in the growth of streaming information for analytics and monitoring the situations like tracking covid-19 patients contacts, vaccination drive, pre-polls, exit polls and election results, etc. Interestingly, the connected world has become one of the main source of data generation. It has continuously generated humongous amount of heterogeneous information. Processing and analysing the large piles of information is becoming cumbersome. In conventional processing, the data coming from multiple sources are not taken into consideration for analysing the data. But when you combine the data from multiple sources it may provide more meaningful insights. For example, the air pollution data is analysed along with healthcare data it can help us to understand the impact of pollutants in people's health. Many streaming techniques and paradigms are available to improve the streaming analytics. Generally, we have real time and batch processing paradigm in distributed environment. There are various stream processing platforms available to stream the data in an efficient manner. Each one of them has its own advantages and disadvantages with respect to a specific context. Sensor data is streamed from diverse devices in a distributed environment. It generates schematically different format of heterogeneous data naturally. This data should be analysed within a short span of time which can provide a better insight on the data and its patterns. Within the environmental monitoring [1] domain, time-series information and historical dataset are crucial for prediction models. Figure 1 shows the overall architecture of streaming analytics. Humongous information established the worth of insights derived from processed stream information. Such insights aren't equal. Some insights are additionally valuable shortly when it's happened. There is a necessity to store and manage this information to achieve valuable insights. Stream process permits such situations, providing insights quicker, usually within milliseconds to seconds from the trigger. Sensor device information naturally comes as an endless stream. To process this information, the information must be held on and data ingestion should be stopped for a time period to process the information. After sometime future batch of information must be processed and then we have to take care of aggregating this data across multiple batches for any process. However, streaming paradigm give an answer to the present downside.

We are able to determine the patterns, examine the results and identify the relationship between these multiple information to grasp it during a higher manner and additionally analyse the information from multiple sources at the same time. Stream process naturally fits with temporal data. Also we can identify the trend and patterns over time. The continuous data examples are traffic sensors, health sensors, environmental data and dealing activity logs, etc. Most of the IoT information is temporal data. Hence, it is good to use a programming model that matches naturally. Batch method lets the information build-up and later process them at once whereas stream processing process the data as soon as it arrives. Hence, stream process can work with less hardware. Moreover, stream process allows approximate querying via systematic load shedding. Hence, stream process fits naturally into use cases wherever approximate answers are sufficient. Generally, the information is big, and it's not even potential to store all of it. Our approach helps us to store the required data which can be used by other systems at the same time or near real time.

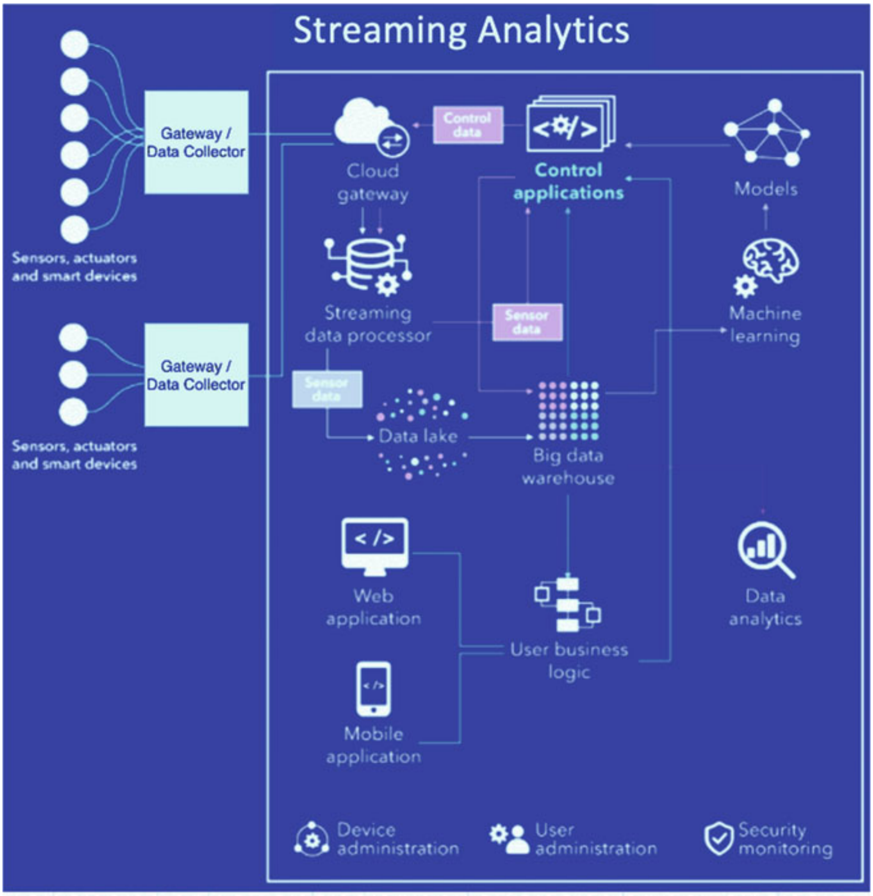


Fig. 1 Streaming analytics overall architecture

Cloud native refers to how the system is built and deployed rather than where the application resides [2]. A cloud native application consists of discrete, reusable components known as microservices that are designed to integrate into any cloud environment. It acts as a building block and it is packaged in a container. It is a small independent loosely coupled services which provide interoperability by integrating different services together. Since streaming application based on IoT pull data from different heterogeneous sources and services, it is well suited with microservices architecture. Event driven architecture makes the microservices-based paradigm suitable for Flink-based streaming. Flink streaming model consumes large amount of data with low latency. The paradigm itself supports data aggregation based on a time-based window and join operation can be performed to combine different data streams into a single stream.

2 Related Work

This section discusses the analysis works associated with microservices design usage in IoT framework, data management problems and approaches related to storage management. The authors Jiang et al. [3] suggested a solution for storage management with relevant to structured, unstructured data and RESTful service for generating platform independent interface. Tomislav et al. [4] had given an outline of specific IoT connected issues like measurability and introduced microservice-based middleware design to provide integration between differing kinds of devices, services and communication protocols. The authors [5] have compared 3 frameworks, particularly Apache Storm, Apache Flink, and Spark Streaming to showcase which platform is appropriate for what reason in the context of streaming application. This paper targeted on evaluating the performance of frameworks in terms of latency, and throughput. The authors [6] used document-oriented databases like MongoDB to facilitate the storage of semi-structured data. They mentioned regarding sharding which helps in scaling the infrastructure along with the expansion of data. Apache Spark framework is combined with MongoDB to improve the potency of data ingestion process. The authors [7] have projected an adaptative solution for satisfying the dynamic and heterogeneous requirements that IoT platforms are inevitably facing. In large data sets, wherever stream mining is that the approach of choice, data preprocessing and reduction have become vital methodologies for knowledge discovery [8]. The authors establish the essential role of such methodologies inefficient machine learning systems. The authors have given an adaptative preprocessing approach that benefits the prediction accuracy on real sensory information. The paper addresses adaptation to concept drift [9] within the input data stream. Microservices-based architectural approach [11–13] is strongly emphasized in IoT-based data analytics. Our approach helps in overcoming the problem of scalability, integration of numerous devices in streaming distributed environment.

The following table summarizes the different approaches discussed for streaming data management in few papers (Table 1).

Table 1 Summary

Proposal/Features	Data	Database support	Cloud support	Architectural approach
[3]	Structured, semi structured, unstructured	NoSQL and relational database	Multitenant	Service Oriented Architecture
[4]	Semi structured	No	Yes	Microservices
[6]	JSON format	MongoDB	Yes	Apache Spark
[7]	JSON format	No	Yes	Microservices
[9]	Structured	No	No	Adaptive learning

3 Architecture

This section focusses on various aspects of event driven microservices architecture for stream analytics with respect to architectural characteristics, components, layers and technical solution architecture inclusive of storage.

3.1 Architectural Characteristics of the Solution

While dealing with streaming sensor data in a distributed platform from multiple sensors for analytics, there are few architectural characteristics which have been addressed are High Throughput, Scalability, Availability, Reliability, Modularity, Maintainability and Agility.

High Throughput:

The distributed platform processes the humongous amount of data with low latency. The response time is faster, due to the continuous streaming of data back pressure may be accumulated. Flink can handle this requirement easily.

Scalability:

The data is streamed from different devices with different schema, supporting future diverse devices to be introduced or accommodate new data formats, the distributed framework will be able to scale up or scale down using Microservices architecture which is vertically scalable along with Flink wherein each process runs as a task in a worker node providing high scalability.

Availability:

The data must be available at all the times even if any node fails, Big Data paradigm itself replicates data in different nodes. Thus, data will be available in the backup node during failures.

Reliability:

The streamed data must be reliable with respect to consistency and data quality.

Modularity:

The complex problem of streaming real time or near real-time analytics should be broken down into simple smaller modules. The tasks are containerized so that any changes happen in one task or component should not affect the remaining.

Maintainability:

The solution should not be a solution for a limited requirement of only one type of raw data, the solution should provide flexibility for modification to each component of the architecture.

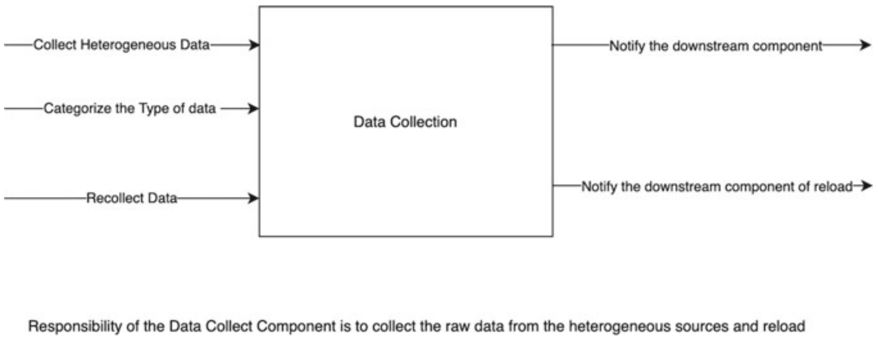
Agility:

An agile approach is a dynamic architecture willing to facilitate changes to the systems, better utilization of resources, and timely completion of processes and control of the implementations, integrations of distributed components on the core layer of the system. Microservices architecture provides agility to the approach.

3.2 Architectural Components:

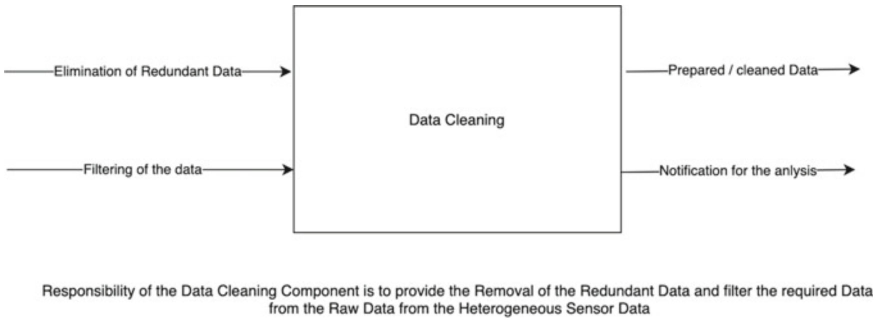
Microservices are implemented as an independent component. These independent components are represented in Figs. 2, 3, 4, 5, 6 and 7. The component architecture of our proposed solution is given below.

1. Data collection component would be the component to collect the heterogeneous data from different type of sensors. Here, heterogeneity refers the different schema of sensors. The data collected from this component are raw data that is unstructured data
2. Data Cleaning is the process of eliminating redundant data and filtering.
3. Data Exploration is the initial step in the process of data analysis, where users explore a large data set in an unstructured way to uncover initial patterns.
4. Feature engineering is the process of transforming raw data into features that better represent the underlying problem in the predictive model. It is the cleaned data ready for analytics and this data will be retained and stored in the cloud.
5. Predictive modelling is the process of doing forecast analysis to find interesting patterns in the data.
6. Data visualization is the process of reporting the result of analytics to the user using various charts and graphs.


Collect Heterogeneous Data
Categorize the Type of data
Recollect Data
Data Collection
Notify the downstream component
Notify the downstream component of reload

Responsibility of the Data Collect Component is to collect the raw data from the heterogeneous sources and reload

Fig. 2 Data collection component



Responsibility of the Data Cleaning Component is to provide the Removal of the Redundant Data and filter the required Data from the Raw Data from the Heterogeneous Sensor Data

Fig. 3 Data cleaning component

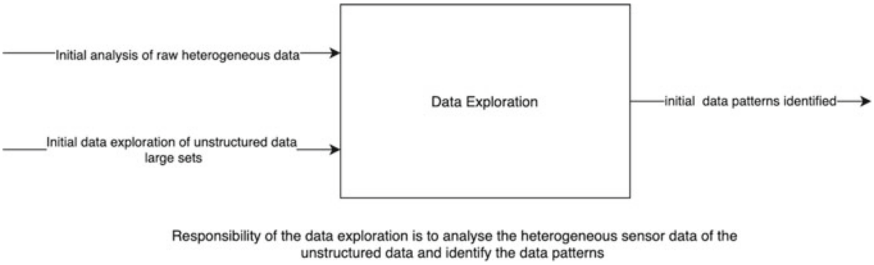

Initial analysis of raw heterogeneous data
Initial data exploration of unstructured data large sets
Data Exploration
initial data patterns identified
Responsibility of the data exploration is to analyse the heterogeneous sensor data of the unstructured data and identify the data patterns

Fig. 4 Data exploration component

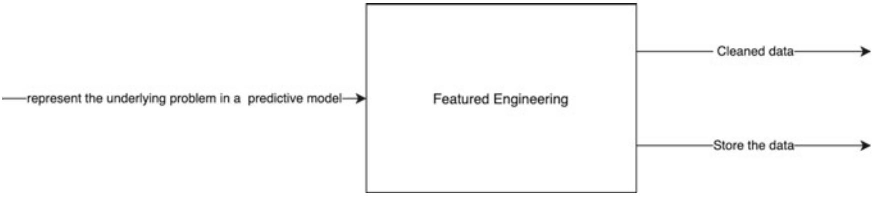


Fig. 5 Feature engineering component

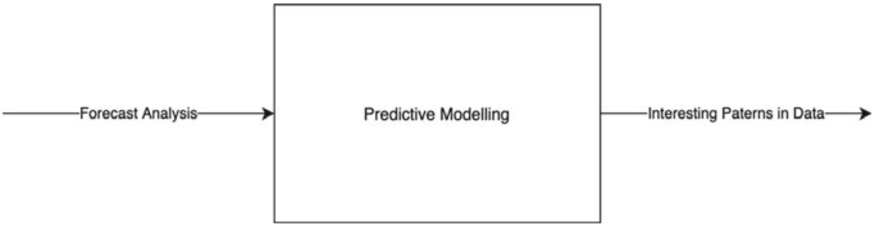


Fig. 6 Predictive modelling component

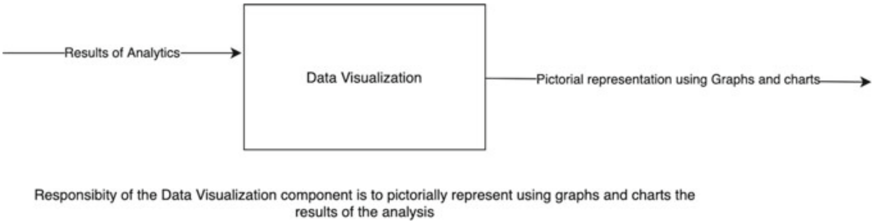

Results of Analytics
Data Visualization
Pictorial representation using Graphs and charts
Responsibity of the Data Visualization component is to pictorially represent using graphs and charts the results of the analysis

Fig. 7 Data visualization component

3.3 Technical Solution Architecture

The brief description of the experimental components of the architecture is enlisted below.

1. Event-based Streaming on Flink–Apache Flink is used for streaming the data from multiple sources.
2. Microservices–The tasks collect, analyse and report mentioned in the Fig. 8 are considered as microservices which are modular with interaction with components like code functions, data and other resources.
3. Dynamic Integration–Apache Kafka is used as a message queueing system for producing data and publishing to the relevant topics. Flink Kafka consumer is used to subscribe and consume a topic for the loosely coupling of the solution and providing an integration layer for integration with any third-party distributed systems.
4. Analytics–Analysis of data will be performed by a Machine Learning Component algorithm this can be implemented by a python component. Sensor data is a time-series data. Forecast analysis can be performed using deep learning technique to find long time trend, seasonal impact and correlation between feature vectors and independent vectors.
5. Reporting of the information after the analysis stage will require an efficient dashboard with charts to articulating/presenting the findings of the predictive analysis. For example, line chart can be used to display the trend and seasonal changes in the environment and regression plot can be used to study the correlation of different vectors.

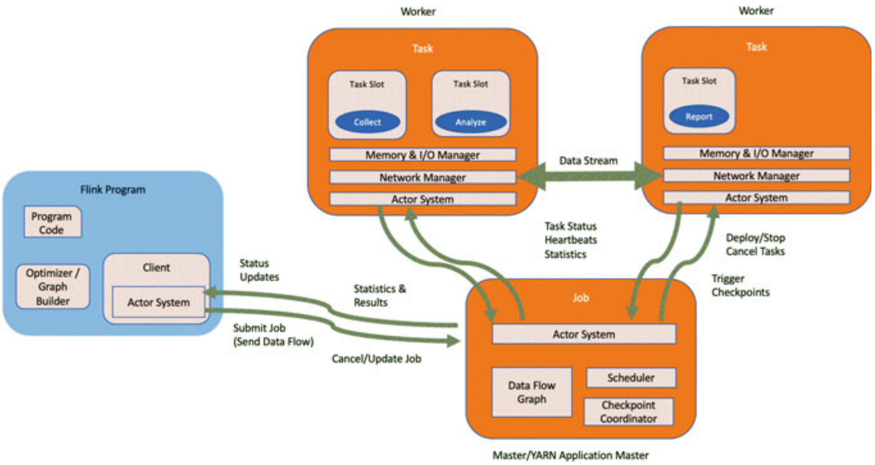


Fig. 8 Streaming architecture

4 Process Workflow

The Flink runtime consists of two types of processes: [10] a JobManager and one or more TaskManagers. The JobManager is the coordinator of the Flink system, while the TaskManagers are the workers that execute parts of the parallel programmes. Each task has a thread. For the experimental setup a test bed is created to stream the environmental data for every hour. The temperature and humidity values are streamed through multiple devices to demonstrate redundant data. Apache Kafka has multiple topics and several partitions to facilitate parallel processing. Flink--based consumer task named as collect consumes the topics with the message and group the data, aggregates the data according to the timestamp with the key value to remove the redundancy. The air quality data telemetry is streamed from different sensor devices with multiple parameters. This data is combined with respect to the timestamp value. The weather data is also joined to create a single stream of data for each timestamp. The analyse task is responsible for data exploration. It is a python-based programme which cleans the data like unwanted data, replaces the missing values with suitable value, removes incorrect data and extracts only the required features for analysis. This data is uploaded to the cloud for analytics. So by doing aggregation and preprocessing, we retain only the required data. This approach reduces the amount of storage required for storing unwanted and unclean data also it reduces the cost incurred by direct streaming. Next, the report task gets invoked by a thread and generates visual presentation of analysed data.

5 Experimental Setup and Results

The layers of the experimental setup have been showcased in Fig. 9. It also showcases the software tools used in the test bed creation. The Sensor Data Collection is important as there is a collection agent to collect the heterogeneous data produced from different sensors of the multi parameters that need to be measured, the collector is a single point of collection of this data the internal software pattern for implementation could be based on a singleton design pattern.

The next layer is the storage of the raw data which will need to be explored and feature engineered before it reaches the analytics layer for predictive analysis of this information.

There is a critical layer for the integration this will support the integration and interfacing with external third-party systems and software for the different functionalities to be achieved. The reporting layer is a very critical layer which reports the finding of the analysis.

A test bed is created to collect the environmental data using sensors–MQ135, DHT11, MQ131, DHT22. The data is collected using Arduino and Raspberry pi. Think Speak is an IoT platform which uses the MQTT protocol to publish the data in the platform, MQTT Flink Bridge will integrate to the streaming platform, Flink

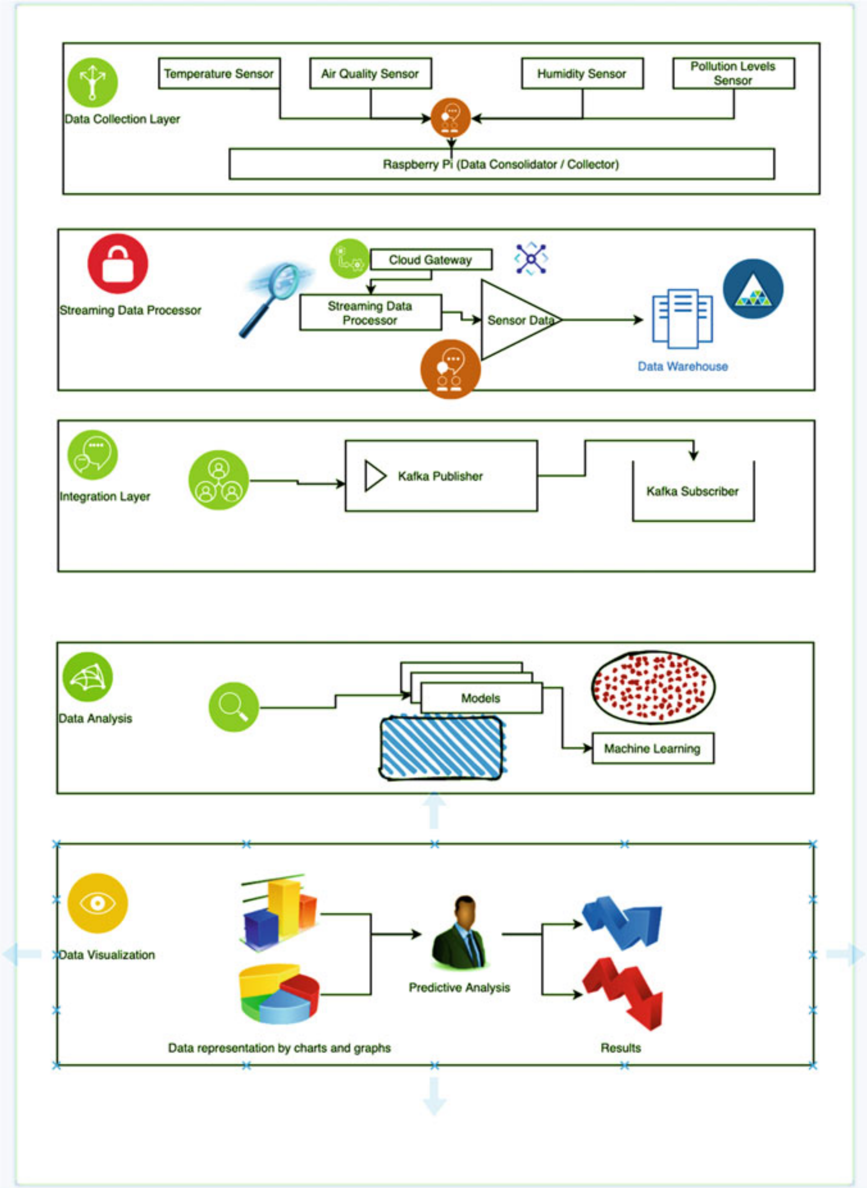


Fig. 9 Experimental setup

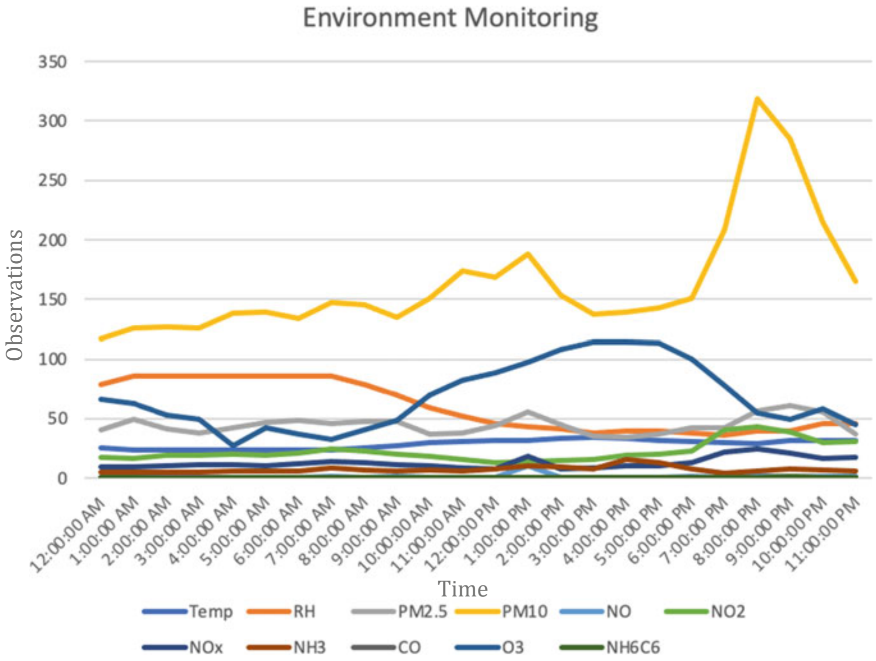


Fig. 10 Result of environment data streamed for a day

streaming consumer will receive this data and publish it using the Kafka integrator layer to the MongoDB. After which the data preprocessing takes place and forecast predictive analytics is performed. The results are visualized using charts and graphs on a dashboard. The sample data is collected in a single day for a city. It is combined as a single stream before analytics is shown in Fig. 10 since predictive forecast analysis and reporting is in development stage.

6 Conclusion

Due to the enormous amount of sensor data generated through various sources, handling that data becomes more important. It can provide a new insight to the analytics. Past few years many researches are going in this direction to address the challenges in streaming, processing and managing the data in real time. Conventional methods can't perform analytics on the sensor data due to its unstructured nature. In our approach, the real-time data is processed on the fly and only the data ready for analytics will be stored for future use. This approach reduces the cost involved in streaming and storage of sensor data. Since we use microservices based event driven architecture with Flink for streaming, we can achieve scalability and modularity. Microservices will allow the components to scale vertically and its containerized architecture allows modularity. Also, the cloud environment is horizontally scalable. By applying predictive analysis on the stored data, we can find interesting patterns, trends and insights. Our approach provides a data stream which can be used for machine learning directly by storing the data in a data store in cloud. The real-time data is stored as historical data for later use in MongoDB. It can be used to prepare a predictive model using deep learning technique to provide more insights and also we can study the relationship between the feature vectors and target variable which can be represented with charts and graphs on an analytical dashboard. This helps us to provide sensor data analytics as a service on cloud with less infrastructure and low cost. There are still few limitations when considering industry graded implementation. The data was obtained from a constrained environment which focus only on a particular city for real-time streaming. There may be challenges when we expand this to country-level monitoring. Data representation and annotation may differ from our approach. In future, we try to address this issue. There has been a gap in academic research and practical approach. Our approach tries to bridge the gap by providing a containerized loosely coupled architectural solution which makes it possible for practical implementation in industry.

References

1. A. Akanbi, M. Masinde, A distributed stream processing middleware framework for real-time analysis of heterogeneous data on big data platform: case of environmental monitoring. Sensors (Basel) (2020). https://doi.org/10.3390/s20113166. PMID: 32503145; PMCID: PMC7308861
2. IBM Cloud Native Applications. https://www.ibm.com/cloud/learn/cloud-native#toc-what-is-cl-OOTvI6Ql
3. L. Jiang, L.D. Xu, H. Cai, Z. Jiang, F. Bu, B. Xu, An IoT-oriented data storage framework in cloud computing platform. IEEE Trans. Industr. Inf. **10**(2), 1443–1451 (2014). https://doi.org/10.1109/TII.2014.2306384.(2014)
4. T. Vresk*, I. Čavrak** *, Končar elektronika i informatika, Zagreb, Architecture of an interoperable IoT platform based on microservices. MIPRO 2016, May 30-June 3, Opatija, Croatia (2016)
5. H. Nasiri, S. Nasehi, M. Goudarzi, Evaluation of distributed stream processing frameworks for IoT applications in smart cities. J Big Data **6**, 52 (2019). https://doi.org/10.1186/s40537-019-0215-2
6. S. Prajwol, I.-S. Maria, T. David, Sensor data management in the cloud: data storage, data ingestion, and data retrieval. Concurrency Computat.: Practice Exper. **30**, e4354 (2017). https://doi.org/10.1002/cpe.4354
7. B. Luca, D. Giorgio, R. Stefano, R. Matteo, A flexible IoT stream processing architecture based on microservices. Information **11**(12), 565 (2020). https://doi.org/10.3390/info11120565
8. S. García, S. Ramírez-Gallego, J. Luengo et al., Big data preprocessing: methods and prospects. Big Data Anal **1**, 9 (2016). https://doi.org/10.1186/s41044-016-0014-0
9. Ž. Indrė, G. Bogdan, Adaptive preprocessing for streaming data. IEEE Trans. Knowled. Data Eng. **1** (2012) in press. https://doi.org/10.1109/TKDE.2012.147
10. Apache flink. https://flink.apache.org/
11. W.U. Bin-feng, Design of IoT middleware based on microservices architecture. Comput. Sci. **46**(6A), 580–584 (2019)

12. S. Zhelev, A. Rozeva, Using microservices and event driven architecture for big data stream processing. in *AIP Conference Proceedings*, vol 2172. (2019), pp. 090010.https://doi.org/10.1063/1.5133587
13. M. Saqlain, M. Piao, Y. Shim, J.Y. Lee, Framework of an IoT-based industrial data management for smart manufacturing. J. Sens. Actuator Netw. **8**(2), 25 (2019). https://doi.org/10.3390/jsan8020025

FGTD: Face Generation from Textual Description

Kalpana Deorukhkar, Kevlyn Kadamala, and Elita Menezes

Abstract The majority of current text-to-image generation tasks are limited to creating images like flowers (Oxford 102 Flower), birds (CUB-200–2011), and common objects (COCO) from captions. The existing face datasets such as Labeled Faces in the Wild and MegaFace lack description while datasets like CelebA have attributes associated but do not provide feature descriptions. Thus, in this paper, we build upon an existing algorithm to create captions with the attributes provided in the CelebA dataset, which can not only generate one caption, but it can also be extended to generate N captions per image. We utilize sentence BERT to encode these descriptions into sentence embeddings. We then perform a comparative study of three models-DCGAN, SAGAN, and DFGAN, by using these sentence embeddings along with a latent noise as the inputs to the different architectures. Finally, we calculate the Inception Scores and the FID values to compare the output images across different architectures.

Keywords Generative adversarial networks · Face generation · Text to image generation · Caption creation · Natural language processing · Deep learning · Text to face generation

1 Introduction

In the field of Generative Adversarial Networks, there has been a lot of advancement since its inception in 2014 [1]. In 2016, we saw a new type of GAN architecture that could generate images from text [2]. This text-to-image problem can be considered as the reverse of a text captioning problem. Similar to text caption, text-to-image helps to understand the semantic relationship between text and image. Text-to-face synthesis is a text-to-image subdomain, aiming to synthesize face images based on human descriptions. For example, it involves describing facial features like "Black

K. Deorukhkar · K. Kadamala (✉) · E. Menezes
Father Conceicao Rodrigues College of Engineering, Mumbai, Maharashtra 400050, India

K. Deorukhkar
e-mail: kalpanas@fragnel.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_43

Hair” and “Oval Face” to the RGB pixel space as stated in [3]. While there has been a lot of advances in the text-to-image domain [4–8], the same coverage is not shared in the domain of text-to-face [3, 9].

Datasets in the text-to-image domain include the Oxford 102 Flower [10] which contains semantic descriptions of flowers and the CUB-200–2011 [11] which contains descriptions of birds while the Common Objects (COCO) [12] broadly describes the object and its context. However, none of these datasets contains any physical description of faces that are required for generating faces. While extensive research has already been conducted to create facial datasets the existing face datasets such as Labeled Faces in the wild [13] and MegaFace [14] lack description. While the CelebA [15] dataset has a large number of images, they have a list of 40 attributes associated with it. These attributes cannot be directly used as they would produce poor-quality semantic vectors. Therefore, we opted in for creating our algorithm for the dataset. Using these attributes, we generated random but meaningful sentences while avoiding grammatical errors.

The method of text-to-face follows a similar procedure when compared to text-to-image. It requires an encoder that encodes the sentences to convert them to a semantic vector. This semantic vector is then fed to the GAN with a noise vector to conditionally generate images.

While working with this methodology, it is important to generate high-quality semantic vectors. Existing solutions to generate sentence embeddings include Skip-Thought [16] which trains an encoder-decoder architecture to predict the sentences around it. The Universal Language Model [17] can be fined tuned on an existing dataset from which embeddings can be obtained. Sentence BERT [18], a modification of the pre-trained BERT [19] network can also be used for the same task. These models can be used to capture the facial features while maintaining semantic consistency.

Once, the text has been encoded, the next stage includes the generation of images. The GAN models we have studied and evaluated in this paper include the following architectures: DCGAN [20], DFGAN [4], Self-Attention GAN [21]. Keeping resource restrictions in mind, these models have successfully facilitated face generation from textual descriptions. While performing this task we need to ensure that the images generated match their textual description. For this purpose, the discriminator not only has to distinguish whether the images are real or fake but also determine whether the given image-text pair match, as given in [2]. To evaluate the performance of GANs we used three scoring metrics: (1) Inception Score [22], (2) Fréchet Inception Distance (FID) score [23], and Clean FID [24]. We then tabulated these results which could be used as a reference for other research purposes (Fig. 1).

Hence, our aim in this paper can be divided into three sections. (1) To create a dataset of facial images with rich textual descriptions. (2) To compare the different text-to-image architectures. (3) To log and evaluate these models using Weights and Biases [25].

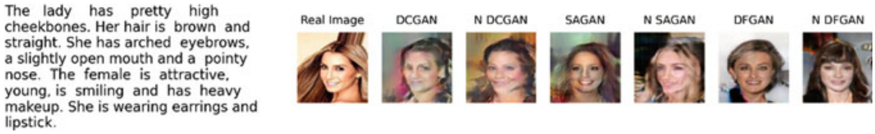

The lady has pretty high cheekbones. Her hair is brown and straight. She has arched eyebrows, a slightly open mouth and a pointy nose. The female is attractive, young, is smiling and has heavy makeup. She is wearing earrings and lipstick.
Real Image
DCGAN
N DCGAN
SAGAN
N SAGAN
DFGAN
N DFGAN

Fig. 1 Comparison of the generated images from the DCGAN, SAGAN, DFGAN for both single captions as well as N captions along with its corresponding ground truth image

2 Related Work

There has been significant research done in the field of Generative Adversarial Networks. Research has been conducted in the fields of audio [26], video [27], and text [28]. However, in this section, we shall mainly focus on two domains: (1) text-to-image (2) text-to-face.

2.1 Text-To-Image

Text-to-image synthesis was a novel concept when it was introduced by Scott Reed et al. [2] in 2016. They made use of DCGAN [20] which was conditioned on text features encoded by a hybrid character-level convolutional recurrent neural network. Since then, this field has seen significant progress in generating high-quality images which also maintain consistency with their descriptions. Zhang et al. introduced a method that adopts the method of stacking multiple generators and discriminators [6]. Later they introduced StackGAN-v2 which offered a more stable training behavior than StackGAN-v1 by jointly approximating multiple distributions [29].

With an improvement in the quality of images, the focus was now shifted to improving the similarity between the output images and the input descriptions. AttnGAN introduced the concept of an attentional generative network. By paying attention to the related terms in the natural language description, they were able to synthesize fine-grained details at different subregions of the image [7]. The concept of attention is also seen in [21] along with spectral normalization [30]. They reported that the self-attention module is effective in modeling long-range dependencies while spectral normalization helped stabilize GAN training. Meanwhile, Zizhao Zhang et al. proposed hierarchical-nested adversarial objectives inside the network. This methodology computes the matching-aware pair loss and the local image loss at different image resolutions [8]. In [4], the authors proposed a simpler and more efficient method to generate realistic and text-matching images. It generates high-resolution images by using a target aware discriminator to generate high-quality images and at the same time maintain a text-image consistency without introducing an extra network (Table 1).

Table 1 Comparison of inception score between the different text-to-image architectures on the CUB, Oxford, and COCO datasets

Models	CUB	Oxford	COCO
GAN-INT-CLS	2.88 ± .04	2.66 ± .03	7.88 ± .07
StackGAN	3.70 ± .04	3.20 ± .01	8.45 ± .03
AttnGAN	4.36 ± .03	-	25.89 ± .47
HDGAN	4.15 ± .05	3.45 ± .07	11.86 ± .18
DFGAN	5.10	-	-

2.2 Text-To-Face

Text-to-face synthesis follows a procedure similar to that of text-to-image, however, the research done in this domain is quite limited. Datasets such as Labeled Faces in the Wild [13], MegaFace [14], and CelebA [15] lack the textual descriptions that are associated with the images. Nevertheless, the CelebA dataset has about 200 k images of human faces and provides different attributes for different faces. There have been approaches to convert these attributes to meaningful descriptions [3], however, we found that the caption creation process lacks variation and clarity. Meanwhile, there exists a crowdsourced dataset called Face2Text [31] with 400 images and their descriptions, however, it is not freely available to the general public.

Some of the initial research done in this domain was conducted by Xiang Chen et al. where they introduced FTGAN. Here, they proposed to train the text encoder and the image decoder at the same time. Although this architecture performed well and produced high-quality images, they were unstable during training [9]. Meanwhile, the author of [32] made use of an LSTM network to encode the textual descriptions into a summary vector. This embedding vector is fed as an input to the generator, while for the discriminator, it is fed to its final layer. The training procedure for this GAN is similar to the ProGAN [33] paper where it increases its spatial resolutions, layer by layer. In [3], the authors used the DCGAN [20] architecture with a matching-aware discriminator [2]. They also made use of Skip-Thought Vectors [16] to encode text into embeddings. While their model did not face mode collapse, they only generated images of 64 × 64 resolution, which is quite low compared to other work [9, 32].

As a result, to provide a clear and descriptive caption we decided to create our own algorithm. We also work on the ideas presented previously to train and evaluate different architectures to produce a table of results.

3 Background

In this section, we provide an overview of caption creation, sentence vectors, GANs, and their different architectures that we have built upon. Finally, we talk about the three different scoring metrics that we use for evaluation.

3.1 Caption Creation

In [9] the authors built a dataset called the SCU-Text2Face. This dataset is based on CelebA [15] and contains 1000 images. Here, for each image, there are five descriptions, however, these descriptions were given by different people. On the other hand, [3] proposed to build an algorithm to generate captions based on the attributes given by CelebA. This involved creating six groups of facial characteristics in response to six questions that describe the face in a step-by-step manner, beginning with the facial outline and ending with the facial features that decide its appearance.

3.2 Sentence Vectors

As input to the generator, we need to provide a semantic vector of the sentence. To achieve this, we used Sentence BERT [18]. Sentence BERT fine-tunes BERT in a Siamese/triple network architecture. It is a modification of BERT [19] to derive semantically meaningful sentencing embeddings. The authors showed that the different methods for obtaining sentence embeddings using BERT gave poor results on tasks like textual similarity. They also compared the computational efficiency of SBERT to GloVe embeddings [34], InferSent [35], and Universal Sentence Encoder [36]. Compared to InferSent and Universal Sentence Encoder it is 9% and 55% faster, respectively. This is due to the smart batching strategy that is used in which sentences with similar lengths are grouped together. They are then padded to the longest element in the mini-batch, reducing the overhead in computing the padding tokens.

3.3 Generative Adversarial Networks

In GANs, we train two models simultaneously: The Generator (G) is responsible for capturing the data distribution and the Discriminator (D) is responsible for identifying if the sample came from the training data or Generator. This framework adopts the minimax two-player game strategy. The training procedure for the G is to maximize the probability of the D making a mistake and for D is to maximize the probability of assigning the correct label to both the training examples and examples generated from the generator [1].

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z))] \tag{1}$$

3.4 Model Overview

DCGAN. The architecture comprises of strided convolutions in the discriminator and fractional-strided convolutions in the discriminator. In the discriminator, it uses LeakyReLU [37] as an activation function for all layers. In the Generator, it uses ReLU [38] as an activation function in all layers except for the output layer. DCGAN is a more stable set of architecture for training Generative Adversarial Networks and the authors provide substantial evidence for the same. However, it does suffer from some forms of model instability like mode collapse [20].

DFGAN. The architecture of the proposed DFGAN comprises a generator, discriminator, and a pre-trained text encoder. DFGAN generates images with a high-resolution directly by one pair of generator and discriminator and combines the text information and visual feature maps through multiple Deep text-image Fusion Blocks (DFBlock) in UPBlocks. Armed with Matching-Aware Gradient Penalty (MA-GP) and one-way output, the model can generate more realistic and text-matching images [4].

SAGAN. In [21], the authors introduced a self-attention mechanism into convolutional GANs. They introduced this mechanism into the generator as well as the discriminator. While AttnGAN [7] used attention over word embeddings with an input sequence, it did not apply self-attention over internal model states. SAGAN, however, learns to efficiently find global, long-range dependencies within internal representations of images. In order to stabilize training, they also proposed the use of spectral normalization [30]. This technique imposes global regularization and is also computationally light.

3.5 Evaluation and Scoring

In order to evaluate the images generated, Inception Score [22] and Fréchet Inception Distances [23] are used. IS uses the Inception model [39] to calculate the KL divergence between the conditional distribution and the marginal distribution. A higher Inception Score indicates that higher quality of images has been generated as well as each image is part of a particular class indicating higher diversity. However, as mentioned in [3], the images in CelebA[15] have high intraclass similarity due to similar facial features being present thus, making inception score a poor choice for evaluation. Hence, as an additional metric for evaluation, we included FID which computes the Fréchet distance that is used to calculate the distance between the feature vectors of real and generated images. Lower scores, in this case, indicate that the images generated are more realistic. We also noted that the authors in [24] found that FID calculation involves steps that produce inconsistencies in the final metric. As a result, we included Clean FID scores as our third evaluation metric.

4 Methodology

4.1 Process Flow

The process flow for our training methodology is given in Fig. 2 . The first step is creating the dataset. The corresponding attributes of each image are converted into a textual description. The final dataset contains the image id from CelebA [15] along with its corresponding description. Each batch of the dataset contains batches of images with its respective text, providing us with an image-text pair. The training steps are as follows:

1. The text from the image-text pair is encoded into a semantic vector with the help of Sentence BERT [18].

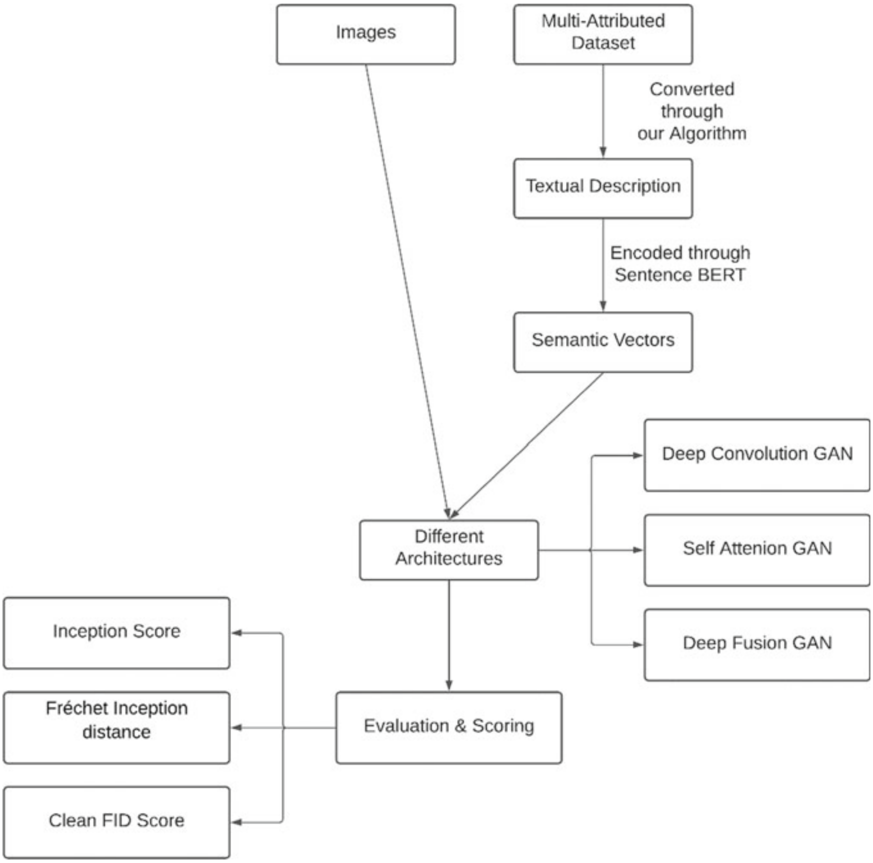


Fig. 2 Process flow for the training methodology

2. These semantic vectors are then used as the inputs to the different generator and discriminator models.
3. After each epoch, the losses are monitored with the help of Weights and Biases [25].
4. Once the training completes, each architecture is evaluated with the Inception Score [22], FID values [23], and Clean FID values [24].

4.2 Caption Creation

Similar to the authors in [3], we have divided the attributes present in CelebA [15]into six categories (see Table 2). This helps in segregating facial features into similar categories which makes it easy to add variation in their descriptions. We designed our algorithm with the aim to create "N" captions using the given list of attributes. We achieved this by randomly choosing elements from a given list of options. As a result, we were able to achieve several variations of a certain phrase which enables the network to learn a wider variety of sentences.

Each category in our algorithm has its own function. The input to each function is the corresponding attributes to the category. The function then outputs a sentence that describes these attributes in a meaningful way. Each function has its own base case and scenarios, and to introduce variations, different choices of words are used. However, while doing so we ensured that there is no compromise in the grammatical structure. When a particular category has multiple attributes associated with it, sentence structure is randomly chosen and in cases of binary attributes, an equal probability is given.

Table 2 Attribute categories for caption creation

Category	Attributes
Face structure	Chubby, Double chin, Oval face, High cheekbones
Facial hair	5 o'Clock shadow, Goatee, Mustache, Sideburns
Hairstyle	Bald, Straight hair, Wavy hair, Black hair, Blond hair, Brown hair, Gray hair, Receding hairline
Facial features	Big lips, Big nose, Pointy nose, Narrow eyes, Arched eyebrows, Bushy eyebrows, Mouth slightly open
Appearance	Young, Attractive, Smiling, Pale Skin, Heavy Makeup, Rosy Cheeks
Accessories	Wearing earrings, Wearing hat, Wearing lipstick, Wearing necklace, Wearing necktie, Eyeglasses

Algorithm 1 Caption Generation for Facial Structure

```
faceAttributes          # List of attributes present in facial structure category
isMale                  # Whether the face is of a male
function GenerateFacialStructure(faceAttributes, isMale)
  features = {Chubby: […], HighCheekbones: [...], OvalFace: […], DoubleChin: […]}
  if isMale then
    sentence = Pick a random sentence with a male gendered noun
  else
    sentence = Pick a random sentence with a female gendered noun
  if len(faceAttributes) = 1 then
    attribute = faceAttributes[0]
    sentence = sentence + random(features[attribute])
  else
    for attribute ∈ faceAttributes do
      sentence = sentence + random(features[attribute])
return sentence         # Output sentence for this category
```

Before each function is executed, we segregate and store each attribute of the image into its respective category list. This list is then passed as inputs to the function and the output descriptions are appended into a string. The final string that describes the image contains the accumulated output of all the category functions. The variation helps in generating N captions per image, thus providing us with different phrases for the same image in each loop.

To ensure equal distribution of attributes in each training batch we balance the dataset as follows:

- First, we count the number of times each attribute was present.
- Then, using the counts we calculated the attribute weights.
- Finally, for each attribute associated with an image, we calculate the sum of the attribute weight. We use this as our image weights.

4.3 Sentence Encoding

In order to convert text into embeddings, SBERT [18] is used. Each sentence from the description is passed to the model and their embeddings are stored in a list. After each sentence is parsed, the embedding list is then averaged and reshaped to (1, 768). For batches of images, the (1, 768) output is stacked to form a batch of embeddings with the shape (|B|, 768) where B is a batch set.

4.4 Network Architectures

In this paper, we have trained three models: (1) DCGAN [20], (2) SAGAN [21], and (3) DFGAN [4]. Once the embeddings have been obtained, they are passed as the inputs to the generators of these models along with a noise vector of dimension 100.

DCGAN. The generator of the DCGAN encodes the text embeddings with the help of a fully connected layer. The outputs from the fully connected layer are then concatenated with the noise vector and reshaped into a vector of shape (|B|, 356, 4, 4). The model architectures are similar to the architectures mentioned in [3]. The concatenated vector is then passed through a set of transposed convolutional layers that allows this output vector to be upsampled into an image of size 128 × 128. The discriminator consists of convolutional layers that are responsible for downsampling the image. The text embeddings are passed to a fully connected layer, expanded and then concatenated with the outputs of the second layer. This concatenated output is then passed to the final layer of the discriminator, producing outputs ranging from 0 to 1. The learning rates for the generator and the discriminator are 0.0002 and 0.0001, respectively. The Adam optimizer [40] for the generator as well as the discriminator is set with $\beta_1 = 0.5$ and $\beta_2 = 0.5$ (Fig. 3).

SAGAN. The SAGAN has a slightly different architecture as compared to the architecture in [21]. Here, the generator has two fully connected layers with ReLU activations. The text embeddings pass through these layers reducing from vectors of length 768 to 256 and then finally to a vector of length 100 as the output of the final fully connected layer. This vector is then multiplied to the input noise and reshaped to a vector of size (|B|, 100, 1, 1). It passes through layers as stated in the SAGAN paper. For the discriminator, before the second last layer, the text embeddings are passed to a fully connected layer followed by a ReLU activation layer. This is concatenated to the output vectors of the previous layers. The resulting vector is then passed to two convolutional layers after which the output of the final convolutional layer passes

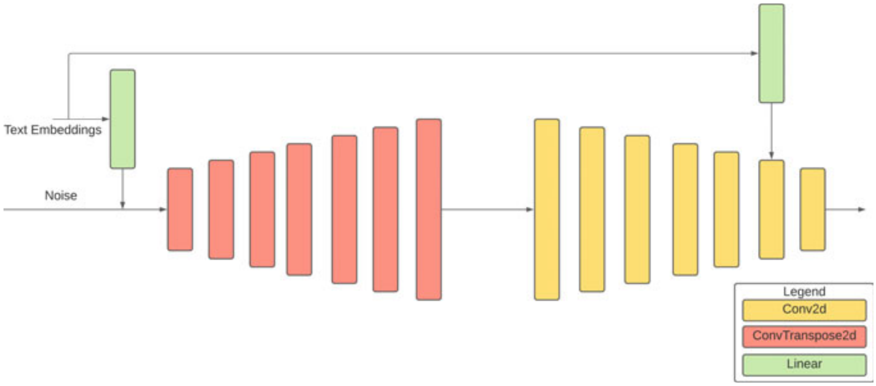


Fig. 3 Overview of deep convolution GAN which produces output images of size 128 × 128

through the sigmoid layer, generating outputs ranging from 0 to 1. Images of size 128 × 128 are generated. For the SAGAN, the learning rates of the generator and discriminator are 0.0001 and 0.0004, respectively. For both the models, the Adam optimizer [40] with $\beta_1 = 0$ and $\beta_2 = 0.9$ is used (Fig. 4).

DFGAN. The architecture of the DFGAN that we have used for training is very similar to the one used by the authors in [4]. However, the images generated are of size 128 × 128. In order to accommodate this reduced image size, the last block of the generator and the discriminator responsible for the generation and validation of 256 × 256 sized images have been omitted. A matching-aware gradient policy was added to the discriminator which helped in improving the quality of the final image. For the DFGAN, the Adam optimizer [40] is set to $\beta_1 = 0$ and $\beta_2 = 0.9$ and the learning rates for the generator and discriminator are 0.0001 and 0.0004, respectively (Fig. 5).

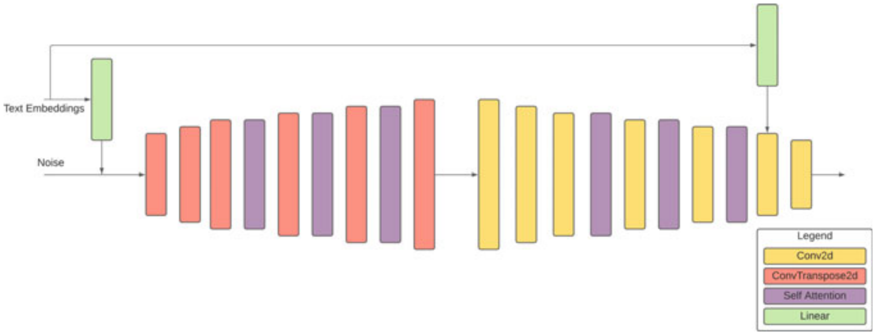


Fig. 4 Overview of self-attention GAN which produces output images of size 128 × 128

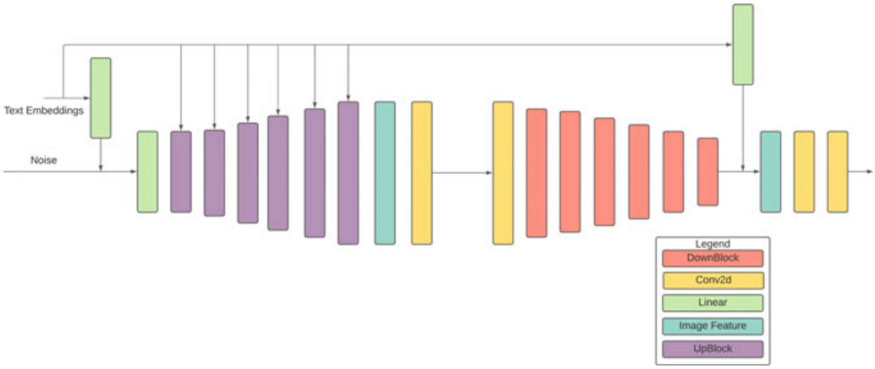


Fig. 5 Overview of deep fusion GAN which produces output images of size 128 × 128

5 Evaluation and Results

In order to calculate the inception score and FID values, we made use of tools provided in [41]. This library provides a fast and reliable evaluation of GANs in PyTorch. In addition to this, the authors in [24] provided a tool to calculate FID scores which removed the inconsistencies present in earlier FID calculations [23]. Hence, as an additional metric, we have included Clean FID scores in the evaluation table (see Table 3).

We initially trained our models on 10 k images. However, it was observed that DCGAN suffered from mode collapse and could not recover over 100 epochs. We then increased the dataset size to 20 k images. This resulted in stable image generation for all the models. The DCGAN and SAGAN models were trained for 20 epochs, while DFGAN was trained for 15 epochs. The loss patterns for the models are shown in Fig. 6. These values were logged during training with the help of weights and biases [25]. This provided us with a dashboard that helped in model versioning and evaluation.

Table 3 Performance comparison between the different architectures on 1 caption and 5 captions datasets

	GAN	Inception score	Fréchet inception distance	Clean FID
1 Caption	DCGAN	2.840 ± 0.062	87.146	87.580
	SAGAN	2.342 ± 0.029	114.512	115.256
	DFGAN	2.865 ± 0.041	109.140	106.453
5 Caption	DCGAN	2.732 ± 0.055	90.268	90.331
	SAGAN	2.855 ± 0.054	95.052	95.656
	DFGAN	3.455 ± 0.075	88.748	87.462

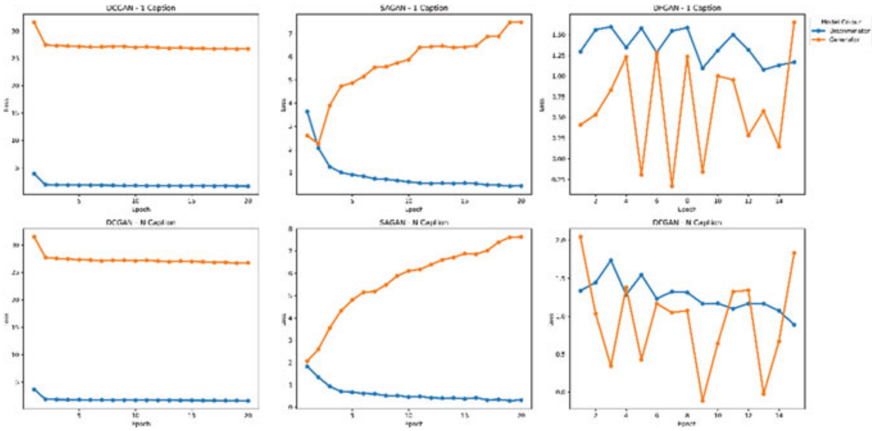
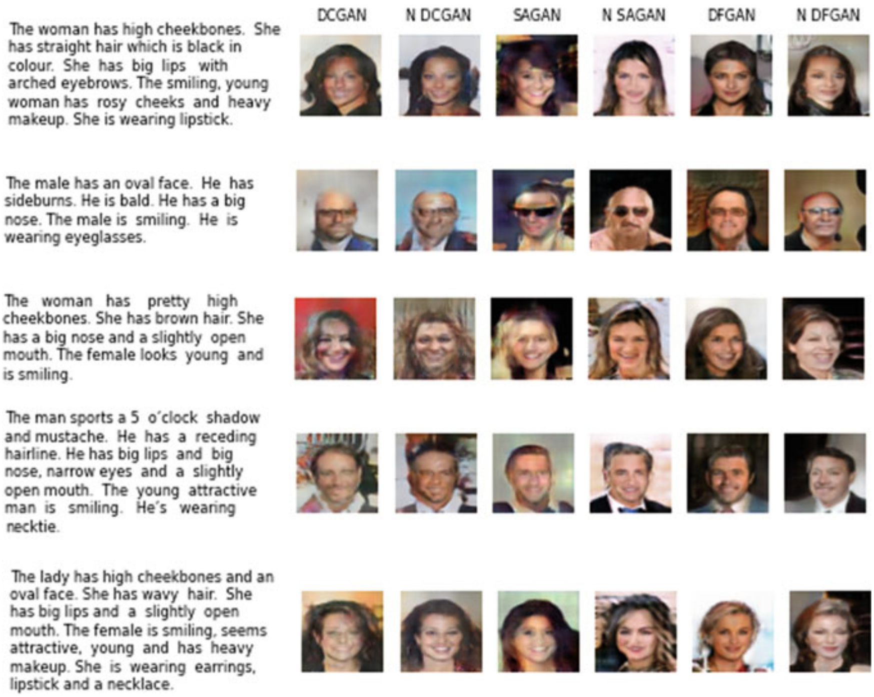


Fig. 6 Loss versus epoch values comparison for the different architectures



The woman has high cheekbones. She has straight hair which is black in colour. She has big lips with arched eyebrows. The smiling, young woman has rosy cheeks and heavy makeup. She is wearing lipstick.

The male has an oval face. He has sideburns. He is bald. He has a big nose. The male is smiling. He is wearing eyeglasses.

The woman has pretty high cheekbones. She has brown hair. She has a big nose and a slightly open mouth. The female looks young and is smiling.

The man sports a 5 o'clock shadow and mustache. He has a receding hairline. He has big lips and big nose, narrow eyes and a slightly open mouth. The young attractive man is smiling. He's wearing necktie.

The lady has high cheekbones and an oval face. She has wavy hair. She has big lips and a slightly open mouth. The female is smiling, seems attractive, young and has heavy makeup. She is wearing earrings, lipstick and a necklace.

Fig. 7 Faces generated by the different architectures from a given text prompt

The output images (Fig. 7) show a close resemblance to their respective captions. For example, in the first row characteristics like high cheekbones, young, and smiling can be identified. In row 2 features like bald, oval face, and wearing sunglasses are clearly visible in the images. Attributes like smiling with a slightly open mouth are identifiable in rows 3 and 5. The architectures have also been able to generate other accessories like a necktie which is pretty evident in the images on row 4.

6 Conclusion and Future Work

In our work, we presented an algorithm to generate captions based on the CelebA dataset. This algorithm provides grammatically correct and meaningful sentences. We used these captions and tested different architectures of Generative Adversarial Networks to dive deeper into the lesser addressed domain of text-to-Face. Due to the imbalanced distribution of the images in the CelebA dataset, images that contain less than 5 attributes or more than 12 attributes in their descriptions are under-represented. As a result, the images generated for these descriptions do not share the same quality when compared to the other images.

We found that by using SBERT, we were able to generate high-quality semantic vectors compared to the previous attempts using skip-thought vectors. From the three architectures, we identified that Self-Attention GANs and DFGANs produced higher quality images, however, due to the complex nature of their model architectures, they take more time to train when compared to the DCGAN. We also found that, for a given input, the DCGAN would generate batches of similar images, while the DFGAN and SAGAN would generate a wider variety of images while being semantically correct. We believe it is due to the concatenation strategy of the noise with the output of the fully connected layer.

We then evaluated the images generated by the models with the help of three metrics, Inception Score, FID, and the newly published Clean FID scores. Finally, we tracked our training runs with the help of Weights and Biases as a step toward reproducibility.

We believe that this work can be further improved by:

1. Introducing a better dataset balancing strategy that considers very short and extremely long descriptions.
2. Increasing the training steps for these models.
3. Extending the resolution of images to 256 × 256, 512 × 512 or further.
4. Using a transformer-based model like DALL-E [42].

Ethics

We recognize how important of a role ethics plays in the development of AI. Datasets like CelebA have an unbalanced distribution of attributes, potentially leading to bias. For this reason, we have balanced the dataset to ensure that all attributes are well represented to eliminate bias. We also recognize that certain attributes like “attractive” are subjective. However, we are considering these labels as features for our model and not as a beauty standard.

Acknowledgements We would like to thank our mentor Prof. Kalpana Deorukhkar for her constant support and guidance throughout the project.

References

1. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks (2014). arXiv preprint arXiv:1406.2661
2. S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, H. Lee, Generative adversarial text to image synthesis. in *International Conference on Machine Learning* (PMLR, 2016), pp. 1060–1069
3. O.R. Nasir, S.K. Jha, M.S. Grover, Y. Yu, A. Kumar, R.R Shah, Text2FaceGAN: face generation from fine grained textual descriptions. in *2019 IEEE Fifth International Conference on Multimedia Big Data* (BigMM) (IEEE, 2019), pp. 58–67
4. M. Tao, H. Tang, S. Wu, N. Sebe, X.Y. Jing, F. Wu, B. Bao, Df-gan: deep fusion generative adversarial networks for text-to-image synthesis (2020). arXiv preprint arXiv:2008.05865

5. T. Hinz, S. Heinrich, S. Wermter, Semantic object accuracy for generative text-to-image synthesis (2019). arXiv preprint arXiv:1910.13321
6. H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D.N. Metaxas, Stackgan: text to photo-realistic image synthesis with stacked generative adversarial networks. in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 5907–5915
7. T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, X. He, Attngan: fine-grained text to image generation with attentional generative adversarial networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 1316–1324
8. Z. Zhang, Y. Xie, L. Yang, Photographic text-to-image synthesis with a hierarchically-nested adversarial network. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 6199–6208
9. X. Chen, L. Qing, X. He, X. Luo, Y. Xu, FTGAN: a fully-trained generative adversarial networks for text to face generation (2019). arXiv preprint arXiv:1904.05729
10. M.E. Nilsback, A. Zisserman, Automated flower classification over a large number of classes. in *2008 Sixth Indian Conference on Computer Vision, Graphics and Image Processing* (IEEE , 2008), pp. 722–729
11. C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie, The caltech-UCSD birds-200–2011 dataset (2011)
12. T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: common objects in context. in *European Conference on Computer Vision* (2014), pp. 740–755
13. G. Huang, M. Mattar, H. Lee, E.G. Learned-Miller, Learning to align from scratch. in *Advances in Neural Information Processing Systems* (2012), pp. 764–772
14. I. Kemelmacher-Shlizerman, S.M. Seitz, D. Miller, E. Brossard, The megaface benchmark: 1 million faces for recognition at scale. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 4873–4882
15. Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild. in *Proceedings of the IEEE International Conference on Computer Vision* (2015), pp. 3730–3738
16. R. Kiros, Y. Zhu, R. Salakhutdinov, R.S. Zemel, A. Torralba, R. Urtasun, S. Fidler, Skip-thought vectors (2015). arXiv preprint arXiv:1506.06726
17. J. Howard, S. Ruder, Universal language model fine-tuning for text classification (2018). arXiv preprint arXiv:1801.06146
18. N. Reimers, I. Gurevych, Sentence-bert: sentence embeddings using siamese bert-networks (2019). arXiv preprint arXiv:1908.10084
19. J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding (2018). arXiv preprint arXiv:1810.04805
20. A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks (2015). arXiv preprint arXiv:1511.06434
21. H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial networks. in *International Conference on Machine Learning* (PMLR, 2019), pp. 7354–7363
22. T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans (2016). arXiv preprint arXiv:1606.03498
23. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium (2017). arXiv preprint arXiv:1706.08500
24. G. Parmar, R. Zhang, J.Y. Zhu, On buggy resizing libraries and surprising subtleties in FID calculation (2021). arXiv preprint arXiv:2104.11222
25. L. Biewald, Experiment tracking with weights and biases (2020). https://www.wandb.com/
26. J.Y. Liu, Y.H. Chen, Y.C. Yeh, Y.H. Yang, Unconditional audio generation with generative adversarial networks and cycle regularization (2020). arXiv preprint arXiv:2005.08526
27. E. Kahembwe, S. Ramamoorthy, Lower dimensional kernels for video discriminators. Neural Netw. **132**, 506–520 (2020)
28. J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, J. Wang, J, Long text generation via adversarial training with leaked information. in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32(1) (2018)

29. H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D.N. Metaxas, Stackgan++: realistic image synthesis with stacked generative adversarial networks. IEEE Trans. Pattern Anal. Mach. Intell. **41**(8), 1947–1962 (2018)
30. T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks (2018). arXiv preprint arXiv:1802.05957
31. A. Gatt, M. Tanti, A. Muscat, P. Paggio, R.A. Farrugia, C. Borg, K.P. Camilleri, M. Rosner, L. Van der Plas, Face2Text: collecting an annotated image description corpus for the generation of rich face descriptions (2018). arXiv preprint arXiv:1803.03827
32. A. Karnewar, blog:https://medium.com/@animeshsk3/t2f-text-to-face-generation-using-deep-learning-b3b6ba5a5a93 Last accessed 26 June 2020
33. T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of gans for improved quality, stability, and variation v. arXiv preprint arXiv:1710.10196
34. J. Pennington, R. Socher, C.D. Manning, Glove: global vectors for word representation. in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP) (2014), pp.1532–1543
35. A. Conneau, D. Kiela, H. Schwenk, L. Barrault, A. Bordes, Supervised learning of universal sentence representations from natural language inference data (2017). arXiv preprint arXiv:1705.02364
36. D. Cer, Y. Yang, S.Y. Kong, N. Hua, N. Limtiaco, R.S. John, N. Constant, G.-M. Céspedes, S. Yuan, C. Tar, Y.H. Sung, Universal sentence encoder (2018). arXiv preprint arXiv:1803.11175
37. A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models. Proc. ICML **30**(1), 3 (2013)
38. V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines. In: Icml (2010)
39. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2818–2826
40. D.P. Kingma, J. Ba, Adam: A method for stochastic optimization (2014). arXiv preprint arXiv:1412.6980
41. A. Obukhov, M. Willylulu, S. Zhydenko, J. Kyl, E.Y.-J. Lin, High-fidelity performance metrics for generative models in PyTorch (2021). https://github.com/toshas/torch-fidelity. Last Accessed 1 March 2021
42. A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, Zero-shot text-to-image generation (2021). arXiv preprint arXiv:2102.12092

Two Levels of Security for Protection of Images Copyright

Sheimaa A. Hadi, Suhad A. Ali, and Majid Jabbar Jawad

Abstract Due to illegal manipulation and image processing attacks, digital image copyright protection is receiving a significant research attention**.** This paper has introduced a blind invisible watermarking method to protect the copyright of digital color images. This method is based on the combination of digital transforms (DWT, DCT) in the frequency domain. The embedding process involved in this method is based on partitioning the host image into 16 × 16 non-overlap blocks, and the chaotic maps are used to generate random numbers in order to choose the appropriate blocks for the inclusion process involved in the purpose of increasing the security of proposed system. As for the extraction process, it is carried out in a way that does not require the presence of original image but rather follows the same embedding protocol to extract the embedded and encrypted watermark. To raise the security level, a hybrid encryption method has been proposed by using the chaotic map and DNA coding for encrypting the watermark before embedding it. Experimental results evident the good imperceptibility. In addition, the proposed method effectively resists common "image processing attacks."

Keywords Discrete cosine transform · Discrete wavelet transform · Chaotic map · DNA encoding · Normalized correlation (NC) · Image watermarking · Copyright protection

1 Introduction

Nowadays, sending and receiving information is a part of everyday life. With the development of communication and technology, the popularity of the Internet spreads everywhere [1–3]. Everyone can access and share this information via multimedia (images, video, and audio) as the volume of media has started to increase dramatically on the Internet [4]. The advantages of digital multimedia storage and processing

S. A. Hadi · S. A. Ali (✉) · M. J. Jawad
Department of Computer Science, College of Science for Women, Babylon University, Hillah, Iraq

S. A. Hadi
e-mail: shaymaa.hadi@student.uobabylon.edu.iq

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_44

have shown explosive growth especially after the introduction of modern multimedia technologies to the information market [5]. Since, the media is digital and has the ability to store huge amount of information, which can be easily processed. Furthermore, there are several factors that have contributed to the improvement of the quality of information services, such as data transmission, sharing, and ease of copying [6].

A digital image transmitted over insecure channels may be exposed to risks such as attacks, where attackers have free and easy access to digital images over the Internet [7]. Medical images, personal images, corporate brand images, and military images are all the examples of digital images. This is why, while transferring them, their safety, security, and proof of ownership become critical, and measures must be available to assure these methods. Digital watermark is considered as an alternative solution to prevent illegal copying practice [8]. To guarantee the copyright of these images, three requirements must be met in the image watermarking for copyright protection system; they are imperceptibility, robustness, and security. Imperceptibility is achieved when the embedding process does not reduce the quality of the host image. The robustness is fulfilled, when resistant embedded watermark modified by the attackers can extract with acceptable quality. Finally, security is achieved when it becomes more difficult for the attacker to know how to embed the watermark and extract it [9]. Watermarking is one of the most important technologies to satisfy all the above requirements. In general, watermarking can be applied in two domains either in spatial domain or in frequency domain. In spatial domain, the watermark is inserted in the pixels of image directly, while in frequency domain, the watermark is inserted in the coefficients obtained by frequency transformation of the image [10].

This paper presents a robust blind watermarking method depending on the properties of DWT and DCT transforms for choosing the best locations. The chosen location is used for embedding bit watermark with low distortion and good robustness against attacks. The structure of this research paper is arranged as follows. Section 2 explains the relevant work. The proposed watermarking scheme is explained in Sect. 3. Experimental results are described and explained in Sect. 4. Section 5 describes the conclusions.

2 Previous Works

In this section, several watermarking algorithms which proposed to deal with the protection of color images are highlighted.

In [11], Adel Jalal Yousif suggested a non-blind image watermarking system based on YCbCr color space and discrete cosine transform. The hiding operation of watermark is done in the Y-component image. Use the scale factor in the proposed scheme in order to obtain the desired trade-off between perceptibility and robustness. The method was tested against a number of attacks to find out the efficiency of the proposed method.

In [12], Fauzi et al. suggested a non-blind watermarking technique depending on applying the DCT transform on the color host image. The embedding activity is done by dividing the channel into 8*8 non-overlap blocks. Then, DCT is applied to blocks and inserting watermark bit in DC coefficients. To obtain a better robustness, it is recommended to insert the watermark in the blue channel.

In [13], Mohammad Moosazadeh suggested a watermarking system based on "JPEG YCbCr color space." The embedding activity depends on the relationship between the DCT coefficients. The suggested method uses "TLBO"; which automatically determines the embedding parameters and appropriate position for embedding the watermark. This suggested method is tested and introduced good imperceptibility of watermarked images and good robustness various attacks.

In [14], Xiaochen Yuan proposed a watermarking method for protection of color images based on "quaternion discrete Fourier transform (QDFT)." The qQuaternion QR (QQR) decomposition" is used in the proposed method. Several procedures are done in the suggested method. Firstly, the "QDFT and QQR decomposition" are implemented on the host image, respectively. Secondly, dividing the matrix generated by the "(QQR)" into blocks and calculate the entropy. The block with a high value of entropy is chosen to embed the secret bits. Finally, the selected blocks are used to embed the bits of secret watermark depending on the quantization index modulation scheme. The experimental results illustrate that the suggested approach obtains the highest robustness against several attacks such as rotation, scaling, salt and pepper, median filter, and JPEG compression.

In [15], Jun et al. presented image watermarking method based on "discrete cosine transform (DCT)" and "just noticeable distortion (JND)." Color complexity features and orientation diversity are taken into regard. In the suggested method, the host is converted to YCbCr color space. The embedding operation is done in the Y-channel. The Y is splitting into non-overlapped blocks with size 8 × 8. Then, the bits of the watermark are embedded in the blocks by modifying the DCT coefficients based on the perceptual JND model.

3 The Proposed Color Image Watermarking System

The proposed system aims to preserve the copyright of color image. The proposed system consists of two procedures. The first one is implemented on the sender side, while the second one is implemented on the receiver side.

3.1 The Sender Side

In the sender side, two procedures are applied that include the watermark encryption and embedding.

3.1.1 Watermark Encryption Procedure

To enhance and increase the security aspects of the system, an encryption method based on DNA philosophy is used to encrypt the watermark (secret logo) before including it into the original (host) image as illustrated in [16].

3.1.2 Blind Watermark Embedding Procedure

This procedure explains embedding the encrypted watermark in the image. Figure 1 illustrates the embedding activity.

The host color image consists of three channels, namely R, G, and B. So, it is divided into three bands, namely R_img, G_img, and B_img. The embedding process is done by embedding two bits of watermark in the R_Img and G_img and one bit in the B_img image. There are two steps for embedding activity, which can be listed as follows.

A. Scrambling of the Block

In this step, the band is divided into non-overlapping 16*16. In order to increase the security of the proposed system, a scrambling process is done on the blocks according to Eq. (1):

$$\text{Newb} = (K \times \text{length}(w)) \bmod \text{orginalb} + 1 \tag{1}$$

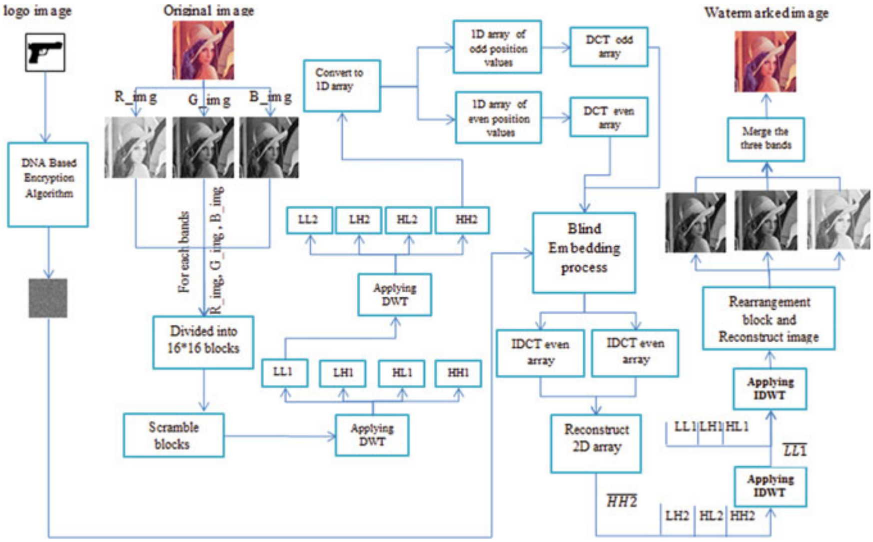


Fig. 1 Embedding watermark activity

"where (K) is a prime number used for blocks scrambling, (orginalb) is the total number of blocks, (length(w)) is a counter secret watermark, and (Newb) is the new obtained block number."

For more robustness, each block is converted from the spatial domain to the frequency domain by applying a combination of DWT and DCT. Two levels of a discrete wavelet transform (DWT) are applied to the selected block. The first level DWT is implemented on the block to obtain four sub-bands, namely ("LL1, LH1, HL1, and HH1"). Then, the DWT is applied on the LL1 sub-band in the second level to obtain the new "four sub-bands (LL2, LH2, HL2, and HH2)." To ensure more imperceptibility, the watermark bits are embedded in band HH2. The HH2 sub-band is converted into a 1D array and then separated into two 1D arrays containing the values of odd locations (i.e., O_array) and the other containing the values of even locations (i.e., E_array). Then, a discrete cosine transform DCT is applied to each array.

B. Embedding Process

The embedding process is done depending on equations given in (2) and (3):

$$O_{\mathrm{dct(idx)}} = \left(\frac{1}{2} \times \left(O_{\mathrm{dct(idx)}} + E_{\mathrm{dct(idx)}}\right)\right) + (\mathrm{fac} \times \mathrm{sce}) \tag{2}$$

$$E_{\mathrm{dct(idx)}} = \left(\frac{1}{2} \times \left(O_{\mathrm{dct(idx)}} + E_{\mathrm{dct(idx)}}\right)\right) + (E \times \mathrm{sce}) \tag{3}$$

For R_img and G_img, idx = 3
For B_img, idx= 3,4

Where O_dct and E_dct represent the odd and the even array after applying DCT on them; Sec is equal to 1 or −1 depending on the value of watermark as shown in (4):

$$\mathrm{Sec} = \begin{cases} -1 \text{ if } W(j) = 0 \\ 1 \quad \text{otherwise} \end{cases} \tag{4}$$

fac represents a scaling factor, and this value is changing depending on the smoothness degree of the host image block to ensure impressibility. To measure the block smoothness degree, the standard division (std) of the block is computed to select the appropriate value of the scaling factor (fac) according to Eq. (5):

$$\mathrm{STD} = \sqrt{\frac{\sum_{i=1}^{n} (x - x')^2}{n}} \tag{5}$$

The selected blocks are divided into three sets of bocks (smooth, partial smooth, and coarse) by comparing the standard division of the block with a threshold (T) according to the following:

$$\text{if td 18 the block is smooth}$$

$$\text{else if std} \geq (18)\&\&\ \text{Std} < (24)$$

$$\text{the block is partially smooth}$$

$$\text{else}$$

$$\text{the block is coarse}$$

According to the above comparison, the high value is assigned to fac when the block is coarse, while the low value is assigned when the block is smooth.

For smooth block, fac = 20
For partially smooth block, fac = 24
For coarse block, fac = 30

After each embedding process, IDCT is applied on the watermarked block to obtain the sub-band (HH2). Finally, the IDWT is applied, then rearranging the scrambled blocks. The watermarked image is resulted by combining the three watermarked images (R_img, G_img, and B_img).

3.2 *The Receiver Side*

On the receiver side, the recipient, in turn, takes applied two procedures including extraction and decryption processes.

3.2.1 Extraction Procedure

The extraction procedure is shown in Fig. 2. The secret watermark is extracted by performing several steps.

The recipient divides the watermarked image into three images R_img, G_img, and B_img. For each band, the same steps are applied. The image band is divided into non-overlapping blocks of size 16 × 16. Next, these blocks are scrambled according to the previously mentioned Eq. (8) and stored in an array. The embedded watermark pixels are extracted from blocks that hid these secret pixels by applying a DWT to the block. It will produce four bands (LL1, LH1, HL1, and HH1) for each block, then applying the second level of DWT on the LL1 to result in the other four bands (LL2, LH2, HL2, and HH2). The HH2 coefficients obtained from the second level of DWT are saved in an array to be used in extraction process. An array will be divided into two 1D arrays. The first is one for saving the coefficients in the odd locations

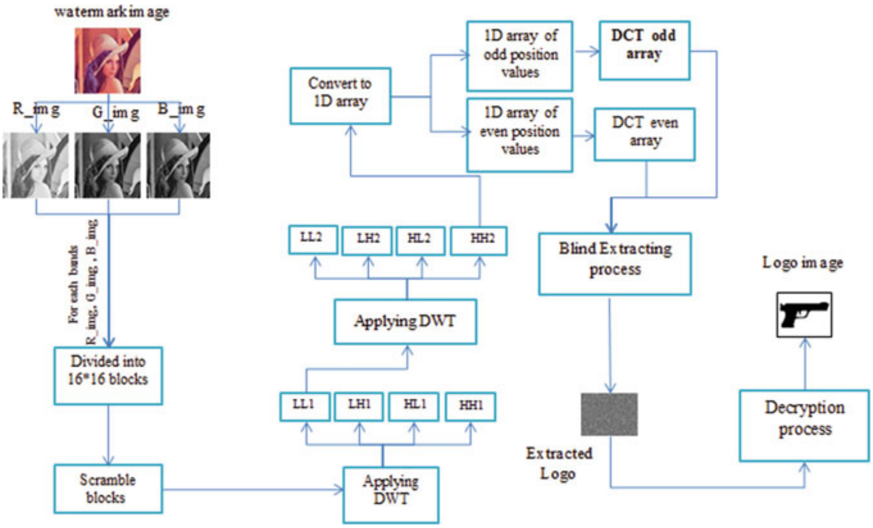


Fig. 2 Extracting of watermark activity

(O_array), and the second is for saving the coefficients in the even locations value (E_array). To complete the extraction process, the DCT is applied on each of these arrays to obtain (O_dct, E_dct) arrays. Finally, the extraction (Eq. (6)) is implemented to obtain the pixels' values of the inserted watermark (encrypted watermark).

$$\text{Wbit}(i) = \begin{cases} 1 \text{ if } O_{\text{dct(idx)}} > E_{dct(idx)} \\ 0 \text{ otherwise} \end{cases}, \tag{6}$$

For R_img band and G_img band, idx = 3.
For B_img band, idx = 3, 4

The extraction process is done by extracting one bit from the block of (R_img), one bit from the block of (G_img), and two bits from the block of (B_img).

3.2.2 Decryption of the Extracted Watermark

To obtain the embedded watermark image, a decryption process must be performed on the extracted watermark image as illustrated in [16].

4 Experimental Results

This discusses the obtained results after applying the suggested system.

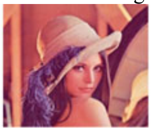
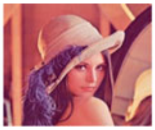
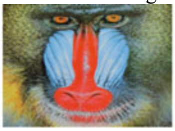
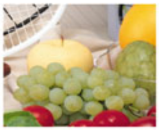
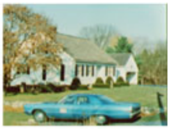
"Original Image"	**"Watermarked Image"**	**"Original Image"**	**"Watermarked Image"**
Lena Image 	PSNR=46.5789 	Baboon Image 	PSNR=43.2661
Pepper Image 	PSNR=47.4624 	Airplane Image 	PSNR=47.3058
Fruits Image 	PSNR=46.6396 	House Image 	PSNR=47.1028 

Fig. 3 PSNR values of test images

4.1 Imperceptibility of Watermarked System

The PSNR measure is used to evaluate imperceptibility after implementing the embedding procedure. The computation of the PSNR metric is very easy and fast as shown in the formula for calculating the PSNR in Eq. (7).

$$\mathrm{PSNR} = 10\log_{10}\frac{\left(2^{L}\right)^{2}}{\mathrm{MSE}} \tag{7}$$

where L represents the number of bits required to represent image pixels.

PSNR is calculated using the original and watermarked images. Figure 3 shows the watermarked images and PSNR values for different standard images when applying blind watermarking system. The cover image has size 512 × 512, and the watermark has size 1024 bits.

4.2 Robustness Measures of the Proposed System Against Attacks

To measure the robustness of the proposed blind watermark system, the test watermarked images must be subjected to various types of attacks. Then, the encrypted watermarked image, which previously included in the cover image, is extracted. Two parameters are used to define the degree of robustness of the proposed embedding method, namely “normalized correlation (NC) and bit error rate (BER).”NC and BER are calculated between the original and the extracted watermarks according to the following equations:

$$\mathrm{NC} = \frac{\sum_{i=1}^{x} \sum_{i=1}^{y} w(i, j)w'(i, j)}{\sum_{i=1}^{x} \sum_{j=1}^{y} w(i, j)^2} \tag{8}$$

BER computes the error rate between the recovered watermark and the original watermark. BER value can be calculated according to Eq. (9).

$$\mathrm{BER} = \frac{1}{X * Y} \sum_{i=1}^{X} \sum_{j=1}^{Y} |w(i, j) - w^{'}(i, j)| * 100\% \tag{9}$$

Where w and w' represent cover image and watermark, respectively. The smaller BER value reflects the better robustness.

Figure 4 shows the NC values for different standard images after attacking the watermarked images with different attacks.

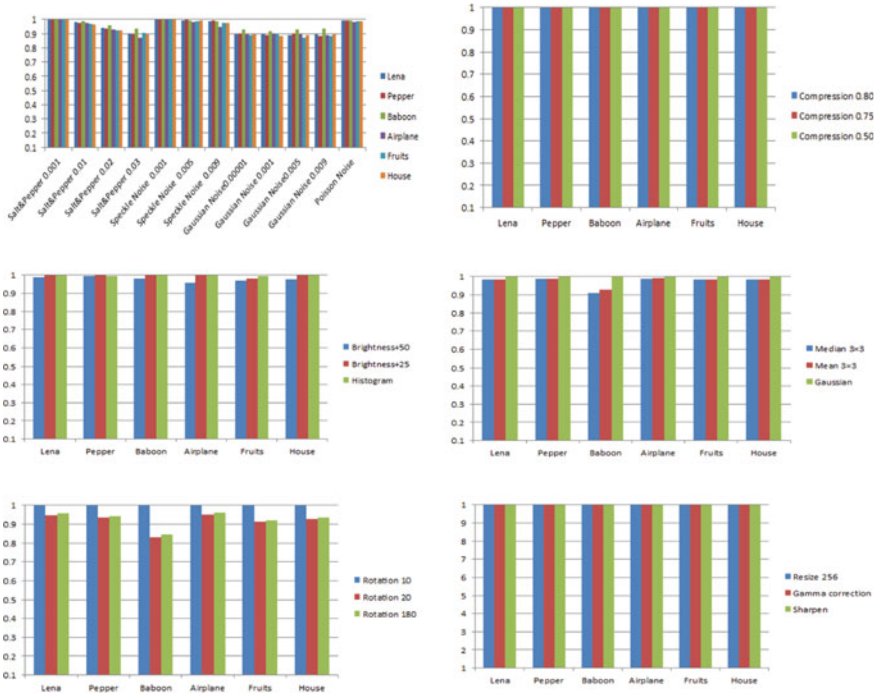


Fig. 4 Different attack

5 Conclusions

In this paper, we suggested an invisible robust watermarking system for copyright protection of the digital color image. A combination of the chaotic map and DNA encoding is used for encrypting before inserting the watermark to add a layer of security where watermark retrieving is performed without the need for the original image. The parameters of the quadratic chaotic map are used as secret keys. The security of the proposed scheme is also confirmed as only the one with the correct parameters can extract the watermark. The embedding process is implemented into the frequency domain by applying both DWT and DCT transforms, respectively, on the cover image. The embedding process uses a scaling factor to increase robustness against different attacks. The large value of the scaling factor can cause a blocky effect in smooth blocks while obtaining good robustness against attacks. Therefore, in this paper, it adopts a technique based on the standard deviation value to choose the variable value of the scaling factor to make a trade-off between robustness and imperceptibility. The proposed system is tested against different types of attacks and provides very good strength. Also, the system fulfilled the security and imperceptibility requirements for any watermarked system.

References

1. R. Sinhal, I.A. Ansari, “A multipurpose image watermarking scheme for digital image protection. Int. J. Syst. Assur. Eng. Manag. **11**, 274–286 (2020). https://doi.org/10.1007/s13198-019-00855-0
2. V. Suma, A Novel Information retrieval system for distributed cloud using hybrid deep fuzzy hashing algorithm. JITDW **2**(03), 151–160 (2020)
3. T. Vijayakumar, R. Vinothkanna, Retrieval of complex images using visual saliency guided cognitive classification. J. Innov. Image Process (JIIP) **2**(02), 102–109 (2020)
4. T. Sumanth, S. Vijay Harisudan, M.N. Tarun Kumar, K. Geetha, A new audio watermarking algorithm with DNA sequenced image embedded in spatial domain using pseudo-random locations, in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, (Coimbatore, 2018), pp 1813–1817. doi: https://doi.org/10.1109/ICECA.2018.8474758.
5. M.R. Khosravi, H. Rostami, S. Samadi, Enhancing the binary watermark-based data hiding scheme using an interpolation –based approach for optical remote sensing images. Int. J. Agric. Environ. Inf. Syst. **9**(2) (2018)
6. P.A. Mendez, R. James, R James," An imperceptible blind image watermarking scheme for image authentication using DNA encoding and multi-resolution wavelet decomposition. Int. J. Eng. Inventions **4**(12), 23–30 (2015)
7. R. Rmana, V.N.K.P. Munaga, D. Sreenivas Rao, Robust digital watermarking of color images under noise attacks. Int. J. Rec. Trends Eng. **1**(1), 334
8. N.F. Mohammed, S.A. Ali, M.J. Jawad, Biometric-based medical watermarking system for verifying privacy and source authentication. Kuwait J. Sci. **47**(3) (2020)
9. R. French-Baidoo, D. Asamoah, S.O. Oppong, Achieving confidentiality in electronic health records using cloud systems **1**, 18–25 (2018). https://doi.org/10.5815/ijcnis.2018.01.03
10. M. Moosazadeha, G. Ekbatanifard, An improved robust image watermarking method using DCT and YCoCg-R color space. Optik—Int. J. Light Electron Opt. **140**, 975–988 (2017)

11. A.J. Yousif, A discrete cosine transform based watermarking scheme for color image using YCBCR space, J. Eng. Sustain. Dev. **22**(06) (2018). doi.org/https://doi.org/10.31272/jeasd.2018.6.1
12. F.A. Rafrastara, A.V. Hadinata, D.R.I.M. Setiadi, E.H. Rachmawanto, C.A. Sari, Copyright Embedding Analysis in Color Image Channel based on Non-Blind DCT Method, in *International Conference on Information and Communications Technology (ICOIACT)* (2019), pp. 185–190. https://doi.org/10.1109/ICOIACT46704.2019.8938427
13. M. Moosazadeh, G. Ekbatanifard, A new DCT-based robust image watermarking method using teaching-learning-Based optimizatio. J. Inf. Secur.Appl **47**, 28–38 (2019). https://doi.org/10.1016/j.jisa.2019.04.001
14. M. Li, X. Yuan, H. Chen, J. Li, Quaternion discrete fourier transform based color image watermarking method using quaternion QR decomposition. IEEE Access **8**, 72308–72315 (2020). https://doi.org/10.1109/ACCESS.2020.2987914
15. J. Wang, W.B. Wan, X.X. Li, J. De Sun, H.X. Zhang, Color image watermarking based on orientation diversity and color complexity. Expert Syst. Appl. **140** (2020). doi.org/https://doi.org/10.1016/j.eswa.2019.112868
16. S.A. Kadhim, S.A. Ali, M.J. Jawad, Binary image encryption based on chaotic and DNA encoding. Lect. Notes Netw. Syst. **201**

Speech Enhancement Using Nonlinear Kalman Filtering

T. Namratha, B. Indra Kiran Reddy, M. V. Deepak Chand Reddy, and P. Sudheesh

Abstract Speech enhancement is today a growing necessity for a wide range of applications in which the noise-free speech signal is important and necessary for the processing to be continued. The main purpose of these speech enhancement techniques is on a higher level is to remove noise from the speech signal. The reverberation component in the corrupted speech signal is also removed using the auto-regressive techniques for better performances. In this paper, unscented Kalman filtering which is an adaptive algorithm is proposed that executes both denoising and dereverberation of the speech recorded in adverse conditions. The algorithm relies on the parameter such as mean and covariance of the state spaces created and updating the concerned measurements to provide the optimal denoised and dereverberated signal. This proposed algorithm is assessed with regard to quality of speech, intelligibility of speech and performance metrics like the figure of merit and cross correlation and is also compared with other denoising and dereverberation techniques. The trial outputs on executing the algorithm using the noisy reverberant speech exhibit the adequacy of the proposed adaptive enhancement algorithm.

Keywords Speech enhancement · Adaptive algorithm · Denoising · Dereverberation · Kalman filtering · Auto-regressive filtering · Unscented Kalman filter · Parameter estimation · Unscented transform · Parameter estimation

1 Introduction

These days, innovation is truly advancing with enormous demand, and the interest for speech enhancement frameworks is clear. Speech improvement in uproarious reverberant conditions, for the audience, is hard and testing. The speech signal is corrupted by the noise and resonation when caught utilizing an inaccessible mouthpiece [1]. A room impulse response will incorporate segments at long postponements, subsequently coming about in resonation and echoes. Reverberation is considered to be a

T. Namratha · B. Indra Kiran Reddy · M. V. Deepak Chand Reddy · P. Sudheesh (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_sudheesh@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_45

convolutive distortion that actuates big haul correlation between successive observations and can be very time-taking with a resonation time [2]. Noise and reverberation can be stationary or non-stationary and inconveniently affect both discourse quality and discourse comprehensibility [2]. Different techniques have been introduced on speech enhancement.

The Kalman filtering is one of them and is a good and dependable speech improvement algorithm. It utilizes the minimum mean square error wisely [3]. Nonetheless, admittance to clean speech and added substance commotion data for the state-space model boundaries for the greater part of the traditional KF-based speech enhancement techniques is needed. In particular, the linear prediction coefficients and the additive noise variance estimation, which is unrealistic in practical speaking to get the noisy speech [4, 5]. Also, the authors in [6] proposed that the fundamental cycle of noise reduction calculation is Kalman filtering. The underlying incentive for KF is dictated by ASS. To get higher exactness, the following calculation is proposed. From the outset, the power spectrum of clean speech is assessed from the spectrum by the KF algorithm. At that point, the acquired power spectrum is filled in for initial value, and Kalman filter calculation is rehashed. On doing this calculation, we acquired greater precision of decrease in noise. It can be repeated at 1.5–2.0 occasion times of constant by taking the noisy speech signal as an input, and fast Fourier transform (FFT) was done to get power spectrum. Using adaptive spectral subtraction (ASS), we get estimates of power spectrum, i.e., noise signal power subtracted is subtracted from mixed signal spectrum [7].

2 Related Works

As per the work done in [8], first a noisy speech signal is given as input, and this input speech signal is assumed as stationary during each frame and processed using three algorithms, which are spectral subtraction, Wiener filter and Kalman filters, and the work suggests that the spectral subtraction can be used only for stationary signals and real-time signals are non-stationary. The Wiener filter is also suitable for stationary signals but denies working on musical noise. To oversee these boundaries, the paper suggests Kalman filtering. When talking about the UKF algorithm, it was first proposed in [9, 10]. In [11], the work proposes that most approaches use the stationary AWGN assumption, but the same of colored noise is believed to be more useful for speech denoising and speech dereverberation. The Kalman filter, because of its flexibility, is widely used for signal enhancement. Kalman filter has a considerable amount of numerical complexity while dealing with colored noise. Moreover, Kalman filtering is a model-based adaptive method, where speech as well as noise is modeled as AR processes. Thus, a major issue in Kalman filtering is the estimation of the AR parameters in the presence of noise. The traditional algorithm utilizes the EM technique to repeatedly calculate the AR boundaries. Unfortunately, its computational complexity is high. The method used in our work is built on spectral subtraction for estimation of AR parameters of clean signal and corresponding noise

[12]. It is computationally efficient and can be easily implemented. The mathematical model for the algorithm of the state-space model and Kalman filter equations was formulated, and the obtained results were compared to the WF method [13, 14].

The work proposed by the authors in [15] is the computer-based algorithms which are generally used for controlling and monitoring a computer where human, digital and analog interactions occur. The cyber-physical systems (CPS) scheme is used in many areas due to its easily available and connectivity features and also offers large amount of storage and computing resources. However, the limitation of this scheme is its large energy consumption. As in [16], spectral subtraction method is applied in the estimation of parameters, musical noise appears in the enhanced speech. To acquire a Kalman filter output with better audible quality, a conceptual post-filter is set at the output of the Kalman filter to decrease the musical noise level. The perceptual filter minimizes signal distortion while constraining the noise spectrum.

3 Methodology

3.1 Flow Process

In the time domain, the distorted speech, $d_k(t)$, is given by $d_k\ (t) = C_k\ (t) * r_k\ (t) + n_k(k)$ where C$_k$ (t) is the clean speech component, r$_k$$(t)$ is the reverberant speech component, and $n_k\ (t)$ is the noise [2]. The time frame index is represented as k. The algorithm holds each time frame bit on its own. In the limits of the algorithm, k is introduced as a variable in the equations that involve multiple time frames [2]. Figure 1 explains the flow process of the algorithm.

The clean speech which is downloaded from the database is processed and is reverberated using the reverb parameters and convolution. The output of the first block in Fig. 1 is the reverberated speech with some given delay, and the magnitude of the speech changes according to the coefficient of reverberation taken [17]. The approach in Eq. 1 is used to do the reverberation process as

$$O(n) = I(n) + aO(n - d) \tag{1}$$

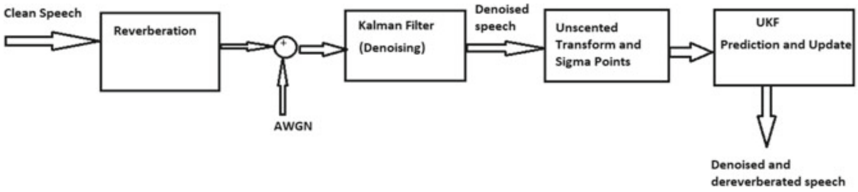


Fig. 1 Flow process

where $i(n)$ is the input audio signal, $O\ (n)$ is the output (echoed) audio signal, d is the echo delay (in samples), and alpha is the coefficient that governs the amount of echo fed back. Then, the reverberated signal is then added with a certain amount of additive white Gaussian noise as shown in Fig. 1. Here, we have the corrupted speech signal that needs to be denoised and de-reverberated.

The corrupted speech is then taken as k reduced time frames or into k smaller time frames that are of a specific period which are called the state spaces. For this process of converting the clear speech signal to state spaces, we use three different windows. They are the rectangular window, the hamming window and the Gaussian window [12]. The proposed algorithm treats each time frame or the state space on its own. Firstly, as in the third block of Fig. 1, each of these frames then undergo the unscented transform in which the sigma points of the first state space are calculated. Then, the statistical mean and covariance of the present state are calculated. Then, the two main steps of the algorithm, the time update and the measurement update steps, are done for the first state space. Being an auto-regressive algorithm, the same is applied to all the k state spaces, i.e., the set of time update equations and measurement update equations given in the following Sect. 3.2 are implemented. The detailed equations to the above algorithm are also mentioned in the Sect. 3.2.

3.2 Unscented Kalman Filtering

3.2.1 Unscented Transform

The unscented transform (UT) is a method for estimating the mean and covariance of RV that goes through a nonlinear transformation [3, 18]. Take into consideration the propagation a RV $\boldsymbol{x}$ into a function $\mathbf{y} = f(x)$. Consider $\overline{x}$ is the mean, and P_X is the covariance of RV x.

Figure 2 explains the steps in the unscented transform step in Fig. 1. The $\overline{x}$ and P_x depicted in Fig. 2 are the mean the covariance of the random variable x, respectively,

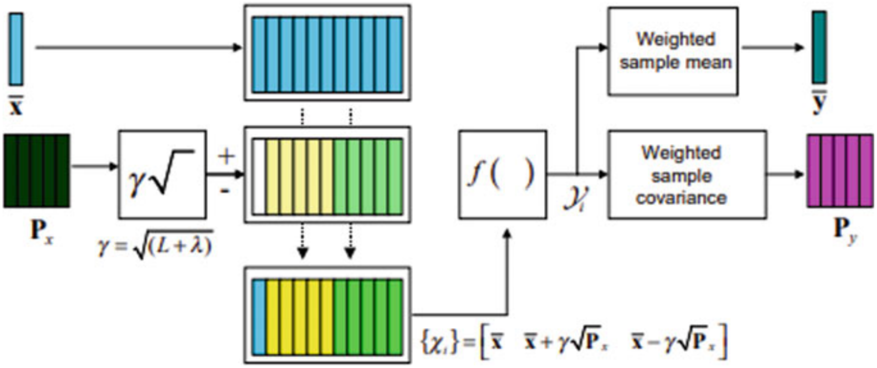

$\overline{\mathbf{x}}$
$\mathbf{P}_x$
$\gamma\sqrt{\ }$
$\gamma=\sqrt{(L+\lambda)}$
+
-
$f(\ )$
$\mathcal{Y}_i$
Weighted sample mean
Weighted sample covariance
$\overline{\mathbf{y}}$
$\mathbf{P}_y$
$\{\chi_i\}=[\overline{\mathbf{x}}\quad \overline{\mathbf{x}}+\gamma\sqrt{\mathbf{P}_x}\quad \overline{\mathbf{x}}-\gamma\sqrt{\mathbf{P}_x}]$

Fig. 2 Diagram of UT

then the sigma points are calculated which are then propagated through non-linear function. Then, the weighted sample mean and weighted sample covariance are calculated for further process [19].

To evaluate the mean and variance of $\boldsymbol{y}$**,** we initiate a matrix X_i of $2L + 1$ sigma vector X_i, relating to the following Eqs. 2–4 as shown in Fig. 2.

$$X_0 = \overline{x} \tag{2}$$

$$X_i = \overline{x} + \left(\sqrt{P_{\boldsymbol{X}}(L + \lambda}\right)_i, i = 1, \ldots, L \tag{3}$$

$$X_i = \overline{x} + \left(\sqrt{P_{\boldsymbol{X}}(L + \lambda}\right)_i, i = L + 1, \ldots, 2L \tag{4}$$

where $\lambda = \alpha^2(L + k) - L$. α is a coefficient that governs the sigma point spread around $\overline{x}$ and is generally set to a positive minor value (e.g., $1 \leq \alpha \leq 1e - 4$). k is a constant that is generally equal to 0 or 3-L and β is used for integration [20]. The initial information of the distribution of $\boldsymbol{x}$ (for Gaussian distribution $\beta = 2$ is ideal), $\left(\sqrt{P_X(L + \lambda}\right)_i$ is the ith column of the square root of the matrix. These sigma vectors undergo transition throughout as in Eq. 5,

$$y_i = f(X_i)i = 0, 1, 2, \ldots 2L \tag{5}$$

And using Eqs. 5–10, the weighted sample mean and covariance of the posterior sigma points are used to approximate the mean and covariance of $\boldsymbol{y}$ [21],

$$\overline{\boldsymbol{y}} \approx \sum_{i=0}^{2L} W_i^{(m)} y_i \tag{6}$$

$$P_y = \sum_{i=0}^{2L} W_i^{(c)} \{y_i - \overline{\boldsymbol{y}}\}\{y_i - \overline{\boldsymbol{y}}\}^T \tag{7}$$

With weights W_i are

$$W_0^{(m)} = \lambda/(\mathrm{L} + \lambda) \tag{8}$$

$$W_0^{(c)} = \lambda/(\mathrm{L} + \lambda) + \left(1 - \alpha^2 + \beta\right) \tag{9}$$

$$W_i^{(m)} = W_i^{(c)} = 1/\{2(\mathrm{L} + \lambda)\} \tag{10}$$

A diagram representing the steps in unscented transform is depicted in Fig. 1. Consider that, it varies considerably from the Monte-Carlo sampling methods that need more sample and orders of magnitude to propagate through a precise distribution of state [22, 23]. The illusionary simple way through with the UT leads to an approximation that are nearly equal to the third order of Gaussian inputs for all nonlinearities [14, 24]. For non-gaussian inputs, approximation is reduced precisely to 1st or 2nd order and the selection of α and β with the exactness of third order and other higher order moments are found.

3.2.2 Unscented Kalman Filter Equations

The UKF is a clear augmentation of the UT to the recurring assessment, when the state RV is reclassified due to the addition of the original state and noise variables: $x_k^a = \left[x_k^T V_k^T n_k^T\right]$. The UT sigma point choosing scheme (in Eq. 4) is put in to the new state random variable to determine the respective sigma matrix, X_k^a [2]. Then, the equations are initialized as shown in Eqs. 11–14. So, however, no conspicuous computation of Jacobians is important to execute this calculation. Moreover, the general number of calculations is a similar request as the EKF.

Initialize with

$$\hat{X}_0 = \mathbb{E}[X_0] \tag{11}$$

$$P_0 = \mathbb{E}\left[\left(X_0 - \hat{X}_0\right)\left(X_0 - \hat{X}_0\right)^T\right] \tag{12}$$

$$\hat{X}_0^a = \mathbb{E}\left[X^a\right] = \left[\,\hat{X}_0^T\ 0\ 0\,\right] \tag{13}$$

$$P_0^a = \mathbb{E}\left[\left(X_0^a - \hat{X}_0^a\right)\left(X_0^a - \hat{X}_0^a\right)^T\right] = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & R^v & 0 \\ 0 & 0 & R^n \end{bmatrix} \tag{14}$$

Calculation of sigma points:

$$X_{k-1}^a = \left[\hat{X}_{k-1}^a \hat{X}_{k-1}^a + \gamma\sqrt{P_{k-1}^a}\hat{X}_{k-1}^a - \gamma\sqrt{P_{k-1}^a}\right] \tag{15}$$

The time update equations are given from Eq. 16–20:

$$X_{k|k-1}^x = F\left[X_{k-1}^x, u_{k-1}, X_{k-1}^v\right] \tag{16}$$

$$\hat{X}_k^- = \sum_{i=0}^{2L} W_i^{(m)} X_{i,k|k-1}^x \tag{17}$$

$$P_k^- = \sum_{i=0}^{2L} W_i^{(c)} \left[X_{i,k|k-1}^- - \hat{X}_k^-\right]\left[X_{i,k|k-1}^- - \hat{X}_k^-\right]^T \tag{18}$$

$$y_{k|k-1} = H[X_{k|k-1}^x, X_{k-1}^n] \tag{19}$$

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_i^{(m)} y_{i,k|k-1} \tag{20}$$

The measurement update equation is from Eqs. 21–25:

$$P_{\hat{y}_k \bar{y}_k} = \sum_{i=0}^{2L} W_i^{(c)} [y_{i,k|k-1} - \hat{y}_k^-][y_{i,k|k-1} - \hat{y}_k^-]^T \tag{21}$$

$$P_{x_k y_k} = \sum_{i=0}^{2L} W_i^{(c)} [x_{i,k|k-1} - \hat{x}_k^-][y_{i,k|k-1} - \hat{y}_k^-]^T \tag{22}$$

$$K_k = P_{x_k y_k} P_{\hat{y}_k \bar{y}_k}^{-1} \tag{23}$$

$$\hat{x}_k = \hat{x}_k^- + K_k\left(y_k - \hat{y}_k^-\right) \tag{24}$$

$$P_k = P_k^- - K_k P_{\hat{y}_k \bar{y}_k} K_k^T \tag{25}$$

where $x^a = \left[x^T v^T n^T\right]$, $X^a = \left[(X^x)^T (X^v)^T (X^n)^T\right]^T$, $\gamma = \sqrt{(\mathrm{L} + \lambda)}$, where $R^v is$ the process noise variance, R^v is the measurement noise covariance, and W_i are the weights that are calculated in Eq. 4. The measurement is then updated in each time frame of the speech taken [2]. Then, all the time frames are then augmented to get back the denoised and dereverberated clean processed speech.

4 Experimental Results

In this section, the simulation results we obtained from the approach detailed in the above section, i.e., the UKF algorithm are discussed. There were few .wav files on which we performed the algorithm under various windowed processing like the rectangular window, hamming window and the Gaussian window. The results we obtained are plotted as wave forms. There are two wave files on which this algorithm was performed. Let the names be Speech A.wav and Speech B.wav. The SNR was precalculated for later use in the comparisons. The waves were then reverberated, and the observation noise was added to both. The observation noise added to all the wave forms is additive white Gaussian noise (AWGN).

After processing the two waveforms through the algorithm and getting the results, we calculated the parameters such as the figure of merit and the correlation between SNR of the processed output and the precalculated SNR of the clean speech for three different number of iterations in the algorithm. A table is given below with the particular details of the figure of merit and correlation for the above two wave forms.

Table1 shows analysis of the performance metrics FOM and correlation between the input and the output of the algorithm proposed.

Table 2 shows the comparison between the SNR values for the different windows—rectangular, hamming and the Gaussian windows used for chopping and the number of iterations performed on both speech A and speech B.

5 Conclusion

In this project, speech enhancement technique using Kalman filtering has been implemented. The objective was to design an effective method to process a noise invaded and reverberated speech in adverse environments. We were able to perform the denoising and dereverberation on the corrupted speech. The proposed algorithm can be used in the cases of nonlinear systems, where in most of the algorithms, this is not possible. Also, this algorithm is time-efficient. So, it can be used for mediocre length speeches. Here, the proposed algorithm, unscented Kalman filtering, uses three windows—rectangular, hamming and Gaussian for the chopping of the signal before processing, and from Table 1, the results significantly differ from each window for every iteration. The performance is slightly increasing with the increasing number of iterations in any window up to a certain number of iterations. Then, there is fall in both the performance metrics—figure of merit and the correlation taken in this report. This is due to the repeated denoising and dereverberation, which causes a damage to the intelligibility of the desired output. Then, Table 2 compares the SNRs of the outputs of different windows under different number of iterations.

Table 1 Performance metrics

Number of iterations	Parameter	Feared_or_respected.wav			Movie-05.wav		
	window	Rectangular	Hamming	Gaussian	Rectangular	Hamming	Gaussian
15	Figure of merit	0.824	0.837	0.831	0.506	0.398	0.5564
	Correlation	0.457	0.443	0.448	0.461	0.432	0.4560
30	Figure of merit	0.822	0.824	0.827	0.498	0.401	0.7104
	Correlation	0.480	0.463	0.472	0.478	0.461	0.4703
50	Figure of merit	0.833	0.830	0.844	0.432	0.421	0.5503
	Correlation	0.475	0.459	0.472	0.468	0.460	0.4561

Table2 SNR comparison

		Speech A.wav			Speech B.wav		
Window	Iterations	15	30	50	15	30	50
Rectangular		18.15	18.23	17.8	7.4	7.75	7.6
Hamming		18.44	18.56	18.2	5.82	5.83	5.78
Gaussian		18.31	18.42	18.1	8.13	8.23	8.09

References

1. M. Mosallaei, Performance evaluation of instrumentation sensor network design using a data reconciliation technique based on the unscented Kalman filter, in *2007 IEEE Conference on Emerging Technologies & Factory Automation (EFTA 2007),* 09/2007.
2. N. Dionelis, M. Brookes, Modulation-Domain Kalman filtering for monaural blind speech denoising and dereverberation. IEEE/ACM Trans Audio, Speech, Lang Process **27**(4), 799–814 (April 2019). https://doi.org/10.1109/TASLP.2019.2894909.
3. A.UmaMageswari, J. Joseph Ignatious ,R. Vinodha, A comparitive study of Kalman Filter, Extended Kalman Filter And Unscented Kalman Filter For Harmonic Analysis of the non-stationary signals International Journal of Scientific & Engineering Research (2012)
4. M. Fujimoto, Y. Ariki, Noisy speech recognition using noise reduction method based on Kalman filter, in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference,* vol. 3. Pp. 1727–1730. https://doi.org/10.1109/ICASSP.2000.862085
5. M.G. Muthukrishnan, P. Sudheesh, M. Jayakumar, Channel estimation for a high mobility MIMO system using Particle filter, in *İnternational Conference on Recent Trends in Information Technology* (2016), pp.197–207
6. F. Asano, S. Hayamizu, T. Yamada, S. Nakamura, Speech enhancement based on the subspace method. IEEE trans. Speech Audio Proc. **8**(5), 97–507 (2000)
7. M.A.A. El-Fattah, M.I. Dessouky et al., Speech enhancement with an adaptive Wiener filter. Int. J. Speech Technol. **17**, 53–64 (2014)
8. S.F.Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoustics, Speech Sig. Proc. **ASSP-27**(2), 113–120,1979
9. B. Cornelis, M. Moonen, J. Wouters, Performance analysis of multichannel wiener filter-based noise reduction in hearing aids under second order statistics estimation errors. IEEE Trans. Audio Speech Lang. Process. **19**(5), 1368–1381 (2011)
10. R.E. Kalman, A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. **82**(D), 35–45 (1960)
11. B.L. Sim, Y.C. Tong, J. Chang, C.T. Tan, A parametric formulation of the generalized spectral subtraction method. IEEE Trans. Speech Audio Proc. **6**(4), 328–337 (1998)
12. S.J. Julier, J.K. Uhlmann, H. A New Extension of Kalman filter to Nonlinear systems, *ın proc ofAreoSense*: *The 11th int. symp. on Areospace/Defence sensing. Simulaton and controls* (1997)
13. K.K. Paliwal, A. Basu, A speech enhancement method based on kalman filtering, ın *Proc. ICASSP,* vol 12 (1987)
14. S. Braun, E.A.P. Habets, Linear prediction based online dereverberation and noise reduction using alternating Kalman filters, in *IEEE/ACMTrans. Audio, Speech, Lang. Process.*, vol. 26, no. 6 (2018), pp. 1119–1129
15. S. Haoxiang Wang, Smys, secure and optimized cloud-based cyber-physical systems with memory-aware scheduling scheme. J Trends Comput. Sci. Smart Technol. (TCSST) **2**(03), 141–147 (2020)
16. V.R. Balaji, Maheswaran S, M. Rajesh Babu, M. Kowsigan, Prabhu E., Venkatachalam K, *Combining statistical models using modified spectral subtraction method for embedded system Microprocessors and Microsystems*, vol 73(2020),102957, ISSN 0141-9331

17. J. Wei, L. Du, Z. Yan, H, Zeng, *Improved Kalman Filter-Based Speech Enhancement* (2003)
18. E.A. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)* (2000)
19. A. Suraj, S, A, G., S, S. Chakravarthy, R. Ramnathan, Routing in wireless sensor network based on swarm ıntelligence, in *3rd International Conference on Trends in Electronics and Informatics* (2019)*,* pp. 215–217
20. T. S. Kavya, T. Peng, Y.M. Jang, E. Tsogtbaatar, S.B. Cho, Face Tracking Using Unscented Kalman Filter, in *2020 International Conference on Electronics, Information, and Communication (ICEIC)* (2020)
21. R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing Communications and Control Symposium (Cat No 00EX373) ASSPCC-00* (2000)
22. B.S. Atal, Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. **47**(1), 65 (1970)
23. J. Ramnarayan, J.P. Anita, P. Sudheesh, Estimation and Tracking of a Ballistic Target Using Sequential Importance Sampling Method. Commun. Comput. Inf. Sci. **746**, 387–398 (2017)
24. R.G. Reddy, R. Ramnathan, An Empirical study on MAC layer in IEEE 802.11p/WAVE based Vehicular ad hoc Networks. Procedia Comput Sci **143**, 720–727 (2018)

An Agriculture Supply Chain Model for Improving Farmer Income Using Blockchain Smart Contract

Banupriya Sadayapillai and Kottilingam Kottursamy

Abstract Farmers are the backbone of our society, and they support the whole world's population. Though, farmers are not getting adequate profit due to the intermediaries and inefficient supply chain system. Farmers will begin producing more once they stand to benefit from higher farm produce rates, which is not the case at the moment. Technology innovation in the agricultural sector is essential for the development of efficient and productive food systems. In this paper, we propose a novel system for automating and controlling the supply chain process through blockchain and smart contracts, while maintaining total trust between the parties. In particular, we design smart contracts with two parameters, Minimum Purchase Rate (MPR) and Percentage Profit Share (PPS), to monitor the cost of product and to increase the farmer's income. The payment history is recorded in a distributed ledger and can be used to trace back to the source of the products. Furthermore, the proposed framework is constructed and tested in an Ethereum blockchain network. Finally, the analysis results show that our solution is feasible and improves the farmer's income.

Keywords Blockchain · Distributed ledger · Ethereum · Smart contract · Supply chain management · Food traceability

1 Introduction

Supply chain management (SCM) has evolved into automated and increasingly dynamic networks, and they are now an important way of creating opportunities in the real world. However, it is difficult to monitor data provenance and preserve traceability through a network. Food and commodity consistency are becoming more important to consumers. Traditional supply chains depend on third-party trade and are centralized. Traditional centralized systems lack transparency, openness, and

B. Sadayapillai (✉) · K. Kottursamy
School of Computing, SRM Institute of Science and Technology, Chennai, Tamilnadu, India
e-mail: bs9093@srmist.edu.in

K. Kottursamy
e-mail: kottilik@srmist.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_46

auditability. The interaction of government policy with the management of the agricultural supply chain needs multi-level research. Government approaches to agriculture are influenced by economic growth, economic concerns, standards of national and sub-national bodies, local environmental factors, and legacies. Research on farmers' income maximization, crop production, policies, and physical infrastructure needs to look at all these aspects and efforts to combine this puzzle into a complete picture of how these factors interact.

Most of the agriculture policies have primarily focused on three policy instruments to raise farmers' incomes: reducing costs of production through input subsidies, raising agricultural productivity through improved seeds and agricultural activities, and enhancing operational prices, while stabilizing revenues through the minimum support price and purchasing. More specifically, the policy has continued to concentrate on having farmers a bigger proportion of the marketed surplus, which has intensified fears about the condition of farming markets, which is the focus of our research.

The final price of farmer products is influenced by a variety of factors as sale timing, location, volume, and operational processes such as grading, quality assessment, price determination, weighing system, and scheduling and payment method. The location of the first sale often differs greatly. As a result, middlemen become extremely strong, and farmers often find themselves at a disadvantage despite being the producers. Farmers want an environment that allows them to fix the price of their products in compliance with customer demands by engaging and selling directly in the market. At the moment, the supply chain should be improved to get suppliers as close to customers as possible to benefit the farmers and customers. Blockchain technology [1] can address these natures of the problem by the use of a publicly licensed ledger system that offers an eco-system. Blockchain allows information to flow openly, but in a licensed way, removes the supply chain risk associated and thus reduces the overall supply chain expense, whereas agility and adaptability are enhanced.

The major contribution of this paper is as follows:

- We propose a novel agriculture supply chain framework to ensure transparency in product pricing, automate the supply-chain management, control and govern the transactions using a smart chain.
- We implement the framework in solidity language with two global parameters called Minimum Purchase Rate (MPR) and Percentage Profit Share (PPS).
- Finally, we demonstrate the result that the proposed framework is feasible and improves the farmer's profit.

2 Literature Review

Researchers across the world are developing various agricultural systems based on blockchain technology to increase overall revenue in the agriculture industry. Many studies, proposals, and advocacy attempts have been made in the literature to investigate, suggest, or promote the use of blockchain technology during pre-harvesting and post-harvesting. The blockchain has been used in agriculture, which covers crop insurance, smart farming, the food supply chain, and food commodity. Various intelligent agriculture solutions based on blockchain technology, IoT and AI were developed and executed. The basic information and data on natural resources which sustain all forms of agriculture and the agro-food systems. The blockchain-based concepts related to ICT-based technologies were reviewed in Lin et al. [2]. They also presented a model ICT e-agriculture system based on blockchain technology for usage at the local and regional levels. Other studies [3] and [4] explore the applications of blockchain in the agriculture sector.

In paper [5], the author suggested a lightweight, smart greenhouse farm employing blockchain architecture. Greenhouse IoT devices that operate as a central blockchain for energy efficiency. This smart greenhouse farming combines IOT technology with blockchain for secure communication infrastructure. In paper [6], the author proposes a similar general-purpose smart agricultural paradigm based on blockchain and IoT. The design and architecture backbone are a framework for rebuilding relationships among blockchain participants. Marketing vendors can use a mobile device to communicate blockchain data from planting to sales. Torky and Hassanein [7] and Singh and Singh [8] implemented a similar approach, precision agriculture, to keeps track of agricultural characteristics and then regulates them. In paper [9], the author suggested electronic farming based on blockchain. This approach is included in the blockchain's comprehensive circular agricultural model of the ecological farm. To expand the data set available for communication, the blockchain network automatically captures and posts data through various types of smart devices. This could help with issues like asymmetrical information, untrustworthy third-party entities, and insufficient organic food traceability. Solution provided in this paper helps farmers to access actual information on seed quality, climatic and environmental data, payments, soil moisture, demand and sale price, etc., from a unified system.

AgriBlockIoT [10] is a reliable and consistent blockchain-based food supply chain management system. Suggested blockchain-based traceability solution seamlessly connected to IoT devices that supply the generation and consumption of digital data. Both Ethereum and Hyperledger Sawtooth blockchain technologies were used to achieve traceability. Another recent initiative [11] has suggested blockchain technology for food safety and traceability. The analysis dealt with the large data created by the electronically coded (EPC) tags in the offset and online data management system. A blockchain-based food safety solution for greenhouse farms also was suggested in Patil et al. [5]. The lightweight solution guarantees the safety and confidentiality of intelligent greenhouses. Patil et al. [12] presented a blockchain-based rice supply chain management system to boost traceability by collecting data at several stages of the construction.

A system of traceability of the food supply chain using radio frequency identification (RFID), communication technology for non-contact identification is presented in Kumar and Iyengar [13]. It can track products through the whole supply chain with trustworthy information. The usage of blockchain ensures dependable and legitimate manufacturing, process, stock, and distribution records in the system. In paper [14], the new deployment of blockchain, IoT, and fugitive logic in a whole traceability shelf-life management system for managing fresh food is presented to include a BIFTS (blockchain IoT-based food traceability system). Lightweight and sprayed qualities are used in the blockchain to satisfy the food traceability requirements, while an integrated consensus process is designed which takes into account shipment transit, stakeholder evaluation, and shipment volume. The blockchain data flow is then coordinated with IoT technology deployment at the traceable unit level.

KRanTi [15] is another agricultural food supply chain over the 5G network using BC embedding. By boosting network infrastructure efficiency, the 5G network improves data connectivity. We have an efficient loan system in KRanTi that enables farmers to buy the required agricultural raw product of higher grade without an immediate payment constraint. In paper [16], a decentralized platform is proposed to buy and sell agricultural production through a link between farmers and persons interested in investing in their fields, through IoT and machine learning to forecast illnesses in agricultural products, and through constant quality monitoring and crop health. Bakare et al. [17] addressed the distribution of direct cash transfers to farmers through blockchain smart contracts technology directly to provide transparency, deduplication, reduced delays, and reduction of fraud in the existing government subsidy system. Hassija et al. [18] suggested a cost-effective secure blockchain architecture for constructing a farmers' community and multi-sourcing data to assist farmers' communities. The majority of the efforts described in Yadav et al. [19], Juma et al. [20], Kamble et al. [21], Bodkhe et al. [22], Saurabh et al. [23], and Kramer et al. [24] are also concentrated on improving or increasing the agriculture supply chain.

3 Blockchain and Smart Contract

Blockchain technology is a decentralized electronic system that provides accountability, tracking, and protection for improving rapidly in alleviating some global supply chain management issues. A blockchain is a set of blocks in which each block contains the data/transactions, a hash of the data, and the previous block hash as shown in Fig. 1. Hash is the output of a one-way mathematical function with a fixed-size value. Any single bit modification in the stored data can be identified by comparing the previous hash value stored in the next block. The copy of the blockchain is maintained in each node in the peer-to-peer (P2P) network, which makes the system more secure. Each node validates the data against the rules written in the smart contract. A smart contract specifies various states of a business entity and controls the processes that transfer the object between them. Buyers and sellers communicate and trade themselves using smart contracts on a blockchain platform. For example, the sender must own the funds with his identity to transfer them. This condition is written in the smart contract and deployed into P2P nodes. In each fund transfer call, the condition is automatically verified by P2P nodes.

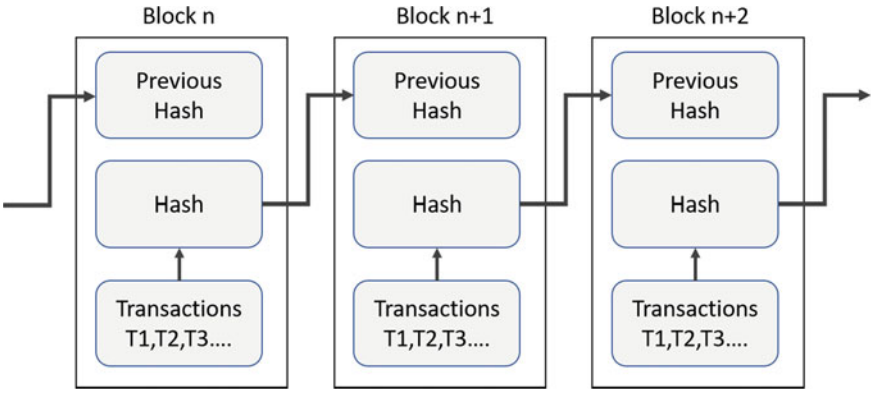


Fig. 1 Blockchain structure

4 Proposed Model

The proposed architecture overview is depicted in Fig. 2. The proposed framework consists of 4 entities: distributed network, government entity, end-user, and certificate authority.

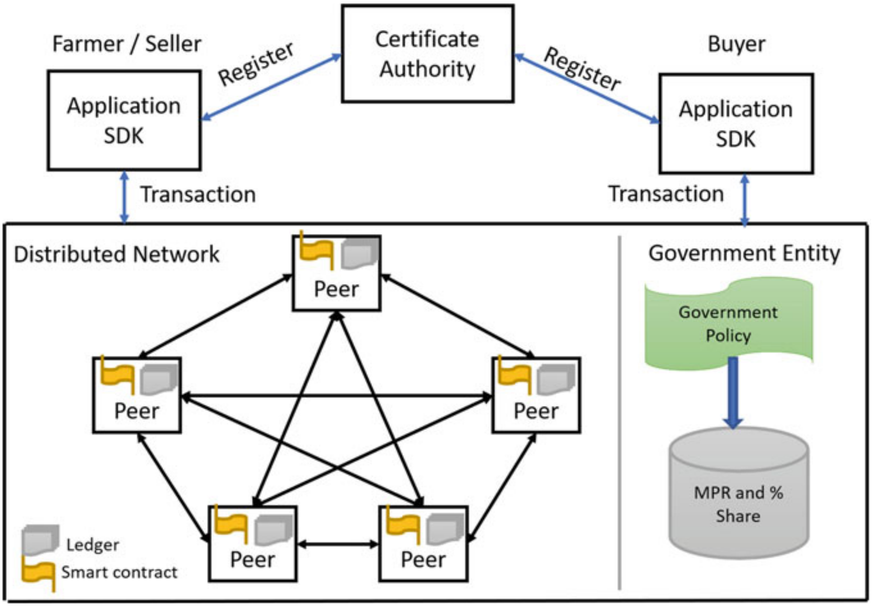


Fig. 2 Blockchain architecture overview

(1) The distributed network consists of P2P nodes and maintains the ledger and smart contract in each node. The ledger is an immutable blockchain that contains the history of transactions.

(2) The government entity provides the global parameter value for each product and regularly updates it. Two parameters, minimum purchase rate (MPR) and percentage of profit share (PPS), are maintained in a government database for a variety of vegetables, fruits, grains, and so on. The data can be updated based on real-time situations and policies.

(3) The end-user can be a farmer, trader, or consumer who can trade the products in the distributed network. The farmer produces a new product and initiates the transaction. The trader can be transporters, distributors, wholesalers. The transactions happen within the trader until it reaches the consumer. The transaction cycle ends when the product reaches the consumer.

(4) The certificate authority generates cryptographic key pairs (public and private addresses) to end-users and assigns roles to each end-user based on their identity. The public address is used to identify the end-user, and the private address is used to transfer the product from one user to another user.

We assume that all the transactions related to agriculture products are happening only through the proposed model under the supervision of a government entity. The government entities regularly update the database based on real factors and policy.

4.1 Ledger

The transactions from farmer to consumer are recorded permanently in the ledger with a timestamp. When the farmer adds the product, the transaction is recorded in the blockchain as shown in Fig. 3. At each stage, the common smart contract conditions are verified automatically. Consumers can trace the source of the product and check the cost of the product. Smart contracts resolve the issue of storing essential data required at various levels of the supply chain and keeping it accessible by all supply chain participants. Smart contracts help to control the network and simplify the processes. It can never be modified once it is implemented inside the blockchain.

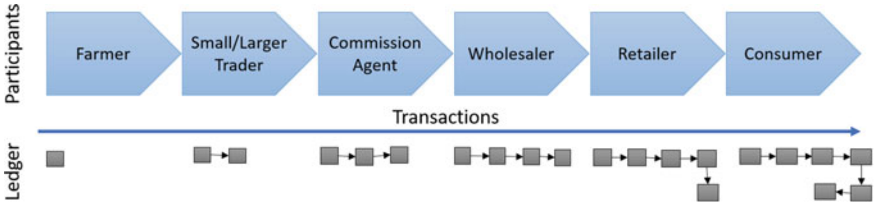


Fig. 3 Transaction in blockchain

Table 1 Smart contract functions and conditions

S. No	Functions	Description	Conditions
1	AddProduct()	To add new product	Caller–Farmer
2	ReadyToSell()	Seller sets the expected price	Caller–Owner, Price > = MSP
3	AgreeToBuy()	Buyer agrees to buy and set a bid	Caller–Any caller, bid > = MPR, Balance of caller > = bid
4	AgreeToSell()	Seller agrees to sell for the bid	Caller–Owner, bid > = MPR
5	Transfer()	Transfer the product ownership to the buyer and Share the farmer's profit	Caller–Owner, bid > = MPR, Balance of buyer > = bid
6	GetProduct()	Get the details of the product	Caller–Any caller
7	GetBalance()	Get the balance of the address	Caller–Any caller

4.2 Smart Contract

The smart contract consists of various state variables and functions to automate the trading. The proposed smart contract consists of 7 functions as shown in Table 1. The smart contract is available in each node in the distributed network. The smart contract executes the functions only when the written conditions are met. The smart contract automatically checks the minimum purchase rate (MPR) whenever the product is transferred from the farmer to any buyer. The transaction completes only when the agreed price is greater than or equal to the MPR given in the database. The MPR parameter value helps the farmer to get the minimum cost for the product as per the government norms.

4.3 Transactions

The transactions are initiated by the end-user to trade the product in the distributed network. Each end-user generates public and private keys from the certificate authority before trading. End user-initiated transactions can be categorized into types: farmer to trader and trader to trader. The transaction cycle of each product contains one farmer to trader transaction and many trader to trader transactions. The MPR is verified in each farmer to trader transaction as given in the smart contract conditions. In trader to trader transaction, the profit share of the farmer is calculated based on a percentage of profit (PPS) parameter value. Whenever the middleman profits more, the farmer also benefited economically. The product can be transferred to any number of the user until it reaches the consumer. At each intermediate transaction, the profit amount is transferred to the farmer based on the government policy. The value of

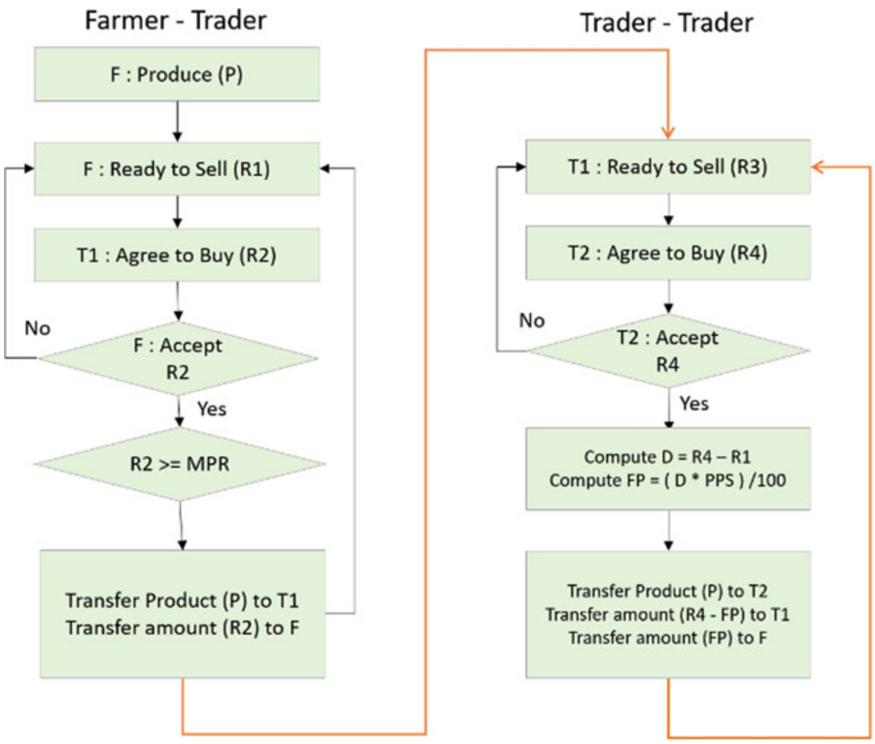


Fig. 4 Transaction flow

PPS is updated based on location, transportation charges, demand and supply of the product, and so on.

The processing flow of farmer to trader and trader to trader is depicted in Fig. 4 with the following labels: F-farmer, $T1$, $T2$—traders, P-product, $R1$, $R2$, $R3$, $R4$—Selling/purchase rate, MPR—Minimum Purchase Rate, PPS Percentage of Profit Share. Initially, the farmer creates the product by submitting all the necessary inputs. The product is tied with the farmer's public address to make the farmer as the owner of the product. The owner calls the ReadyToSell() and sets the expected price for the added product. Any buyer can call the AgreeToBuy() with a specific product and set the bid. The owner of the product calls the AgreeToSell(), which means the owner accepts to sell the product for the bid. Upon agreeing to the deal, the transfer condition written in the smart contract is checked and then the product is transferred to the buyer. Now the product is liked with the buyer's public address, and the amount is transferred to the farmer's account. The actual amount transferred to the farmer account is set to the state variable farmer price. In each function call, the smart contract conditions are verified. For example, when the buyer calls AgreeToBuy(), the smart contract condition checks two conditions as given in Table 1. The bid value must be greater than the MPR, and the buyer's balance must be greater than the bid value.

Similarly, trader to trader transaction is processed by calling ReadyToSell(), AgreeToBuy(), AgreeToSell and Transfer(). During Transfer() call, PPS value is utilized to compute the farmer profit. At each trader to trader transaction, the difference amount ($D =$ Trader selling price–FarmerPrice) is calculated. The PPS percentage of this difference amount [(PPS * D)/100] is transferred to the farmer.

5 Implementation and Analysis

We have created the architecture of the framework and implemented the smart contract using the Ethereum blockchain. The Remix is a web-based IDE that provides a smarter contract with the aid of solidity programming language, Ethereum voucher as wallets loaded with dummy ether cryptocurrency and infrastructure to execute and deploy the contract within the Ethereum blockchain. The MPR of the onion is set as 40 rupees per kg and the PPS is set to 5% in the database. The state variables are maintained using 2 structures called ProdData and farmer price. The ProdData structure stores product identity number, current owner, product name, quantity, price, and timestamp. The farmer price structure stores product identity number, farmer addresses, and farmer prices and used to calculate the farmer profit. The product can be added by the farmer and transferred by the owner.

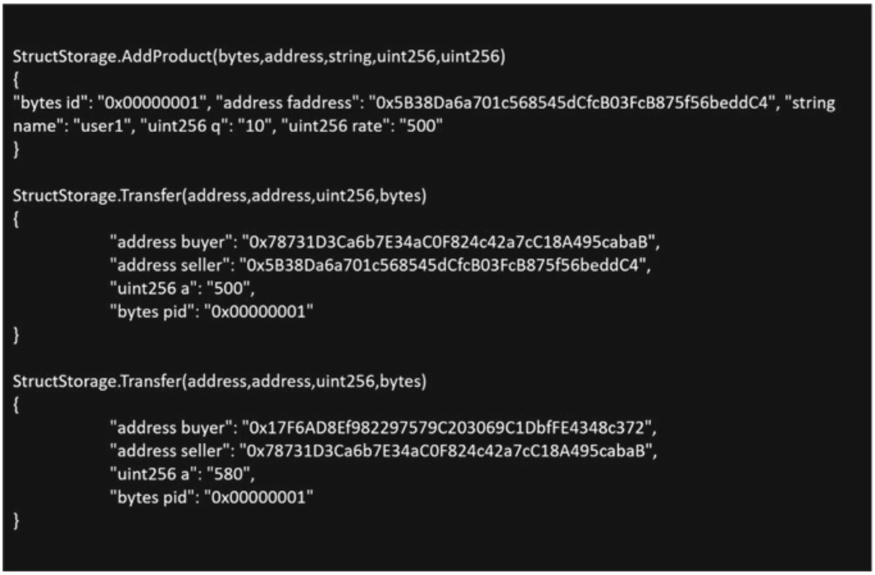
Figure 5 hows the transaction of AddProduct initiated by the farmer, transfer initiated by the farmer, and transfer initiated by the trader. Initially, all the Etherum addresses ($F1$, $T1$, $T2$) hold 2000 as balance.

(1) ***AddProduct***: Farmer ($F1$) added the product by setting the price as 500 for 10 kgs. At this point, there is no change in the balance of the F1, T1, T2 addresses. All these addresses hold 2000 as a balance.

(2) ***Transfer***: Farmer ($F1$) sells the product to the trader ($T1$) and gets 500 for the sale of the product. The transaction is validated against MPR. Now, the balance of $F1$ is 2500 (2000+500), T1 is 1500 (2000-500) and T2 is 2000.

(3) ***Transfer***: Trader ($T1$) sells the product to Trader ($T2$) for 580 and the farmer gets 4 as a profit. When the Trader ($T1$) sells it for 580, the gain value 80. So the PPS 5% (4 Rs) of gain is transferred to the farmer and 95% of the gain (76 Rs) is transferred to the trader. Now, the balance of $F1$ is 2504(2500+4), $T1$ is 2076 (1500+576) and $T2$ is 1420 (2000-580).

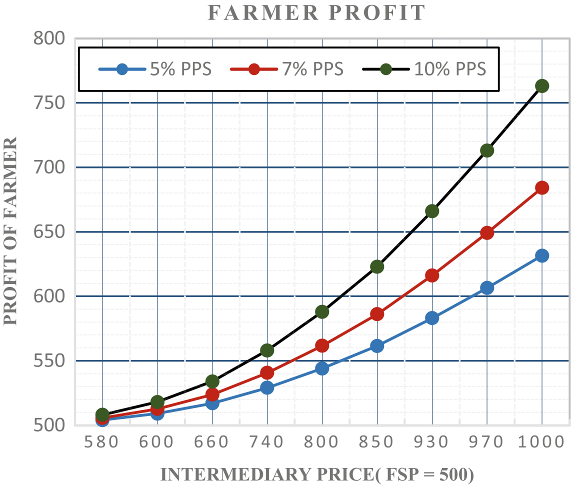
Figure 6 shows the farmer's profit for the various intermediary prices. The farmer selling price is denoted as FSP and it is set to 500. When the PPS is set to 5%, the farmer gets 131.5 as profit from intermediaries. Similarly, when PPS is set at 7 and 10% the farmer gets 184.1 and 263, respectively. The smart contract automatically transfers the farmer's profit from the intermediary's profit.



```
StructStorage.AddProduct(bytes,address,string,uint256,uint256)
{
"bytes id": "0x00000001", "address faddress": "0x5B38Da6a701c568545dCfcB03FcB875f56beddC4", "string
name": "user1", "uint256 q": "10", "uint256 rate": "500"
}

StructStorage.Transfer(address,address,uint256,bytes)
{
                "address buyer": "0x78731D3Ca6b7E34aC0F824c42a7cC18A495cabaB",
                "address seller": "0x5B38Da6a701c568545dCfcB03FcB875f56beddC4",
                "uint256 a": "500",
                "bytes pid": "0x00000001"
}

StructStorage.Transfer(address,address,uint256,bytes)
{
                "address buyer": "0x17F6AD8Ef982297579C203069C1DbfFE4348c372",
                "address seller": "0x78731D3Ca6b7E34aC0F824c42a7cC18A495cabaB",
                "uint256 a": "580",
                "bytes pid": "0x00000001"
}
```

Fig. 5 Smart contract function calls

Fig. 6 Farmer profit share from intermediaries' profit



6 Conclusion

Traditional supply chain management in the agriculture sector is inefficient and failed to provide transparency to the traders and consumers. Numerous technological innovations facilitate overcoming these issues. Though the solutions are not guaranteeing the farmer's profit. We proposed a novel agriculture supply chain framework to ensure transparency in product pricing, automate the supply-chain management, control and govern the transactions using a smart chain. We implemented the framework in solidity language with two global parameters called minimum purchase rate and percentage profit share. The proposed model ensures the minimum purchase rate whenever the product is purchased from the farmer and maximizes the farmer's profit using the percentage of profit share. The proposed smart contract is deployed and tested in the Ethereum blockchain network. Finally, we demonstrated the result that shows the proposed framework is feasible and improves the farmer's profit.

References

1. S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System* (Manubot, 2019)
2. Y.P. Lin, J.R. Petway, J. Anthony, H. Mukhtar, S.W. Liao, C.F. Chou, Y.F. Ho, Blockchain: the evolutionary next step for ICT e-agriculture. Environments **4**(3), 50 (2017)
3. L. Ge, C. Brewster, J. Spek, A. Smeenk, J. Top, F. van Diepen, B. Klaase, C. Graumans, M.D.R. de Wildt, *Blockchain for agriculture and food: findings from the pilot study* , vol 112 (Wageningen Economic Research, 2017)
4. A. Kamilaris, A. Fonts, F.X. Prenafeta-Boldύ, The rise of blockchain technology in agriculture and food supply chains. Trends Food Sci. Technol. **91**, 640–652 (2019)
5. A.S. Patil, B.A. Tama, Y. Park, K.H. Rhee, A framework for blockchain based secure smart green house farming, in *Advances in Computer Science and Ubiquitous Computing*. (Springer, Singapore, 2017), pp. 1162–1167
6. M.S. Devi, R. Suguna, A.S. Joshi, R.A. Bagate, Design of IoT blockchain based smart agriculture for enlightening safety and security, in *International Conference on Emerging Technologies in Computer Engineering* (Springer, Singapore, 2019), pp. 7–19
7. M. Torky, A.E. Hassanein, Integrating blockchain and the internet of things in precision agriculture: Analysis, opportunities, and challenges. Comput. Electr. Agric. 105476 (2020)
8. P. Singh, N. Singh, Blockchain With IoT and AI: a review of agriculture and healthcare. Int. J. Appl. Evol. Comput. (IJAEC) **11**(4), 13–27 (2020)
9. A. Vangala, A.K. Das, N. Kumar, M. Alazab, Smart secure sensing for IoT-based agriculture: blockchain perspective. IEEE Sensors J. (2020)
10. M.P. Caro, M.S. Ali, M. Vecchio, R. Giaffreda, Blockchain-based traceability in Agri-Food supply chain management: A practical implementation, In *2018 IoT Vertical and Topical Summit on Agriculture-Tuscany (IOT Tuscany)* (IEEE, 2018), pp. 1–4
11. Q. Lin, H. Wang, X. Pei, J. Wang, Food safety traceability system based on blockchain and EPCIS. IEEE Access **7**, 20698–20707 (2019)
12. M.V. Kumar, N.C.S. Iyengar, A framework for Blockchain technology in rice supply chain management. Adv. Sci. Technol. Lett **146**, 125–130 (2017)
13. F. Tian, An agri-food supply chain traceability system for China based on RFID & blockchain technology, in *2016 13th International Conference on Service Systems and Service Management (ICSSSM)* (IEEE, 2016), pp. 1–6
14. J. Lin, Z. Shen, A. Zhang, Y. Chai, Blockchain and IoT based food traceability for smart agriculture, in *Proceedings of the 3rd International Conference on Crowd Science and Engineering* (2018), pp. 1–6
15. N. Patel, A. Shukla, S. Tanwar, D. Singh, KRanTi: blockchain-based farmer's credit scheme for agriculture-food supply chain. Trans. Emerging Telecommun. Technol **e4286** (2021)
16. M. Senthilmurugan, R. Chinnaiyan, IoT and machine learning based peer to peer platform for crop growth and disease monitoring system using blockchain, in *2021 International Conference on Computer Communication and Informatics (ICCCI)* (IEEE, 2021), pp. 1–5

17. S. Bakare, S.C. Shinde, R. Hubballi, G. Hebbale, V. Joshi, A blockchain-based framework for Agriculture subsidy disbursement, in *IOP Conference Series: Materials Science and Engineering* , vol. 1110, No. 1 (IOP Publishing, 2021), p. 012008
18. V. Hassija, S. Batra, V. Chamola, T. Anand, P. Goyal, N. Goyal, M. Guizani, A blockchain and deep neural networks-based secure framework for enhanced crop protection. Ad Hoc Networks 102537 (2021)
19. V.S. Yadav, A.R. Singh, R.D. Raut, U.H. Govindarajan, Blockchain technology adoption barriers in the Indian agricultural supply chain: an integrated approach. Res., Conser. Recycling **161**, 104877 (2020)
20. H. Juma, K. Shaalan, I. Kamel, A survey on using blockchain in trade supply chain solutions. IEEE Access **7**, 184115–184132 (2019)
21. S.S. Kamble, A. Gunasekaran, R. Sharma, Modeling the blockchain enabled traceability in agriculture supply chain. Int. J. Inf. Manag. **52**, 101967 (2020)
22. U. Bodkhe, S. Tanwar, P. Bhattacharya, N. Kumar, Blockchain for precision irrigation: opportunities and challenges. Trans. Emerg. Telecommun. Technol. (2020), e4059
23. S. Saurabh, K. Dey, Blockchain technology adoption, architecture, and sustainable agri-food supply chains. J. Clean. Prod. **284**, 124731 (2021)
24. M.P. Kramer, L. Bitsch, J. Hanf, Blockchain and its impacts on agri-food supply chain network management. Sustainability **13**(4), 2168 (2021)

SDN-Enabled Secure IoT Architecture Development: A Review

Shikhar Bhardwaj and Sandeep Harit

Abstract The Internet of things (IoT) is a revolution in the technological future. We expect it to affect everything, from virtually all forms of automation to everyone, living or non-living, and in multiple ways. Every network entity contributes to the generation of huge amounts of data, and network objects exchange these data buckets over the Internet for seamless transfer of control. With the use of numerous digital technologies, each network object receives, interprets, and utilizes the data on account of non-standard intrinsic properties, thus creating an exploit opportunity for malicious users, hence compromising the security and privacy of the data. With such significant effects on society through technological advancements, trust in IoT-based systems is necessary. The key to this trust lies in security advancements and enhancements in IoT connectivity, implying that IoT must exhibit some level of reliable, secure, and private behaviors. This paper proposes the development considerations for secure IoT architectures. The discussion varies over various aspects of IoT architecture features, challenges that arise when adopting an IoT security architecture, security techniques proposed in recent years, and how software-defined networking (SDN) and SDN controllers help alleviate these aforementioned security concerns. This paper also highlights some of the existing industrial IoT frameworks and comparison of certain primitive SDN controllers.

Keywords Privacy · Authentication · SDN controllers · IoT · OpenFlow · SDN · Security · Cyberphysical systems

S. Bhardwaj (✉) · S. Harit
Punjab Engineering College (Deemed to be University), Chandigarh, India
e-mail: shikharbhardwaj.phd19cse@pec.edu.in

S. Harit
e-mail: sandeepharit@pec.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_47

1 Introduction

Nowadays, IoT and cloud are two very closely related future Internet technologies, which go hand-in-hand in non-trivial IoT deployments [1]. Former IT systems have evolved to constitute the future IoT systems [2]. There are three major forces, namely speed, data, and autonomy which act at different levels of IoT, that encompass three complementary concepts:

1. Communicating devices, any standard/non-standard computing device, connect via any wired or wireless medium to a network and can transmit data.
2. Machine-to-machine (M2M) communication refers to machines which, even without any connectivity, can work as autonomous objects.
3. Internet of objects (IoO) refers to passive objects which do not generate any data.

The above concepts must incorporate security. "The S in IoT stands for security." Think about that for a second, as you say, "Wait, there is no S in IoT." That is precisely the point of this statement. Various security features are missing from IoT. An end-to-end secure IoT solution development involves multiple levels that fuse essential security architecture features across four different management layers, namely: device, communications, cloud, and life cycle. There are security considerations across all layers, and these depend on the type of implementation, protocol, and requirement of security.

IoT networks are constantly under the threat of being compromised using various attacks and intrusions. Software-defined networking (SDN) is one of the concepts, that protect the network in a much more flexible and efficient way. SDN consists of controllers that act as a central decisive entity for connecting various nodes and communication among these nodes, which exploit numerous communication protocols. OpenFlow is the most widely used controller/protocol as it is very flexible. It contains the definition of control messages to establish a secure connection and other manipulations within the network.

If one wants to begin the development of a secure IoT network, the present literature lacks standardization for design, development, and deployment aspects. The article establishes this in Sect. 4 as we discuss the present state of secure IoT network development. Each technique has a different approach for scaling up toward a developed architecture resulting in a heterogeneity of products, thus creating a pool of marginally elusive frameworks yet far-fetched non-interoperable frameworks. The article establishes this in Sect. 8.

In this research article, certain aspects of IoT security are discussed along with threats, their mitigation, and the use of SDN and SDN controllers. The article also proposes a new approach toward the development of a secure IoT architecture. The following sections are as described. Section 2 gives an overview of IoT basics. Section 3 describes various IoT security features and challenges of an IoT architecture. Section 4 gives an overview of the existing IoT security techniques in use. Section 5 sheds light on overlooked aspects of IoT security and interoperability. Section 6 proposes the considerations while designing a secure IoT architecture.

Section 7 introduces SDN as a concept. In Sect. 8, we discuss some existing secure industrial IoT frameworks. Section 9 concludes the research discussions and opens the future scope of this research.

2 IoT Basics

This section describes the architecture along with layers in the architecture and the elements present in IoT.

2.1 Architecture

Sadly we do not have a standard architecture. But, after analyzing different IoT architectures, one can look up the layers and characteristics in Fig. 1.

Each IoT layer has a defined role and can be characterized as follows:

- Business layer: It manages and enhances overall IoT system activities and other services such as business model, design, analysis, decision making, monitoring, and implementation. It also maintains users' privacy.
- Application layer: It provides the ability to enable high-quality smart services and covers numerous vertical markets such as surveillance, transportation, and healthcare.

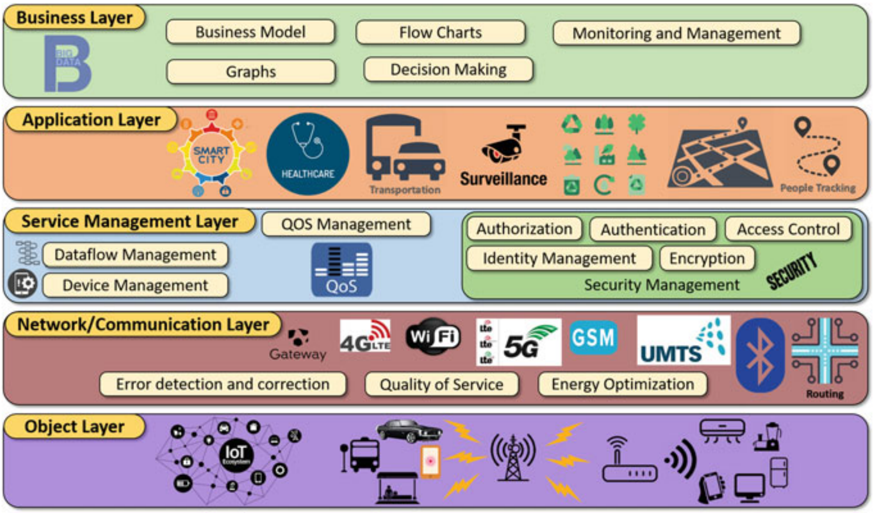


Fig. 1 A General IoT architecture

- Service management layer: This layer identifies the correlation between a service and its requesting entity based on addresses and names. This layer processes the received data, makes correlations, and caters the requested service over appropriate protocols. This layer is responsible for providing data flow and management of devices, quality of service (QoS), and security.
- Network/communication layer: Also known as object abstraction layer, it handles data management processes such as routing, energy optimization, error detection, and correction. It transfers data for computation to the service management layer. Some examples of object abstraction could be in the form of various communication technologies such as GSM, RFID, 3G, UMTS, 5G, Wi-Fi, Bluetooth Low Energy (BLE), ZigBee, and many more.
- Object layer: This layer includes actuators, sensors, and other standardized or non-standardized plug-and-play mechanisms.

2.2 IoT Elements

IoT constitutes various entities that enable core functionalities of IoT communication. Figure 2 depicts these entities [3].

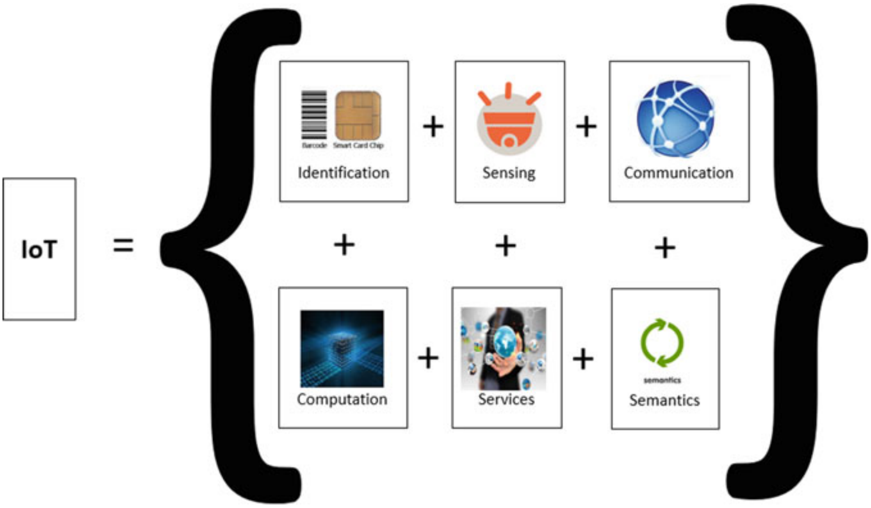


Fig. 2 IoT elements

3 Features and Challenges in IoT Security Architecture

The development of a framework requires an architecture for its description and working. Here we discuss the development considerations for a secure IoT Architecture.

3.1 IoT Security Architecture Features

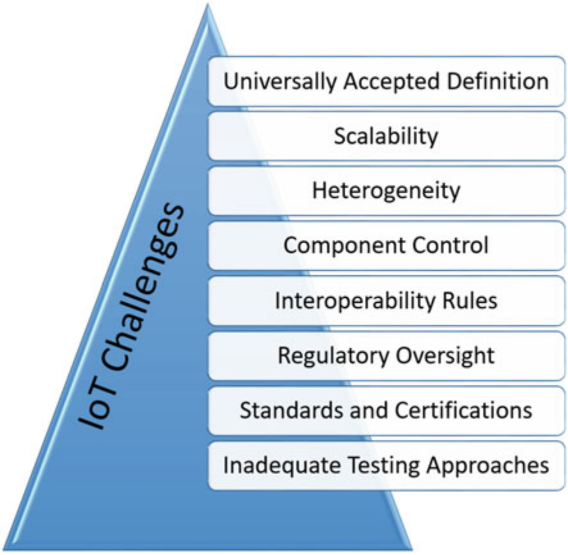
The following features have been introduced to make a complete IoT security architecture:

1. Manufacturers introduce chip security through Trusted Platform Modules (TPMs) that protect sensitive information and credentials which does not allow access to encryption keys outside the chip.
2. Secure booting ensures that only trusted and verified software run on the devices.
3. Intruders may gain physical access to the device. Thus, physical protection offers a guard against tampering, including all the internal circuitry.
4. Data-centric security solutions ensure the safe transition of encrypted data, such that even on interception, it is meaningless, except to the authentic entities (i.e., a person, device, system, or application) possessing the correct decryption key.
5. Encryption of data at rest, i.e., confidential information stored in the cloud, is a must to avoid easy exposure to attacks.
6. Firewalls and intrusion detection systems detect any unwanted access and prevent malicious network layer activities by examining network traffic flows.
7. Verification of other cloud platforms integrity is beneficial in order to prevent malicious activities from third-party application communicating with the cloud services.
8. Digital certificates play a key role in identification and authentication needs at a scale required for the IoT [4].
9. Activity monitoring helps to track, log, and detect suspicious activities.
10. Updates and security patches at regular intervals strengthen resistance against attack by staying up-to-date and fix possible vulnerabilities.
11. When dealing with a significantly large number of devices, a secure remote connection is a must.

3.2 IoT Challenges

The aforementioned desired behaviors have existed in the past IoT practices with definitions from earlier generations of IT and physical systems. Former IoT systems had fewer issues when compared to the present ones. Eight such key differences (Fig. 3) could be described as the following concerns:

Fig. 3 IoT challenges



Universally Accepted Definition: IoT has trust-related issues which could be easily addressed if we all agreed to standardization, i.e., the heterogeneity of devices demands certificate exchange among devices and networks, but lack of any defined structure for these certificates leads to non-compliance and interoperability concerns, thus creating trust issues. Alas, there is no universal definition. IT systems such as cloud computing and fog computing were around for years before a standard definition came into existence.

Scalability: Sensors play an important role in IoT as their abundant presence generates enormous amounts of data for the systems, which many times overwhelms their ability to handle dataflow and workflow to achieve their goals. This data generation demands them to be scalable.

Heterogeneity: IoT consists of various components that vary in size and their development scales to different vendors. If a system is assembled from varying sources (i.e., heterogeneous assembling), it lacks integrity and reliability. IoT trust concerns arise when the homogeneity, integrity, and reliability of the system are not maintained. Also, information and data overload decreases trust in the network as device bottlenecks force them to block connections.

Component Control: With the increase in the use of distributed computing services such as ISP and cloud services, there is a decrease in the sense of ownership. Because of such heterogeneity, there is little to no control over the functioning of a deployed system.

Interoperability Rules: When we fuse two or more environments, it is uncertain whether the resultant environment would surely work without any bugs or crashes.

While scaling a system, new devices need to be attached, thus increasing heterogeneity (as there is a shallow chance that all system requirements could be met with the use of hardware from a single manufacturer) and component quality control (as the components in use may or may not have standardized quality and development of standard drivers is a big concern).

Regulatory Oversight: There are numerous industrial and non-industrial sectors/domains where IoT has no definite future. People involved in politics, governance, and regulation, including NGOs and standardization organizations, have not yet taken up the challenge of creating regulations for technologies. This concern for trust is linked to the lack of supervision and the above-mentioned governance bodies.

Standards and Certifications: Standards are of two types: de-facto and prescriptive, and the IoT community is much more likely to end up with the former as construction of standards takes much time. The same is the case with certifications, as the rapid advancement of certain products in the market cannot be traditionally certified. Safety-critical systems are expensive and require in-depth review. These deployed systems pose as 'gold standard' for future assemblies.

Inadequate Testing Approaches: Testing systems' takes, and because of high market competition and pressure of launching a product first, testing is sometimes skipped. Also, there is no agreed-upon notion of 'how to test networks of things.'

4 Existing Security Techniques

Exploring various technologies used for Web-based IoT security techniques and their comparative analysis is quite important [4, 5]. Stankovic [6] discusses vision for the changes to the future that IoT could make to the world, enumerating eight key research points which alongside inquire about issues inside these subjects, which is somewhat obscure for preliminary research thought. In the connected world, we rely on various apps (that run on android, iOS, Sailfish, and others). Analysis of a popular programming framework reveals that many smart home apps are inherently overprivileged, leaving the risk for remote attacks on users that can cause physical, financial, and psychological harm [7]. They address the possibility of preventing the remote attackers from exploiting overprivilege. A key missing observation is that shortening the privileges can hinder the quickness of the IoT structure. Also, the connection of secure devices requires more resources than a comparatively less secured one.

Investigation of possibilities of sensitive information leakage and analysis of potential threats to privacy in the automatic appliance control (AAC) application is reasonably necessary. The authors in [8] have contributed just the same, along with a proposal of an attribute-based encryption (ABE) key management technique. On the contrary, the system incurs a light overhead over the existing protocol. Introduction of unit and ubiquitous IoT (U2IoT) [9] to address the cybersecurity issues is a step toward security, although it is weak in access time and resource management. Banerjee et al. [10] present an idea of developing a privacy-aware slotted channel access mechanism, to share the wireless channel among the IoT nodes from heterogeneous operators or trusted domains without imparting their identities. It also presents a privacy-preserving TDMA wireless access mechanism. On the contrary, high performance of the system requires high resource utilization, which degrades with less resource availability.

As a network grows, so do the networks vulnerabilities and for industries which scale over large networks, these vulnerabilities pose various threats. Sajid et al. [11] feature some imperative realities about modern SCADA frameworks with accentuation on dangers and vulnerabilities, their administration, and the at present followed rehearses. The target of IoT-based SCADA frameworks is to expand their adaptability, cost productivity, advancement ability, accessibility, and versatility. The content is in-depth; still it does not consider threats that occur because of physical attacks.

IoT has scaled not only to industrial systems but also into our socioeconomic lives. The authors of [12] discuss the implication of IoT which can challenge our legal system's ability to make an appropriate decision when lawsuits arise. They coin a new term, network of things (NoT) which refers to a network of devices in surroundings and proximity. Although by definition, NoT can also be formed over remote locations which affects the definitions of IoT and NoT. Also, it can be challenging to correctly attribute blame when a system interacts with hundreds of leased third-party products and services. Sybil defense schemes [13] include mobile Sybil detection, social graph-based Sybil detection (SGSD), and behavior classification-based Sybil detection (BCSD) with their comprehensive comparisons.

There remains the query regarding inheritance issues with devices that cannot, without much of a stretch, be changed and shape a fundamental piece of a basic framework [14]. Giuliano et al. [15] proposes a calculation that empowers secure access for uni- and bidirectional gadgets using subjective security. This methodology depends on secure key restoration (with no trade in air). The authors introduce a benchmark investigation keeping in mind the end goal to evaluate the viability of their approach regarding other existing models and the security examination as far as regular exploits. The impact likelihood among unidirectional and bidirectional terminals influences the security strategy. With an increment in failures, the inactivity required for bidirectional terminals to accurately convey packet bundles also increments.

Table 1 summarizes the existing security techniques in literature.

5 Inference

The main concern of this research remains over security implementation in IoT networks and devices. In the discussed IoT architectures, no security layer or any other security measure has been mentioned. Security is being neglected again and again in various walks of Internet-based applications. While designing any system, it

Table 1 Survey of existing security techniques

References	Findings	Shortcomings
[6]	Presents a vision for how IoT could change the world in the distant future. Then, eight key research topics are enumerated and research problems within these topics are discussed	Topic discussion is vague for initial research idea. Many important topics such as the development of standards, the impact of privacy laws, and the cultural impact on use of these technologies are outside the scope of this paper
[7]	Analysis of a popular programming framework reveals that many smart home apps are automatically overprivileged, leaving users at risk for remote attacks that can cause physical, financial, and psychological harm. Address the possibility of preventing remote attackers from exploiting overprivilege.	Shortening the privileges can act as a barrier in quickness of IoT structure. Connection of secure devices require more resources than comparatively less secured one
[8]	Investigates possible sensitive information leakages and analyzes potential privacy threats in the automatic appliance control (AAC) application. Proposes an attribute-based encryption (ABE) key management variant	Incurs a light overhead over the existing protocol
[9]	Unit and ubiquitous IoT (U2IoT) to address the cybersecurity issues.	Does not address physical security of devices. Weak regarding access time and resource management
[10]	Proposes an idea is to develop a privacy-aware slotted channel access mechanism using which IoT nodes from multiple operators or trust domains can share wireless channel without mutually exposing their identities. It presents a privacy-preserving TDMA wireless access mechanism for a multitrust-domain network of IoT	Performance is based on high resource utilization which degrades with less resource availability
[11]	Highlights some important facts about industrial SCADA systems with an emphasis on threats, vulnerabilities, management, and the current practices being followed. The objective of IoT-based SCADA systems is to increase their flexibility, cost efficiency, optimization capability, availability, and scalability of such systems	Does not take into account the threats that arise because of physical attacks

(continued)

Table 1 (continued)

References	Findings	Shortcomings
[12]	The authors discuss why the IoT will challenge our legal system to do the right thing when lawsuits arise. They coin a new term network of things (NoT) which specifies to a network of devices in surroundings and close proximity.	The NoT can also be formed over remote locations which affects the definitions of IoT and NoT. No differentiation in judge and the jury. It can be very difficult to properly attribute blame when a system is in contact with hundreds of leased third-party products and services and their interactions
[14]	Raises a question to think about legacy problems with devices that cannot easily be changed and form an integral part of a critical infrastructure	Does not list any particular practice that needs to be taken care of while using various devices
[15]	Proposes an algorithm for secure access for uni- and bidirectional devices by exploiting the cognitive security concept. The security procedure is based on a secure key renewal (without any exchange in air). It presents the benchmark analysis to assess the effectiveness of an approach with respect to other existing standards, and the security analysis in terms of typical attacks	The collision probability among unidirectional and bidirectional terminals affects the procedure largely. The latency required bidirectional terminals to correctly deliver packets increases with collision

is vital to understand the potential threats to it and the addition of appropriate defense mechanisms. Also, various queries need to be answered, like: where should security reside?, where should the security be defined?, should it be a hardware, firmware, or software feature?, does it require a new layer for security protocols?, and would the security protocol affect energy consumption? (Ans: YES, Obviously!) Additionally, the interoperability of different IoT solutions becomes a cumbersome task, as there is no standard architecture.

The remedies could be found in various approaches of networking enthusiasts who tend to use the potential of available resources to their fullest. The answer to threat modeling and mitigation could be found in networking architectures. Thus, we propose the following considerations while designing an IoT security Architecture.

6 Designing an IoT Security Architecture

Designing any architecture requires a detailed analysis of threat models and mitigation of every threat. The following eight steps comprehensively mitigate the threats posed toward any IoT Architecture:

1. Identification of static and dynamic data is necessary. There is a scope of connected IoT devices wherever we find business value of data collection or data.
2. Categorizing IoT devices should be prioritized based on inventory, device discovery, monitoring, remote configuration, and software upgrade. Also, their requirements management is beneficial.
3. Define and organize the new dangers of information misfortune, keeping as a top priority the new vectors that rise because of the discontinuity of inserted working frameworks, systems, and network interfaces.
4. Quantify the danger of unauthorized access to all gadgets. For instance, if a production line automation gadget in an assembling floor or shrewd-therapeutic gadget in a healing facility is compromised, it may have a substantial negative impact on the business.
5. Define the related security activities to be activated, i.e., the conditions under which a compromised device would be removed from the network.
6. Define the big data technique for IoT. Information-oriented security with dynamic data relationship, investigation, and insight stays as a center prerequisite for IoT.
7. Develop protection arrangements for sensor data. The multiplication of sensors will bring about an ever-increasing volume of data being produced, for instance, health data from therapeutic gadgets. The access to and the security of this information will have numerous technical suggestions with restricted direction from existing case law.
8. Ensure the new device associations against targeted attacks and denial-of-service assaults. Undertakings with desired components should achieve it over a substantially extensive configuration of devices.

The SDN can be incorporated to IoT to provide an adaptable, programmable, secure, and dynamic network without disturbing the existing underlying IoT architecture. This answers our query about security placement discussed in Sect. 5.

IoT and SDN are two ever-developing technologies. The IoT aims to provide connectivity among the objects over the Internet. The SDN is responsible for the network management orchestration (open and user-controlled access to communication hardware). It exploits the feature of switches and routers by segregating the control plane and the data plane. The large number of objects leads to network management concerns. Before getting into the implementation of SDN into an IoT architecture, we need to understand this technological concept's basics.

7 SDN Basics

SDN [16] architecture model has three layers Fig. 4.

1. Infrastructure layer: It comprises network devices, e.g., switches (both hardware and virtual), wireless access point, routers.
2. Control layer: It consists of SDN controller(s), e.g., floodlight, NOX, open daylight, POX, MUL, beacon, and many more.
3. Application layer: The application layer includes applications such as virtualizers and software configurations, to configure the SDN workflows. For example, access control, security/traffic monitoring, management of the network, energy-efficient networking.

One of the basic features of SDN design (Fig. 4) is its ability to stretch out the configurations to edge systems (switches, remote routers), by setting up approachable tenets to organize all systems.

An SDN-based architecture comprises of:

– Legacy interfaces, i.e., the physical layer,
– Programmable layer, i.e., SDN-compatible virtual switch and an SDN controller [17]
– OS layer, i.e., operating systems and their applications

All legacy interfaces are connected to a virtual switch, and an SDN controller controls this switch. An SDN Controller can be called the "brain" of the SDN network and is responsible for pushing changes, allowing network administrators to segregate traffic, control flows for optimal performance, and start testing new configurations and applications through an interface, which is a strategic point in SDN.

SDN controllers help in organizing network administration, taking care of all correspondences among applications and systems to viably oversee and modify flows to address evolving issues. At any point when the system control plane is executed in software, as opposed to firmware, administrators can oversee the organized activity with granular access and information. An SDN controller transfers data using southbound and northbound APIs (Fig. 5) to the routers/switches and the applications/business rationale, respectively.

SDN controllers use standard application interfaces such as open virtual switch database (OVSDB), OpenFlow, and others, to observe SDN controller domains. For devices working in an OF environment, support for OpenFlow protocol is a must for their communication to an SDN controller.

- **OpenFlow**: OpenFlow (OF) is considered one of the first SDN standards. For SDN environments, it defines the communication protocol that enables direct interaction of SDN controller with the forwarding plane of network devices, both physical and virtual such as switches and routers, for better adaptation to changing business requirements.
- **OpenFlow Controller**: An OpenFlow controller utilizes an SDN convention, i.e., OpenFlow, to associate and design the system gadgets, for example, routers,

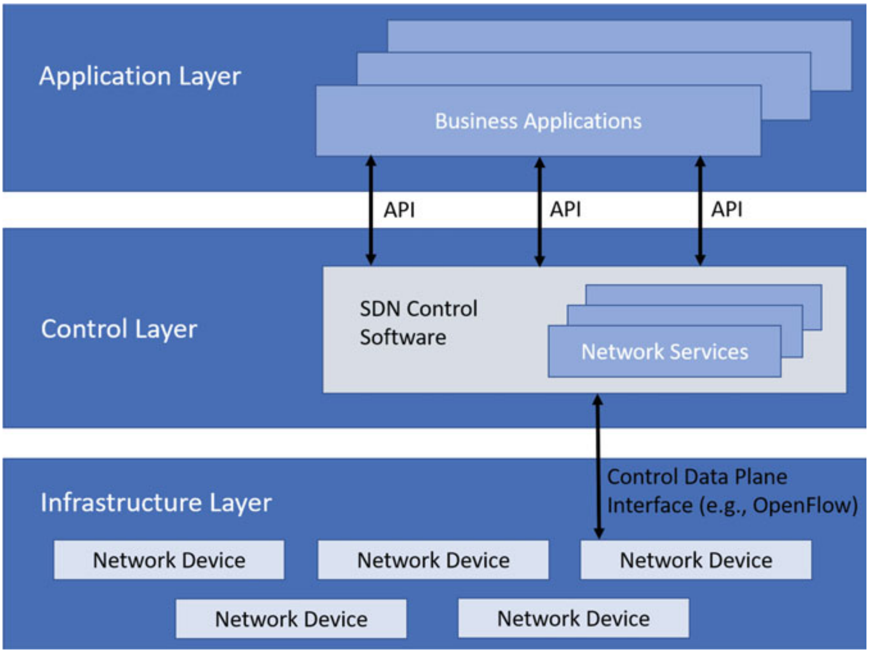


Fig. 4 SDN layers

switches, and so on, to decide the best path for movement of application traffic. OpenFlow controllers go about as a focal control point to supervise an assortment of OpenFlow-empowered system parts. OpenFlow standards outline adaptability by dispensing with exclusive definitions to equipment sellers.

- **OpenDaylight (ODL)**: ODL is an SDN project hosted by the Linux Foundation under open-source license which creates the basis for a strong network functions virtualization (NFV) and advance SDN adoption. The creation poses as a community-led and industry-supported open-source SDN framework. OpenDaylight Project aims to offer a functional SDN platform that directly provisions SDN for users, without any other components. Additionally, vendors and contributors can deliver add-ons and other pieces that can offer more value to OpenDaylight. OpenDaylight utilizes the current open standards while working with leaders like the Open Networking Foundation (ONF). OpenFlow is a prominent example of an SDN protocol supported by OpenDaylight Project, which can incorporate new standards and products as they are created. However, it needs to be considered that SDN is not just a single protocol (i.e., OpenFlow), however beneficial it might be. Therefore, the OpenDaylight Project is intended to configure several SDN interfaces, including, but not static to, OpenFlow.

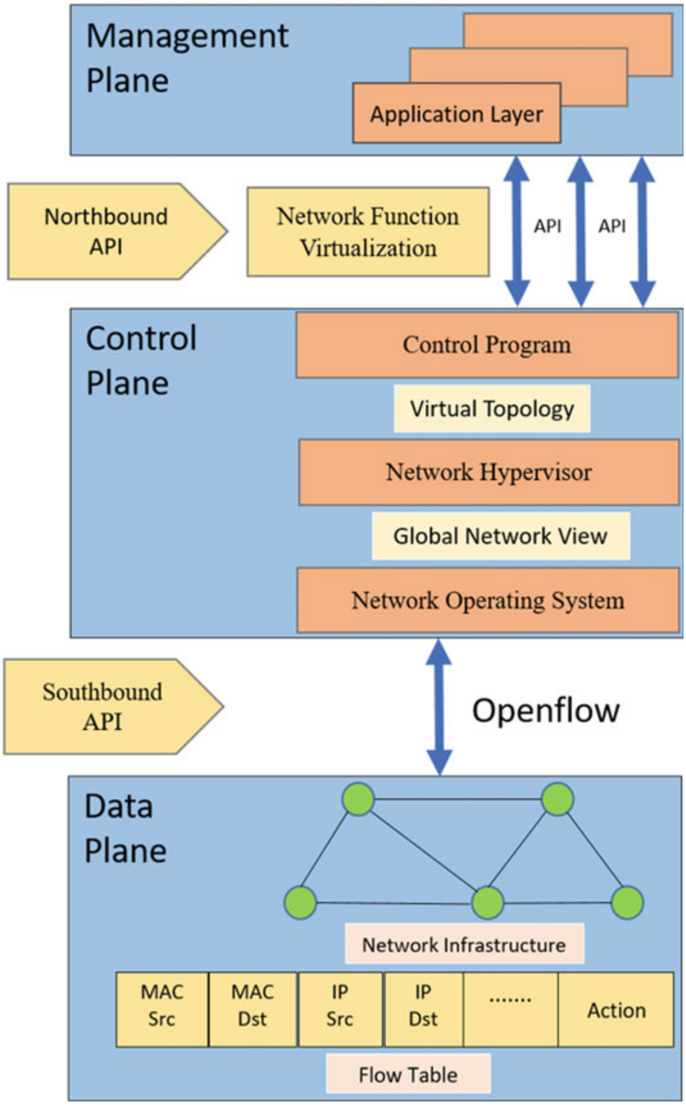


Fig. 5 SDN planes [18]

A brief comparison of various SDN controllers [19] based on the features such as modularity of controller, implementation, support for OpenStack, controller architecture, i.e., whether centralized or decentralized is given in Table 2.

NOX and POX, discussed in the table, have poor documentation as they are propriety and are not widely accessible.

As all controllers operate at equal operational conduct, they do not impact the edge-device liability for the malicious users connecting to the network through them, thus increasing security and decreasing the risk of malicious activities. But the concern over unauthorized network access remains unmoved as the communication network is open to exploits.

8 Existing Secure IoT Frameworks

This section highlights some of the existing IoT security architectures, some of which also incorporate SDN.

8.1 Microsoft Azure Architecture

With a specific end goal to upgrade security best practices, Microsoft prescribes that an average IoT engineering should be segregated into a few segments/zones (Fig. 6). These zones are described as:

- Device
- Field gateway
- Cloud gateways
- Services

Each zone has a custom requirement for its data authentication and authorization. Zones, in like manner, can restrict harm and cut off the effect of low trust zones on higher place stocks in them. Trust boundary isolates each zone. The boundary is shown by the dabbed red line in the graph. It represents the progress of data/control, from one source to the next. Amid this progress, the information/data could be liable to spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privilege (STRIDE).

8.2 IBM SDN-VE

The IBM SDN-VE (Fig. 7) [20] solution provides a complete implementation framework for the network. It offers a core component of the SDN architecture, deployable for data center expansion. With SDN, instead of directly confining each connected

Table 2 Comparison of various SDN controllers

Controller	Features						
	Documentation	Programming language	Modularity	GUI	Platform support	Centralized/ Distributed	OpenStack support
Floodlight	Defined	Java	Basic	Web based	Linux, MAC, Windows	C	N
Beacon	Basic	Java	Basic	Web based	Linux, MAC, Windows	C	N
POX	Poor	Python	Poor	Python + QT4	Linux, MAC, Windows	C	N
NOX	Poor	C++	Basic	Python + QT4	Linux	C	N
MUL	Basic	C	Basic	Web based	Linux	C	Y
OpenDaylight	Elaborate	Java	Advanced	Web based	Linux, MAC, Windows	D	Y
RYU	Basic	Python	Basic	Python	Linux	C	Y

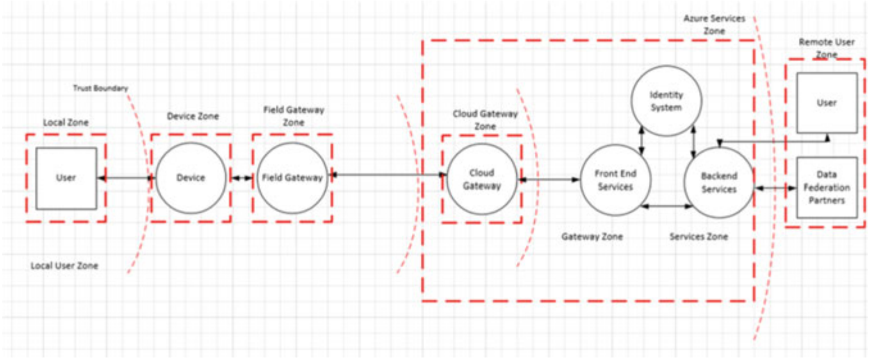


Fig. 6 Azure architecture

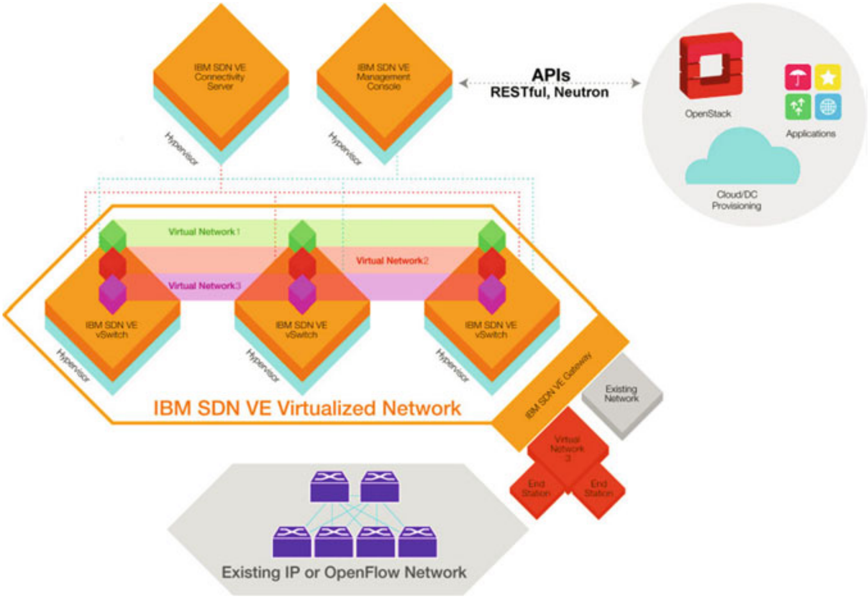


Fig. 7 IBM SDN-VE architecture

device that makes up a network, administrators can establish multiple networks dynamically. It comprises Northbound API, a uniform API enabling the exploitation of underlying network services and the network itself, applications, and platforms of various architectures, which presents the network as a service or infrastructure.

SDN uses high-level control programs to allocate bandwidth and route the data so that administrators can dynamically establish multiple networks. Previously, it had to be directly configured for each connected device that was part of the network. This enables full deploy-ability for data center expansion.

The IBM SDN-VE solution has a single point of control: the SDN unified controller using which resources can be abstracted and utilized in two ways:

1. Overlay: unified network virtualization services based on IBM's Distributed Overlay Virtual Ethernet (DOVE) technology
2. OpenFlow: logical groups (networks), based on OpenFlow technology

8.3 Cisco IoT Security Architecture

Cisco introduces security architecture (Fig. 8) for industrial IoT to strengthen the IoT security. It gives enterprises understanding of their complete IoT environments, including all the networking components [21].

The architecture includes Cyber Vision, which is a security enabler in IoT network and its smart data extraction tool. Cisco Cyber Vision is a software-based security solution for the automated discovery of industrial assets. It can analyze traffic activity from connected components and also prevent threats within operational environments by defining segmentation policies. The architecture aims to simplify IoT cybersecurity management at any of the Cisco-developed devices such as gateways, switches, or routers.

Cisco Software-Defined Access (SD-Access) [22] caters as a solution for Cisco Digital Network Architecture (Cisco DNA) Fig. 9 designed on intent-based networking principles. It isolates user, device, and application traffic with the use of visibility-based automated end-to-end segmentation without altering the design of the physical network.

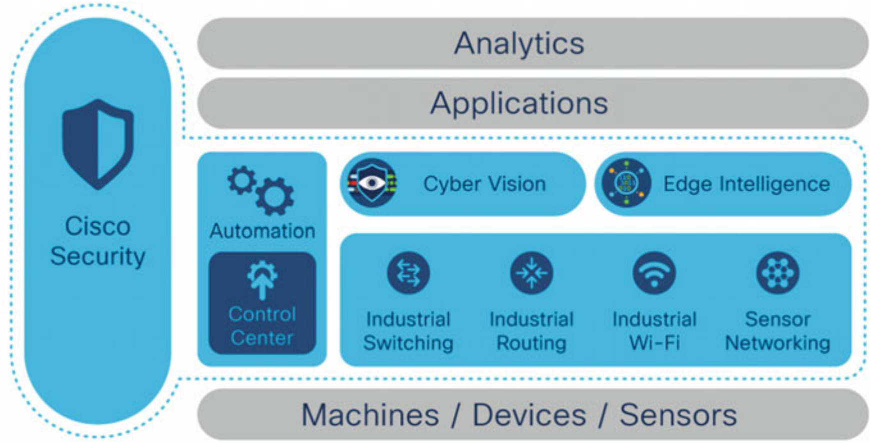


Fig. 8 CISCO secure IoT architecture

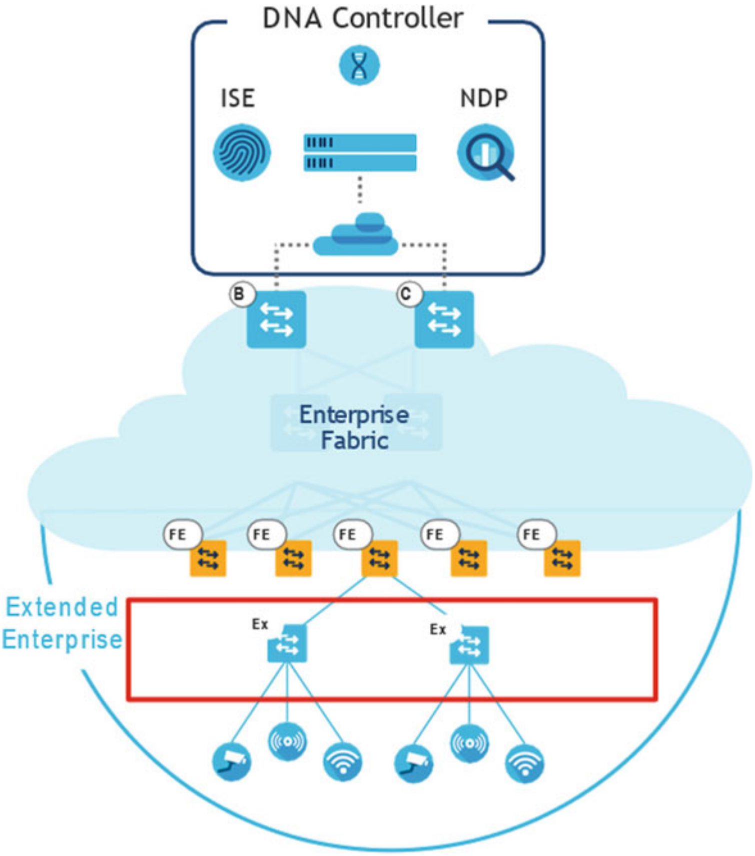


Fig. 9 CISCO-software-defined access architecture for IoT [22]

9 Conclusion

Security is being neglected time and again in various implementations of Internet-based applications. This paper addresses certain aspects of IoT architecture development, keeping *security* as a paramount consideration. For the same, we propose and enumerate eight checkpoints which should be taken into account while planning and designing a framework to comprehend the potential dangers to that framework and include proper barriers as needed. It is especially imperative to plan any task from the beginning on account of security while observing how an assailant may have the capacity to bypass a framework, ensuring that fitting alleviation is set up from the earliest starting point.

During this research, we came across many frameworks that incorporate SDN as an approach to achieve security in IoT networks. The idea of SDN is the simple segregation of a network into different planes, to gain granular access and management of various functions of the network. Although if one tries to take the output of a machine from the zonal approach of Azure architecture and advance with the CISCO DNA controller to devise an input for IBM SDN Virtualized Network, to create a hybrid solution deployable at any platform (on-premise or cloud), interoperability issues will inevitably arise because of presence of private APIs. Thus, these secure IoT frameworks cater to the need of securing the communication architecture in their unique way, but none of the base security frameworks is modular enough to repurpose, replace, or even restructure the other architecture(s). We also provided a brief comparison of various primitive SDN controllers.

References

1. K.D. Chang, C.Y. Chen, J.L. Chen, H.C. Chao, Internet of things and cloud computing for future internet, in *Security-Enriched Urban Computing and Smart Grid*, ed by R.S. Chang, T.h. Kim, S.L. Peng (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011)
2. I. Bojanova, J. Voas, Trusting the internet of things. IT Prof. **19**(5), 16–19 (2017). https://doi.org/10.1109/MITP.2017.3680956
3. M. Burhan, R.A. Rehman, B.S. Kim, B. Khan, Iot elements, layered architectures and security issues: a comprehensive survey. Sensors **18** (08 2018). https://doi.org/10.3390/s18092796
4. S. Phillips, Top 3 security issues in infrastructure iot. IT Professional (2016)
5. M. Coughlin, A survey of sdn security research (2014)
6. J.A. Stankovic, Research directions for the internet of things. IEEE Internet Things J. **1**(1), 3–9 (2014). https://doi.org/10.1109/JIOT.2014.2312291
7. E. Fernandes, A. Rahmati, J. Jung, A. Prakash, Security implications of permission models in smart-home application frameworks. IEEE Secur. Priv. **15**(2), 24–30 (2017). https://doi.org/10.1109/MSP.2017.43
8. D. Li, Z. Aung, J. Williams, A. Sanchez, P3: privacy preservation protocol for automatic appliance control application in smart grid. IEEE Internet Things J. **1**(5), 414–429 (2014). https://doi.org/10.1109/JIOT.2014.2358578
9. H. Ning, H. Liu, L.T. Yang, Cyberentity security in the internet of things. Computer **46**(4), 46–53 (2013). https://doi.org/10.1109/MC.2013.74
10. D. Banerjee, B. Dong, M. Taghizadeh, S. Biswas, Privacy-preserving channel access for internet of things. IEEE Internet Things J. **1**(5), 430–445 (2014). https://doi.org/10.1109/JIOT.2014.2346513
11. A. Sajid, H. Abbas, K. Saleem, Cloud-assisted iot-based scada systems security: a review of the state of the art and future challenges. IEEE Access **4**, 1375–1384 (2016). https://doi.org/10.1109/ACCESS.2016.2549047
12. J. Voas, P.A. Laplante, The iot blame game. Computer **50**(6), 69–73 (2017). https://doi.org/10.1109/MC.2017.169
13. K. Zhang, X. Liang, R. Lu, X. Shen, Sybil attacks and their defenses in the internet of things. IEEE Internet Things J. **1**(5), 372–383 (2014). https://doi.org/10.1109/JIOT.2014.2344013
14. V.G. Cerf, What hath we wrought? IEEE Internet Comput. **21**(4), 103–104 (2017). https://doi.org/10.1109/MIC.2017.2911427
15. R. Giuliano, F. Mazzenga, A. Neri, A.M. Vegni, Security access protocols in iot capillary networks. IEEE Int. Things J. **4**(3), 645–657 (2017). https://doi.org/10.1109/JIOT.2016.2624824
16. D. Kreutz, F.M.V. Ramos, P. Verissimo, C. Esteve Rothenberg, S. Azodolmolky, S. Uhlig, Software-defined networking: a comprehensive survey. ArXiv e-prints (Jun 2014)
17. F. Wang, H. Wang, B. Lei, W. Ma, A research on high-performance sdn controller, in *2014 International Conference on Cloud Computing and Big Data* (2014), pp. 168–174. https://doi.org/10.1109/CCBD.2014.41

18. M. Chahal, S. Harit, Towards software-defined vehicular communication: architecture and use cases, in *2017 International Conference on Computing, Communication and Automation (ICCCA)* (2017), pp. 534–538. https://doi.org/10.1109/CCAA.2017.8229859
19. M. Paliwal, D. Shrimankar, O. Tembhurne, Controllers in sdn: a review report. IEEE Access (06 2018), pp. 1–1 https://doi.org/10.1109/ACCESS.2018.2846236
20. IBM: Ibm software defined network for virtual environments: Network virtualization for the network you have (2013). https://www.research.ibm.com/haifa/dept/stt/papers/QCW03028USEN.PDF
21. V. Butaney, Cisco iot delivers double digit growth across industries (2021). https://blogs.cisco.com/internet-of-things/a-look-back-on-cisco-iot-success-and-investments
22. CISCO: Sd-access (1.1.x) firewall integration (2018). https://www.cisco.com/c/en/us/solutions/collateral/enterprise-networks/software-defined-access/white-paper-c11-741103.pdf

Distributed Identity and Verifiable Claims Using Ethereum Standards

Aju Mathew Thomas, R Ramaguru, and M Sethumadhavan

Abstract Identity management is an inevitable part of the proper delivery of schemes and services to individuals by the government and private organizations. Identity management encompasses the creation and maintenance of identity. It ensures that the right entity gains access to the right resources at the right time for verification. It also involves providing high security, privacy, productivity, and enhanced user experience. Rising incidents of data infringements and identity thefts in a centralized identity system are a growing concern. Blockchain-based identity solutions gained a competitive edge over the present centralized identity system due to the features of Self-Sovereign Identity (SSI) and Verifiable Claims (VC). In this paper, we investigate various blockchain-based identity solutions that are self-sovereign and can create VC. We have also analyzed and proposed to create a user-centric identity and claims using ERC-725 and ERC-780 Ethereum standards powered by IPFS for distributed data storage.

Keywords Distributed identity · DID · Self-sovereign · Verifiable claims · Ethereum · IPFS · ERC-725 · ERC-780

1 Introduction

Expeditious growth in technology has led to the development of establishments in the digital world drastically. Identity management (IM) has become one of the significant systems in this digital era that efficiently creates and manages the digital identity.

A. M. Thomas (✉) · R. Ramaguru (✉) · M. Sethumadhavan
TIFAC-CORE in Cyber Security, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India
e-mail: cb.en.p2cys19001@cb.students.amrita.edu

R. Ramaguru
e-mail: r_ramaguru@cb.amrita.edu

M. Sethumadhavan
e-mail: m_sethu@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_48

These digital identities help uniquely identify the individuals and real-world assets like property, artwork, credentials, and much more. It is a crucial and tedious task to digitally identify individuals, organizations, and devices efficiently with high interoperability among multiple systems to receive and provide services [1]. Security and privacy should be given with utmost attention when creating an entity's identity, and it should comply with the national and international regulations like the European Union's General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), and the proposed Personal Data Protection (PDP) Bill of India. Presently, IM is closely associated with confidentiality, non-repudiation, and other security aspects, which are the underpinning features of any trusted environment.

The primary goal of IM is to ensure that authorized users gain access to appropriate resources as quickly as possible. A few common problems in traditional paper-based identities are loss, theft, and various fraud and impersonation activities. These issues typically happen for government-issued identities like national identity (Aadhaar), birth certificate, financial identity, and educational identity. Digital identities have already become part of our life, but the issues related to security and privacy are a significant concern for corporates and institutions. Digital identities can lower the level of bureaucracy and expedite the processing speed within organizations. It provides greater interoperability among various departments and institutions. However, it can lead to a honeypot for hackers if all this information is available in centralized systems [2].

The primary model of IM is the Isolated Identity Model (IIM), in which both the identity provider and service provider are paired together. In the Centralized Identity Model (CIM), there is no pairing between an identity provider and the service provider. User authenticates with the identity provider before accessing the service from the corresponding provider [5]. In India, the Aadhaar unique identification system employs CIM, in which the user authenticates his or her identity via biometrics before accessing any service. The centralized model suffers from various internal and external attacks. Perhaps, one of the most significant data leaks that ever happened, which affected nearly 120M users, was the Jio Data Leak in 2017. Personally Identifiable Information (PII) like subscriber name, address, phone number, and call records was made publicly available. It was later sold for a significant sum on the dark web by hackers and cybercriminals [3]. Another similar incident that made headlines was Dominos India's massive data breach [4] on May 25, 2021, which exposed the order details for 18 crore pizza orders placed through their services. Customer names, credit card numbers, mobile phone numbers, and location information are included in the 130 TB data dump. The attackers created a web page on the dark web to display the details, and information about the order can be retrieved by entering the email address or phone number, which was later made public.

Single Sign-On (SSO) schemes widely seen on Google and Facebook are based on the Federated Identity Management Model (FIMM), in which multiple service providers form an association with one or more identity providers. It allows users to use the same identity credential for authenticating and accessing multiple services [5].

Self-Sovereign Identity (SSI) [6] is wholly owned, controlled, and supervised by the user without any dependency on any third party. In the SSI model, the user generates her or his own digital identity. It can be multiple identities with several attributes per user, and it purely depends on the use case. It offers users to store their credentials on their phones or in cloud repositories [5]. Users add these identities to the blockchain, which is cryptographically secured. Decentralized Identifier (DID) [7] is one of the main components of SSI, which is the standard for World Wide Web Consortium (W3C) for globally unique, persistent decentralized identifiers. DID is a URL that points to our identity. It is associated with a document that includes public keys, authentication, claims, and service endpoints for signature verification and encryption. It is expressed as the linchpin of SSI and uses blockchain or other Distributed Ledger Technology (DLT) to protect privacy and security concerns. Similarly, a Verifiable Claim (VC) is a piece of information representing a set of claims about specific attributes of an entity. VC is cryptographically secure and trustworthy. It is generally issued by an issuer to a user based on the user's request and can be validated by a verifier using the public address of the issuer DID [7]. Blockchain-based IM solutions are seen as the next big revolution to tackle the problems faced mainly by the government to disburse schemes to qualified beneficiaries. Here, the identity and specific attributes are stored in the blockchain, and these attributes can be verified through a VC. Blockchain ledger is a data structure similar to the linked list where every block is a container linked to each other cryptographically. Blockchain Technology is a decentralized computation and distributed ledger platform to immutably store transactions in a verifiable manner efficiently, through a rational decision-making process among multiple parties in an open and public system [8]. Blockchain technology has seen wider adoption in the last decade disrupting automotive industry [9], digital rights [10], health care [11], manufacturing industries [12], supply chain [13], and e-governance as shown in Fig. 1.

Blockchain Technology is more suitable for IM as this can allow individual users to have complete control toward their own identity and management. DID of the individual user, access log to the DID by any entity, Verifiable Claims, and status are recorded immutably on the blockchain. Blockchain can also be extended to SSO models for data distribution and replication. Current state-of-the-art SSO systems store data in traditional centralized databases and retrieve it using Structured Query Language (SQL). When combined with SSO systems, blockchain technology protects data from tampering and prevents SQL injection. It avoids the single-point-of-failure problem and consumes less CPU power than centralized systems. However, the response time is slower than the traditional centralized SSO systems. This paper has proposed and developed a user-centric DID using the ERC-725 and VC using ERC-780 standards that live on Ethereum blockchain. Our work also highlights a sample use case where this approach can be implemented.

The rest of the paper is organized as follows. Section 2 discusses the state-of-the-art that accentuates different proposals of blockchain-based identity solutions. In Sect. 3, we detail our architecture, implementation using ERC-725 and ERC-780, and use cases. Section 4 summarizes and concludes the paper.

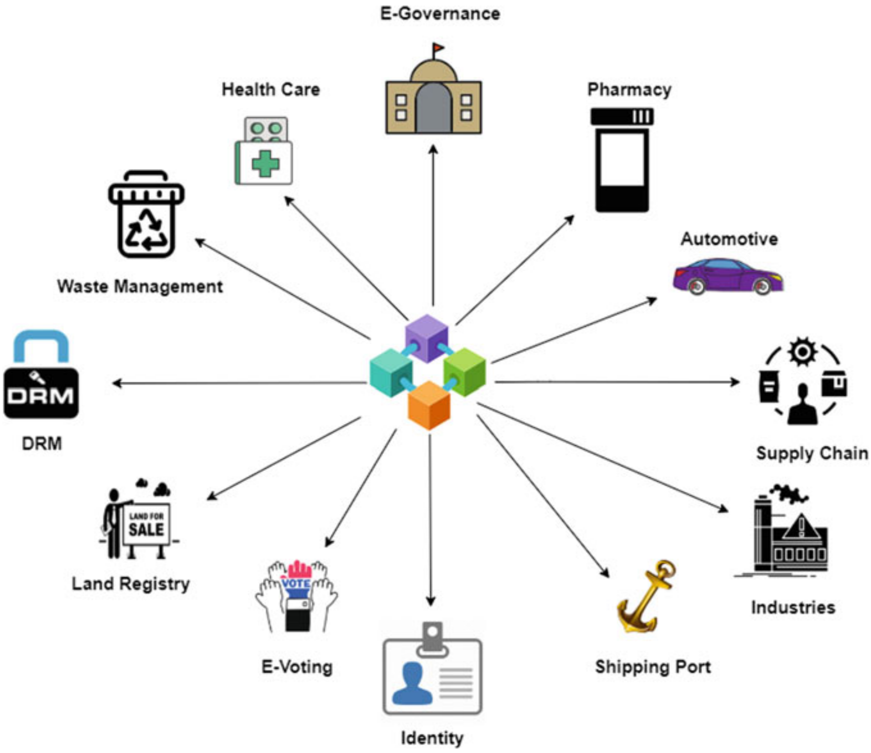


Fig. 1 Application of blockchain in different domains

2 Related Works

This section describes the evolution of various IM models. An entity can have multiple digital identities, denoted with an identity in a specific application domain. The augmentation in online-based applications and services has engendered more secure and privacy-focused digital identity solutions. Shashank et al. [14] conducted an exploratory study on the significance of SSI and highlighted its advantageous characteristics when combined with blockchain technology. The paper also provided an overview of various identity models that evolved and their shortcomings. Additionally, the authors conducted a comparative study of various identity models and their associated benefits when various parameters were considered. Another piece of research that gained traction was Soltani et al. [5] introduction of the Know Your Customer (KYC) 2 framework for banking purposes, which is based on the SSI, GDPR, and privacy by design standards. The Estonian government's initiative to issue an SSI-based electronic ID (eID) to all citizens is worth mentioning. The eID uses blockchain technology to ensure the security of government data and to reduce internal risks. The eID provides users with digital access to all of its e-services, including e-governance, e-tax, e-banking, i-voting, e-health record, e-law, and e-residence, to name a few. The initiative proved to be both time and cost-effective, and easily verifiable [15]. The following section discusses some of the most popularly used identity solutions to date.

2.1 Dock

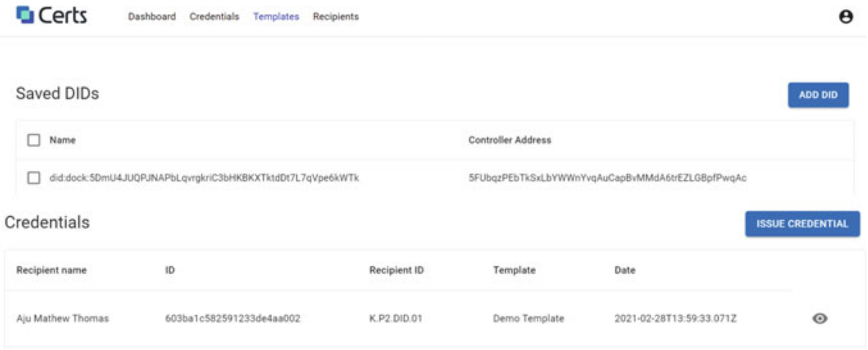
Dock [16] is a decentralized and open-source network for issuing verifiable credential that is cryptographically secured, tamper-resistant, and secured via purpose-built public blockchain. It employs W3C standards to make it interoperable with other W3C compliant solutions.

The issuance of a driving license by a state government and the awarding of an academic degree by a university are two real-world examples where Dock can be used. Issuers can use Dock to generate DIDs that can be stored on chain and used by verifiers to validate credentials issued by the issuing entity. DIDs are embedded in all entity credentials and aid in proving the entity's identity. Figure 2 shows a sample academic credential issued through Dock.

2.2 Evernym's Connect.Me and Verity

Connect.Me [17] is an SSI digital wallet app for mobile devices that enables user to securely share, manage, and store credentials with others. The credential is stored on the user's device and provides trust, privacy, and security. It uses the zero-knowledge proof (ZKP) principle to allow for selective information disclosure to verifiers.

Verity [18] is for issuing and verifying digital credentials. It offers secure and private interaction with end-users in the SSI domain. Some of the outstanding features



Saved DIDs

□ Name	Controller Address
□ did:dock:5DmU4JUQPJNAPbLqvrgkriC3bHKBKXTktdDt7L7qVpe6kWTk	5FUbqzPEbTkSxLbYWWnYvqAuCapBvMMdA6trEZLGBpfPwqAc

Credentials

Recipient name	ID	Recipient ID	Template	Date
Aju Mathew Thomas	603ba1c582591233de4aa002	K.P2.DID.01	Demo Template	2021-02-28T13:59:33.071Z

Fig. 2 Academic degree issued through Dock

of Verity 2.0 are user image as part of credentials and deployment in a private cloud environment. Currently, International Air Transport Association (IATA), a trade body that works with major airlines worldwide, leverages Verity's use for issuing a digitized passport to the passengers for contactless travel amid the COVID-19 pandemic.

2.3 Bloom

Bloom [19] is a back-to-back identity attestation, risk assessment, and credit scoring protocol. Bloom is built on Ethereum blockchain and powered by IPFS for data storage which offers a self-sovereign identity platform with complete identity control and quicker verification. The platform lowers the risk of identity theft and minimizes costs related to customer on-boarding, compliance check, and fraud prevention.

Bloom also provides a radar service that searches the Internet and the dark web and alerts about possible data breaches if the user's data is a part of it. Bloom also provides a strength score which increases whenever more identities and services are verified. Figure 3 shows Bloom IDs of two different users with different attributes like strength and number of breaches.

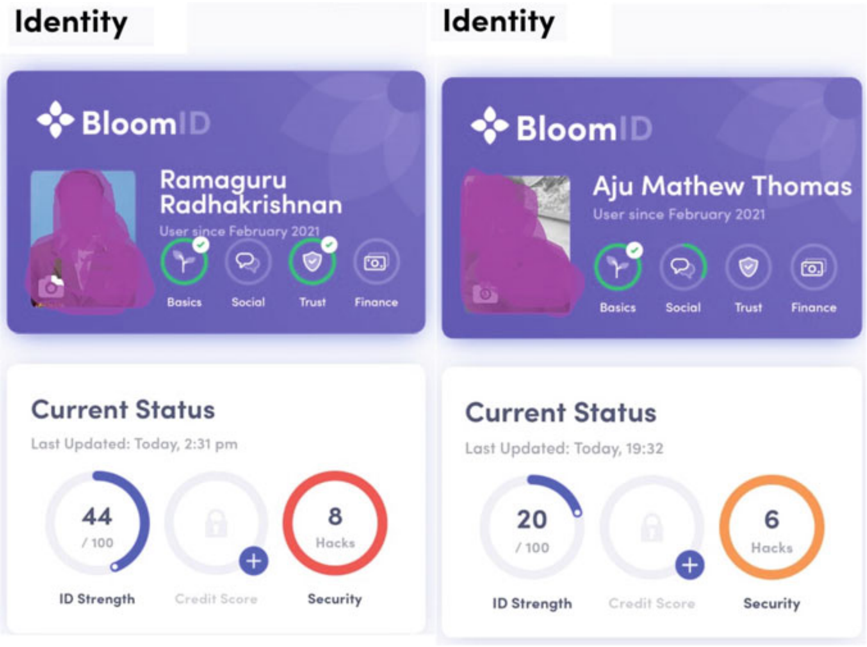


Fig. 3 Bloom ID application

2.4 Microsoft's ION

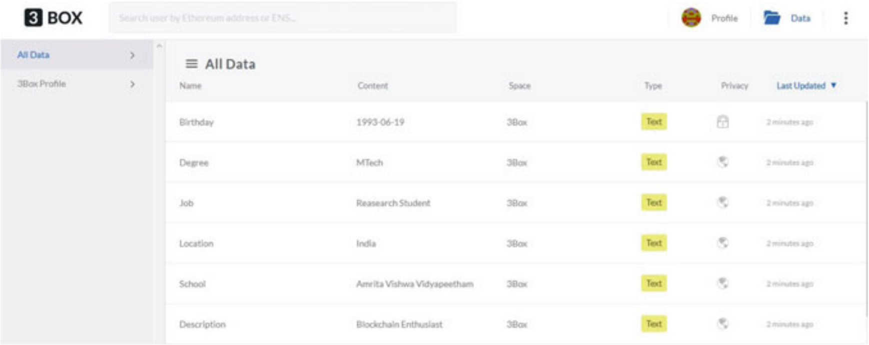
The Identity Overlay Network (ION) [20] is a public permissionless DID network that runs on top of Bitcoin at Layer 1 and uses the Sidetree protocol at Layer 2 to create scalable Decentralized Public Key Infrastructure (DPKI) networks. This protocol allows a user to create globally unique and user-controlled identifiers that comply with the specifications defined by W3C DID. Microsoft Corporation also offers the feature of VC on the Azure Cloud platform. The system offers an efficient and quicker verification process and establishes trust across the subject, issuer, and verifier. Currently, Microsoft aims to implement this methodology on skill verification by presenting the VC on professional platforms such as LinkedIn to speed up the hiring process [21].

2.5 3Box and 3ID Connect

3ID Connect [22] is a secure way for applications to authenticate user-controlled 3Box accounts (identities, profiles, and data stores) and communicate with ceramic-based web applications using blockchain-based wallets. Figure 4 shows how an Ethereum wallet address can be used to create a 3Box profile. 3Box accounts use IPFS for decentralized storage and OrbitDB for data structuring to create DID. The user has authority over the privacy of the information stored in his or her 3Box profile.

2.6 KILT

KILT [23] network is built upon a parity substrate blockchain framework. It uses the KILT protocol that allows the users to issue VC in Web 3.0. These SSI credentials



All Data

Name	Content	Space	Type	Privacy	Last Updated ▼
Birthday	1993-06-19	3Box	Text	🔒	2 minutes ago
Degree	MTech	3Box	Text	🌐	2 minutes ago
Job	Reasearch Student	3Box	Text	🌐	2 minutes ago
Location	India	3Box	Text	🌐	2 minutes ago
School	Amrita Vishwa Vidyapeetham	3Box	Text	🌐	2 minutes ago
Description	Blockchain Enthuslast	3Box	Text	🌐	2 minutes ago

Fig. 4 3Box profile created using Ethereum wallet address

preserve anonymity and are revocable. The user can present this credential to the service providers who are the verifiers. The verifier verifies the credential's validity by comparing its hash with the value stored on the blockchain. KILT protects user privacy by design by allowing them to hide sensitive personal information that is not required for verification.

2.7 Sovrin

Sovrin [24] is an open-source and first global public permissioned ledger for SSI and VC. Sovrin was designed to address the four primary requirements of SSI, namely governance, scalability, accessibility, and privacy. All Sovrin identifiers and public keys are written under a false name by default. Sovrin allows users to create multiple identifiers for each service. Each node in a Sovrin network is regarded as either a validator node or an observer node.

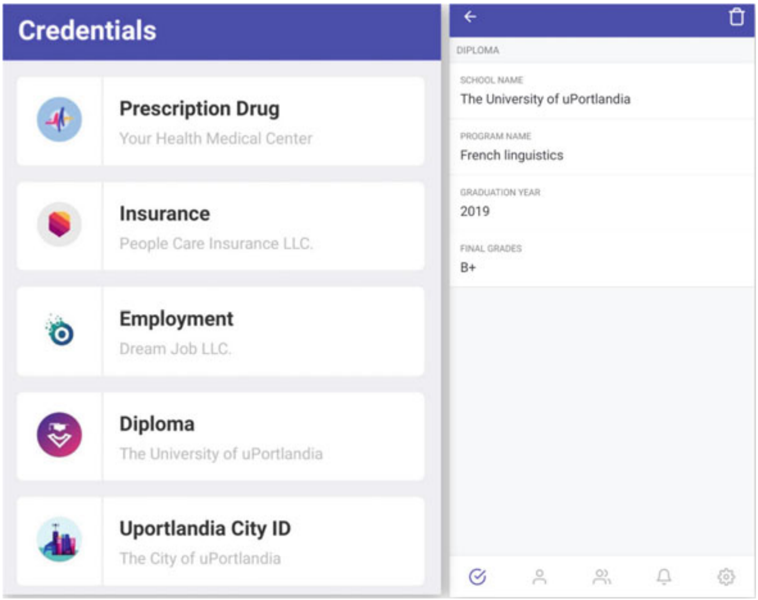
2.8 Serto (uPort)

Serto, formerly uPort [25], is an Ethereum-based SSI solution that aims to provide users with identity through a mobile-based application. The Serto (uPort) identifier is a blockchain Ethereum address that enables users to store their information on their devices, including the private key used for signing and sharing claims. A smart contract links identity attributes to the uPort ID. Figure 5 shows the sample credentials requested from uPortlandia. The contact of another uPort account can be added by scanning its QR code. The contact features the user's photograph and uPort address.

Serto (uPort) has two major drawbacks: (1) It lacks portability because only other uPort identities will attest to the claims; (2) it lacks interoperability because it is primarily intended for use on the Ethereum blockchain. One of Serto's (uPort's) realistic use cases is the issuance and verification of resident claims in Zug [26], Switzerland. Citizens use the uPort app to establish their digital self-sovereign identities and have their VC attested by a government authority after in-person verification to be eligible for government e-services. Serto (uPort) ID has a few advantages, including low infrastructure requirements, GDPR compliance, cost-effectiveness, scalability, and reduced security risk .

2.9 Hyperledger Indy

Hyperledger Indy is one of the frameworks in Hyperledger projects for creating and managing DIDs. Indy represents the concept of SSI that makes it possible and prac-



Credentials

Prescription Drug
Your Health Medical Center

Insurance
People Care Insurance LLC.

Employment
Dream Job LLC.

Diploma
The University of uPortlandia

Uportlandia City ID
The City of uPortlandia

DIPLOMA

SCHOOL NAME
The University of uPortlandia

PROGRAM NAME
French linguistics

GRADUATION YEAR
2019

FINAL GRADES
B+

Fig. 5 Credentials from uPortlandia

ticable for individual users, institutions, and IoT devices by making it interoperable across multiple domains, applications, organizational silos. One of Indy's use cases is digital documents where it allows user to have a secure and versatile digital version of essential documents like passport, driver's license, and birth certificates [27].

2.10 RubiX DID

The RubiX [28] Network uses a deterministic state machine protocol called RubiX ProofChain and custom-build Proof of Harvest (PoH) to address the shortcomings of scalability, privacy, cost-related challenges of existing blockchain protocols along with environmental considerations. A DID uniquely identifies each node that is part of the RubiX Network. RubiX DID is generated using a self-selected 256 $*$ 256 PNG image and the user details that can be verified by peers using a Non-Linear Secret Sharing (NLSS) cryptographic algorithm. Unlike existing models where it provides a Layer 2 solution built on Bitcoin or Ethereum, RubiX DID offers a complete Layer 1 solution.

3 Proposed Work

In this section, we are detailing our proposed framework for Ethereum-based DID. The proposed DID scheme is user-centric, anonymous, revocable, decentralized based on Ethereum standards with storage powered by IPFS.

3.1 Architecture

Various components used in the proposed framework are described as shown in Fig. 6.

Ethereum is a public permissionless and second-generation blockchain that supports smart contracts and user-defined token functionalities. Ether (ETH) is the native cryptocurrency that is used for paying transaction fees in the form of gas. Ethereum offers multiple standards for tokens and identity management. Ethereum also provides permissioned and enterprise-level blockchain.

Smart contract is a self-executing computer or a digital agreement between two or more parties written and executed on top of the blockchain. Smart contracts are not equivalent to legal contracts, but it codes the legal agreement between the parties in the computer understandable language. We have used Solidity programming for implementing smart contracts.

Ethereum blockchain provides multiple standards for Identity and Claim Management. ERC-725, an identity protocol that applies to proxy smart contracts to manage multiple keys and other smart contracts, was used. Figure 6 shows the ERC-725

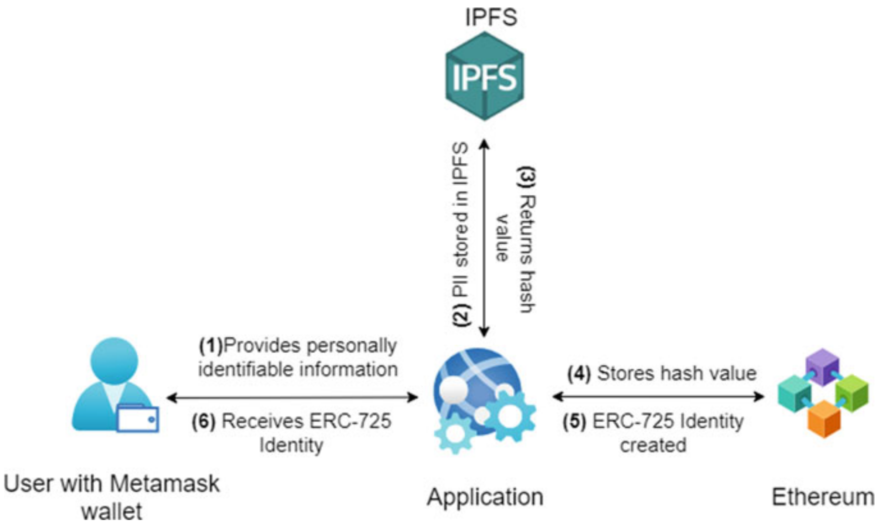

IPFS
IPFS
(2) PII stored in IPFS
(3) Returns hash value
(1)Provides personally identifiable information
(6) Receives ERC-725 Identity
(4) Stores hash value
(5) ERC-725 Identity created
User with Metamask wallet
Application
Ethereum

Fig. 6 Decentralized identity creation using ERC-725

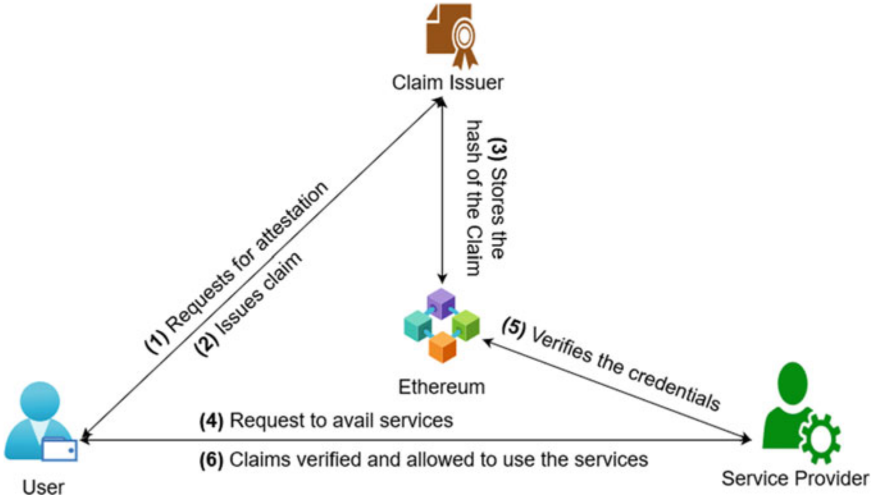
identity account creation process through MetaMask wallet. Identity contracts based on ERC-725 can represent persons, objects, and machines [29]. The ERC-725 (v.2) standard is further divided into two parts:

1. **ERC-725X** is used to execute any random smart contract and can be used to deploy other smart contracts. ERC-725X incorporates the ERC-165 standard for contract interactions. It functions as a validator to check whether the user's smart contract invokes the required smart contract.
2. **ERC-725Y** is used for storing any arbitrary data via a generic key-value store.

ERC-735, an ERC-725 ancillary standard for adding or withdrawing claims, is used along with the older version of ERC-725 (v.1) [29]. Adding and revoking claims became standardized with the introduction of ERC-735, paving the way for a central reference point. ERC-735 leverages the zero-knowledge proof (ZKP) concept and aids in selective disclosure of information with the verifying entity. These features are restricted to Ethereum accounts or smart contracts. The current version of ERC-725 is v.2, and it is not fully compliant with ERC-735.

ERC-780 [30] is a proposed standard in Ethereum for an on-chain claim registry. A Verifiable Claim consists of two parts: a claim and a signature from the identity owner that makes a claim. The on-chain claim is not ideal for holding private information, and in such cases, off-chain claims signed by a DID are used with JSON Web Tokens (JWT), where the issuer signs the claim data and encodes it. Figure 7 shows the claim issuance process using ERC-780.

InterPlanetary File System (IPFS) [31] is a hypermedia protocol and peer-to-peer network system that enables the creation of a distributed file system to store and distribute data. IPFS is a content-addressed protocol that uses the hash value of each



Claim Issuer
(3) Stores the hash of the Claim
(1) Requests for attestation
(2) Issues claim
Ethereum
(5) Verifies the credentials
(4) Request to avail services
(6) Claims verified and allowed to use the services
User
Service Provider

Fig. 7 Claim issuance and verification using ERC-780

file in the global namespace to uniquely identify and locate it. Each IPFS node is identified by a peer ID used to efficiently store and retrieve data within the network using directed acyclic graph (DAG). The file can be accessed only if the hash value of the content addressed is known. IPFS is used in conjunction with blockchain to store large volumes of data to reduce the cost of transactions within the blockchain while also securely storing and identifying them. IPFS is available in both private and public modes. To comply with GDPR requirements, we propose to use private IPFS in our proposed work.

3.2 Insurance Use Case

Figure 8 illustrates the workflow for managing claims between a user and an insurance company via a government-issued DID that adheres to the ERC-725 standard. The user submits a request to the government for a decentralized identity via a dedicated application by using the Ethereum wallet, and after verification, the ERC-725 identity wallet is issued. For better privacy and enhanced security, user-related information is stored in a private IPFS. The chain contains a reference to the hash value returned from the private IPFS. When the user later wishes to obtain a new policy or claim against an existing policy, the user creates a VC by providing the necessary information. The government attests to the user's claim and stores the corresponding hash on the chain. Then, the user attaches the attested claim to the corresponding ERC-725 identity. Finally, the insurance company verifies this claim by comparing its hash value to the chain's hash value. Following successful claim verification, the insurance company can now issue the policy to the requested user.

In this case, the claim can be issued and verified in a matter of minutes to hours, as opposed to the traditional claim process, which typically takes several days to receive a policy from any insurance company. Additionally, the identity, the government-issued claim attestation, and the verification are stored in the blockchain as an immutable record for future auditing and accounting.

3.3 Discussion

We have created an ERC-725 contract which consumed 4150244 gas for deploying in Ropsten test network. Fourteen parameters were selected: name, date of birth, gender, blood group, contact number and address, email ID, and biometric details. Biometric information such as fingerprints (left and right thumb), iris scans (left and right), and a photograph are stored in a private IPFS for enhanced privacy and security under regulatory requirements. On average, it costs 3139050.57142 gas to create a DID for a user. The transaction cost is reduced due to private IPFS; irrespective of the size of the biometric data, the block will store only the IPFS hash value.

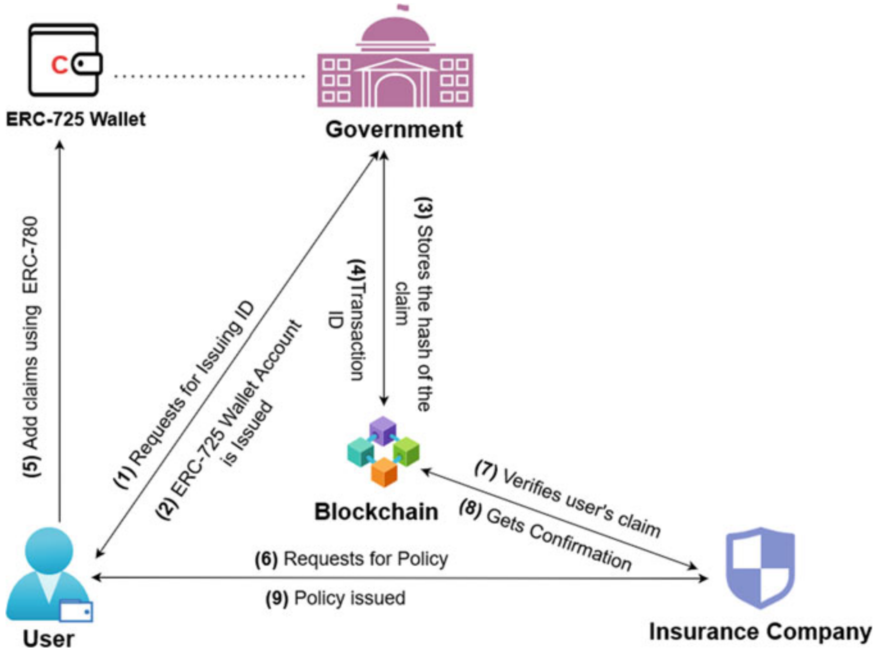

ERC-725 Wallet
Government
(3) Stores the hash of the claim
(4)Transaction ID
(5) Add claims using ERC-780
(1) Requests for Issuing ID
(2) ERC-725 Wallet Account is Issued
Blockchain
(7) Verifies user's claim
(8) Gets Confirmation
(6) Requests for Policy
(9) Policy issued
User
Insurance Company

Fig. 8 Verifiable Claims for policy issuance

Table 1 Comparison between various identity solutions

Parameters/Solutions	Indy	ION	Dock	3Box	uPort	RubiX	ERC-725
Zero-knowledge proof	√	√	√		√	√	√
Layer 1						√	
Layer 2	√	√	√	√	√		√
Verifiable credential	√	√	√		√	√	√
W3C DID compliant	√	√	√	√	√		

We compare different key parameters of current existing solutions with our proposed framework, as shown in Table 1.

4 Conclusion

This paper proposes incorporating a user-centric identity and claims system based on the ERC-725 and ERC-780 standards. However, because identity management is such a vast and intricate subject, there is much room for in-depth research. We highlighted the challenges that currently exist with existing identity management systems and the repercussions that organizations must bear in data infringement. The use of private IPFS to store user data improves privacy and restricts information access within the private network's nodes. By leveraging blockchain technology, our solution improves the overall security of the entire ecosystem. Based on our analysis, it can be concluded that since Ethereum blockchain is widely used and being the second largest market cap, we suggest that a user-centric identity system using the ERC-725 and ERC-780 standards can be considered as a viable alternative for centralized and traditional identities.

The proposed Ethereum Verifiable Claims (ERC-1812) standard can replace the ERC-780 standard for registering off-chain VC. This new standard helps to meet specific regulatory requirements such as KYC compliance and GDPR. Currently, we have used the ERC-780 standard to make on-chain claims, which is not recommended when PII is involved. Additionally, generating claims using ERC-780 is a time-consuming process because it requires key management for each claim issued. When privacy is a consideration, an off-chain claim is preferred. ERC-1056, a lightweight Ethereum DID that uses a wallet address, is recommended for a cost-effective DID creation.

References

1. M. Kuperberg, Blockchain-based identity management: a survey from the enterprise and ecosystem perspective. IEEE Transactions on Engineering Management **67**(4), 1008–1027 (2020). https://doi.org/10.1109/TEM.2019.2926471
2. Blockchain Identity Management: The Definitive Guide (2020 Update). https://tykn.tech/identity-management-blockchain. Last Accessed: 22 Jun 2020
3. J. Anwer, Jio data leak explained: is your data safe, is aadhaar also leaked and other questions answered. India Today, 10 July 2017. https://www.indiatoday.in/technology/features/story/jio-data-leak-explained-is-your-data-safe-has-aadhaar-also-leaked-and-other-questions-answered-1023395-2017-07-10
4. T. Desk, Dominos data breach: name, address, other details of over 18 crore orders leaked. The Indian Express, 25 May 2021. https://www.indianexpress.com/article/technology/tech-news-technology/dominos-data-breach-name-address-other-details-of-over-18-crore-orders-leaked-7328416
5. R. Soltani, U. Trang Nguyen, A. An, A new approach to client onboarding using self-sovereign identity and distributed ledger, in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (Halifax, NS, Canada, 2018), pp. 1129–1136. https://doi.org/10.1109/Cybermatics_2018.2018.00205
6. P. Dunphy, F.A.P. Petitcolas, A first look at identity management schemes on the blockchain, *in IEEE Security & Privacy*, vol. 16, no. 4 (July/August 2018), pp. 20–29. https://doi.org/10.1109/MSP.2018.3111247
7. C. Brunner, U. Gallersdörfer, F. Knirsch, D. Engel, F. Matthes, DID and VC: untangling decentralized identifiers and verifiable credentials for the web of trust, in *2020 the 3rd International Conference on Blockchain Technology and Applications. Association for Computing Machinery* (New York, NY, USA, 2020), pp. 61–66.https://doi.org/10.1145/3446983.3446992
8. R. Ramaguru, M. Minu, Blockchain terminologies. NamChain Open Initiative Research Lab (2021). https://github.com/NamChain-Open-Initiative-Research-Lab/Blockchain-Terminologies

9. R. Ramaguru, M. Sindhu, M. Sethumadhavan, Blockchain for the internet of vehicles, in *Advances in Computing and Data Sciences. ICACDS 2019. Communications in Computer and Information Science*, ed. by M. Singh, P. Gupta, V. Tyagi, J. Flusser, T. Ören, R. Kashyap, vol 1045. (Springer, Singapore, 2019). https://doi.org/10.1007/978-981-13-9939-8_37
10. M. Kripa, A. Nidhin Mahesh, R. Ramaguru, P.P. Amritha, Blockchain framework for social media DRM based on secret sharing, in *Information and Communication Technology for Intelligent Systems. ICTIS 2020. Smart Innovation, Systems and Technologies* ed. by T. Senjyu, P.N. Mahalle, T. Perumal, A. Joshi, vol 195. (Springer, Singapore, 2021). https://doi.org/10.1007/978-981-15-7078-0_43
11. M. Abraham, A.H. Vyshnavi, C. Srinivasan, P.K. Namboori, Healthcare security using blockchain for pharmacogenomics. J. Int. Pharmaceut. Res. **46**, 529–533 (2019)
12. D. Sivaganesan, Smart contract based industrial data preservation on block chain. J. Ubiquitous Comput. Commun. Technol. (UCCT) **2**(01), 39–47 (2020)
13. A.B. Archa, K. Achuthan, Trace and track: enhanced pharma supply chain infrastructure to prevent fraud, in *Ubiquitous Communications and Network Computing. UBICNET 2017. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, ed by N. Kumar, A. Thakre, vol 218 (Springer, Cham, 2018) https://doi.org/10.1007/978-3-319-73423-1_17
14. M.G. Shashank, V. Sangeetha, H. Shilpa, An exploratory study on self-sovereign identity powered by the blockchain technology. No. 5484. EasyChair (2021)
15. M. Kuperberg, S. Kemper, C. Durak, Blockchain usage for government-issued electronic IDs: a survey, in *International Conference on Advanced Information Systems Engineering* (Springer, Cham, 2019)
16. Dock. ”About.” Dock. https://www.dock.io/about. Last Accessed: 30 Jan 2021
17. T. Ruff, Meet connect. me, the first sovrin-based digital wallet. Evernym, 21 June 2019. https://www.evernym.com/blog/connect-me-sovrin-digital-wallet
18. A. Andrade-Walz, Developer documentation—Getting Started. Evernym, 18 Dec 2020. https://www.evernym.com/docs
19. Bloom, What is bloom? Bloom Knowledge Base, 8 Nov. 2019. https://faq.bloom.co/article/5-what-is-bloom
20. Decentralized-Identity/Ion. GitHub https://github.com/decentralized-identity/ion. Last Accessed: 26 Jan 2021
21. Verifiable Credentials Preview by Azure AD | Decentralized Identity Developer Docs. Microsoft. https://didproject.azurewebsites.net/docs/overview.html. Last Accessed: 8 Feb 2021
22. M. Sena (2020) Introducing, 3ID connect—3Box. Medium, 16 June 2020. https://medium.com/3box/introducing-3id-connect-531af4f84d3f
23. KILT Protocol. KILT Protocol–Claim Independence. KILT Protocol. https://www.kilt.io. Last Accessed: 8 Feb 2021
24. Foundation, S., Sovrin-protocol-and-token-white-paper. Sovrin. https://sovrin.org/wp-content/uploads/2018/03/Sovrin-Protocol-and-Token-White-Paper.pdf. Last Accessed: 27 Jan 2021
25. A.E. Panait, R.F. Olimid, A. Stefanescu, Analysis of uPort open, an identity management blockchain-based solution, in *Trust, Privacy and Security in Digital Business. TrustBus 2020*. ed by S. Gritzalis, E.R. Weippl, G. Kotsis, A.M. Tjoa, I. Khalil, Lecture Notes in Computer Science, vol. 12395 (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-58986-8_1
26. P. Kohlhaas, Zug ID: exploring the first publicly verified blockchain identity. Medium, 21 June 2018. https:/medium.com/uport/zug-id-exploring-the-first-publicly-verified-blockchain-identity-38bd0ee3702
27. Indy Use Cases—Hyperledger Indy—Hyperledger Confluence. Hyperledger. https://wiki.hyperledger.org/display/indy/Indy+Use+Cases. Last Accessed: 27 Jan 2021
28. Build Software Better, Together. GitHub, Rubix. https://github.com/rubixchain/rubixnetwork/blob/master/RubiX_WhitePaper.pdf. Last Accessed: 17 Feb 2021
29. ERC725Alliance. ERC725Alliance/ERC725. GitHub. https://github.com/ERC725Alliance/erc725/blob/master/docs/ERC-725.md. Last Accessed: 27 Jan 2021

30. J. Thorstensson, ERC780—an open identity and claims protocol for ethereum. Medium, 19 Oct. 2020, https://medium.com/uport/erc1056-erc780-an-open-identity-and-claims-protocol-for-ethereum-aef7207bc744
31. IPFS Documentation. IPFS Docs, 31 Aug. 2020. https://docs.ipfs.io

Transfiguring Handwritten Text and Typewritten Text

M. Keerthana, P. Hima Varshini, K. Sri Thanvi, G. Vijaya, and V. Deepa

Abstract Detecting and recognizing handwritten characters have become a challenging task nowadays. It comes under pattern recognition. The recognition of handwritten text automatically is widely used in many applications, where to process huge handwritten text. This application is useful for not only transforming handwritten text to editable text or typed text, but also for transfiguring the typewritten text to handwritten text. In transforming the typed text to handwritten text, also includes randomization of handwritten character styles to pragmatize the output. The output text file of both transfigurations is translated to the corresponding audio format for easy output validation by the user and automatically mailed to the user to avoid data loss. Here the whole input is taken by the user and driven via graphical user interface (GUI). This application reduces the time and manual work involved in processing and converting both handwritten and typewritten texts.

Keywords Transfiguring · Randomization · Pattern recognition · Optical character recognition (OCR) · Python pillow (PIL) · Tesseract · Validation

1 Introduction

Recognizing the handwritten characters either letters or digits is one of the effortful areas in pattern recognition. In this project, ‘Transfiguring Handwritten Text and Typewritten Text,’ we aim at building software that aids in analyzing the text or information in a sophisticated and well-defined way. The project recognizes any handwritten character efficiently on a computer with an input given as an image file. The recognized text is then converted into typewritten text through the algorithm used in the software, which is built on the foundation of the concept, optical character recognition (OCR). This project also recognizes the typed text and transforms it into user handwriting styles by randomizing the characters to make it realistic. The

M. Keerthana (✉) · P. Hima Varshini · K. Sri Thanvi · G. Vijaya · V. Deepa
Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

V. Deepa
e-mail: deepa@vrsiddhartha.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_49

output files obtained here are automatically mailed to the user based on the input prompted. Handwriting recognition systems are of two types: Online and Offline. In an online handwriting character recognition system, the handwritten characters of the user are recognized as the user is writing simultaneously. But in an offline handwriting recognition system, the handwriting of the user is available as an image accurate prediction on test data [1]. It is much more effortful to recognize the handwritten characters compared to printed ones. Using 'Transfiguring Handwritten Text and Typewritten Text' software, intelligible handwritten data is digitized and stored in computers, it saves the manual efforts and time to understand and type the text, power consumption techniques lead to a lot of pollution. With proper demand in mind, the agencies can limit themselves to produce only the required amount. Identifying demand also helps these agencies to provide quality power without swells and outages. Commonly, magnetic character recognition (MCR) or optical character recognition (OCR) is to recognize handwritten as well as printed data [2]. Our system comprises of two parts:

(1) GUI Application: This is the front end of the system that prompts the user to intake the file to be detected and transfigured and the mailing-related information. The file to be uploaded by the user is an image file in case of transforming the handwritten text to editable text and a text file in case of transfiguring typed to handwriting text.
(2) Backend of the application: The system backend contains the logic required for detecting and transforming. Tesseract OCR is used for recognizing handwritten or printed characters. In transfiguring typed text into the handwritten text, we input the text file, and the file is read character by character and after reading each character, and Python Imaging Library (PIL) is for pasting each handwritten character image on the plain background each time to make an output. Here we need to give the user the handwriting style images of each character at starting. To make the output realistic, we provide a few various handwriting styles of each character of the user and now the system chooses the handwriting style of an alphanumeric randomly automatic mailing of the output file is done to the user.

The remaining part of the paper is arranged in the following order: Related work is presented in Sect. 2. The detailed description of the proposed prototype and its functions is presented in Sect. 3. In Sect. 4, the result analysis of our proposed work is presented.

2 Related Work

Rohan Vaidya et al. [1] innovative method for offline handwritten character detection using deep neural networks. Image segmentation-based handwritten character recognition system is described by keeping in view of the amount of computational power needed to train a neural network has increased due to the availability of GPU's and another cloud-based service. The segmentation-based approach used here gives no guarantee on the optimum results when tested with different styles of writing. Jamshed Memon et al. [3] provided research directions on the papers published between the years 2000 to 2019 are provided in this by collecting, synthesizing, and analyzing research articles on the topic of handwritten OCR (and closely related topics). They almost chose 176 articles for this study and gave a comprehensive survey by summarizing all the techniques. This survey gave us an idea of exiting techniques used and considered their strengths and weaknesses.

Manoj Sonkusare et al. [4] gave a survey on handwritten character recognition methodologies for English characters. In their paper, they have discussed the global skew correction technique, which corrects the text line to be horizontally aligned while scanning the image. They had concluded stating that much more work has to be done in the area of HCR to build up a practical solution that can be accessible to everyone. They designed an easy-to-use graphical interface with a semi-automatic complete OCR workflow, from image optimization and also encompassed page layout analysis to automatic post-correction. Our application includes beyond these such as line finding algorithm and adaptive classifiers for more accurate results.

Tsehay Admassu Assegie and Pramod Sekharan Nair [5] by considering the digits, the main problems of handwritten digits recognition using machine learning approaches are listed. Here the standard Kaggle digits dataset is used for the recognition of handwritten digits using a decision tree classification approach. Here considering the result of one decision tree may not give optimum results but random forest is suggested on average. Senka Drobac et al. [6] observed that with deep neural networks and randomly sampled training data, it is possible to train one mixed model for the entire dataset that performs better than the monolingual models. Voting with five relatively similar models further reduces error rates and post-correction models correct punctuation, and it is a time-consuming process unlike the application proposed in this paper.

Anshul Gupta et al. [7] the proposed algorithm in order to recognize the offline handwritten English characters produces a satisfactory result with an accuracy rate of 86.46%. As scrutinize to other handwritten techniques, the proposed algorithm takes lesser training time. Mandal et al. [8] proposed algorithms for the line, words, and character segmentation. Each line is split and segmented into words by using a contour tracing algorithm from the extreme left of the paragraph. Words are segmented by considering white pixels in between words. The existing systems mostly has static classifiers, but Python-tesseract we used include adaptive classifiers.

3 Proposed Work

3.1 Creating GUI Interface for the Application:

This project has a graphical user interface to interact with the user and prompts the user for the input via its widgets. This user interface (UI) acts as a way of interaction between the user and the application. This project contains two different user interfaces for these two types of conversions those are converting handwritten or printed image text to editable text and converting typed text to handwritten style. To create an effective interaction between the application and the user, we created the face of the application simply called as user interface. Tkinter is used here to create interactive user interfaces and it is the GUI library in Python. Tkinter provides a fast and easy way for creating GUI applications. First, import the Tkinter module and then create the GUI application root window. Add one or more of the widgets to the GUI application based on your necessity. Generally, the widgets include buttons, labels, canvas, list box, etc. Here the widgets used are labels, buttons, and entry. There are almost 15 types of widgets in Tkinter. Each widget has some set of attributes width, height, bg, fg, active background, command, etc. For button widgets, we have to include 'command' attribute which defines the method to be called or action to be performed after user clicks the button, and then the backend algorithms or techniques or function used will be executed and outputs the result. At last, enter the main event loop to take action against each event triggered by the user. The interface of the first type of conversion has two buttons one for uploading the file and for downloading the output file. The label is used for displaying the steps to be followed. Entry widget is used for the taking mail id of the user in the text format and then another button for clicking to send the mail. In the second type of conversion also, the same set of widgets are used as above. In this way, the project used the Tkinter module and its widgets for prompting the user for the input and made it interactive. It is to summarize that in total for each type of conversion we used three buttons(Upload, download, Listening output, 2 labels one for display the steps to be done and other for a display image for making user interface attractive, one entry widget to take user's mail id and at last one button for sending the mail.

3.2 Transfiguring Handwritten Text to Typed Text

In the first transfiguration, i.e., to convert the handwritten text or printed image text to editable text, the implementation did here is by using Python tesseract OCR. tesseract OCR is a type of OCR engine that has matrix matching. Tesseract engine has flexibility and extensibility of machines and the fact that many communities are active researchers to develop this OCR engine. Tesseract architecture has various parts in order. They are line and word finding, word recognition, static character classifier, linguistic analysis, and adaptive classifiers and word finding uses line finding algorithm, and also it includes base line fitting, fixed pitch detection and chopping and proportional word finding. Word recognition involves chopping joined characters and associating broken characters. Python-tesseract is an optical character recognition (OCR) engine for python as it will recognize and 'read' the text embedded in images with various extensions such as jpeg, png, gif, tiff, and others. Here the user has to upload the file of image format either.jpeg or.jpg or.png as input. It is required to download the tesseract, and also it is ought to set the tesseract executable in the path. 'image_to_string' method is used here to extract the text from the image that contained handwriting or printed text, and this function returns the unmodified output as string from tesseract OCR processing. This output of the function is saved in the output file called 'myOutFile,' and it is of .txt format. Each time the application is run for conversion, the output file and created and the text is saved in the file (without having any output text of the previous run). In this type of conversion, we are identifying English alphabets. The output of OCR is the text, which is stored in a file (myoutFile.txt).

As shown in Fig. 1, user inputs the image file and then the file processing takes place for recognizing the handwritten text or printed text to typed text using above-mentioned methodology. We used OCR algorithm for getting the output, and in this output text file modification, deletion of text can be done, and it is converted to audio for user validation and then mailed to user.

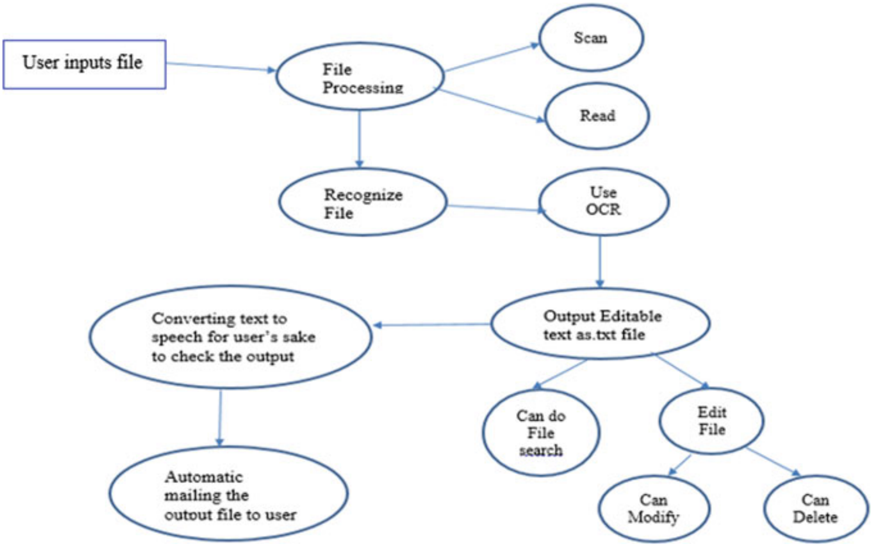

User inputs file
File Processing
Scan
Read
Recognize File
Use OCR
Output Editable text as.txt file
Converting text to speech for user's sake to check the output
Automatic mailing the outout file to user
Can do File search
Edit File
Can Modify
Can Delete

Fig. 1 Architecture for first type of conversion

3.3 Transforming Typed Text to Handwritten Text

In the case of the second transfiguration, i.e., converting the typed text to handwritten text, the implementation is achieved with the Python Imaging Library (PIL). The module named ‘Image’ is a part of the PIL that is used to manipulate the images in ways necessary to achieve the desired output. The methods that are used prominently are Image.open(), Image.paste() and Image.new(), and Image.save(), which allows the program to open images, paste images and do image manipulation and to create new images, and save the manipulated images, respectively. Since the entire mechanism of this function revolves around the idea of pasting respective handwritten symbols on a white background, these functions play a crucial role. When the function is made to run, the users are able to choose an input file, which is of .txt format. Once the file is uploaded, the control of the function is passed to the image_convertor.py. This function performs the role of parsing the input file and performing the actions of mapping each character to the respective but random handwritten characters. These images are opened using the open() method. Pasting the images of the chosen handwritten symbols in the order, they appear in the input text is achieved with the paste() method which includes an attribute, mask which defines how dark the image of the letters should be, hence defining the darkness of the handwritten symbols. As the background image is out of space, a new image is created with the letters continuing. This new image acts as a new paper, with the new() method. Then, all these pages which are images of a bunch of images pasted on them are saved with the save() method. The brilliance of the entire methodology is observed through the spacing and lining techniques and the randomization feature in the program, to perfectly space the symbols and evenly proceed to the next line in a page and also to proceed to the next page and to add a natural feel to the output. This is achieved by a systematic increase in the lining height and also a constantly added space. Once all the characters are replaced, all the pages are combined into a PDF document and are returned back to the main program. Images of each generated page are stored in the project file directory in addition to the generated PDF file, named final.pdf.

As shown in Fig. 2, user inputs the .txt file and then the file processing takes place for recognizing and converting the editable text to handwritten text using above-mentioned methodology. Some of the PIL methods used to obtain the output pdf file as mentioned, and it is then converted to audio for user validation and then mailed to user.

3.4 Randomizing Handwritten Character Styles

The generated output file that has handwriting styled text from the given typed text may vary a little bit from the text written by the human hand because if we consider ‘a’ alphabet, whenever the person writes the letter ‘a’ by the hand he/she might not

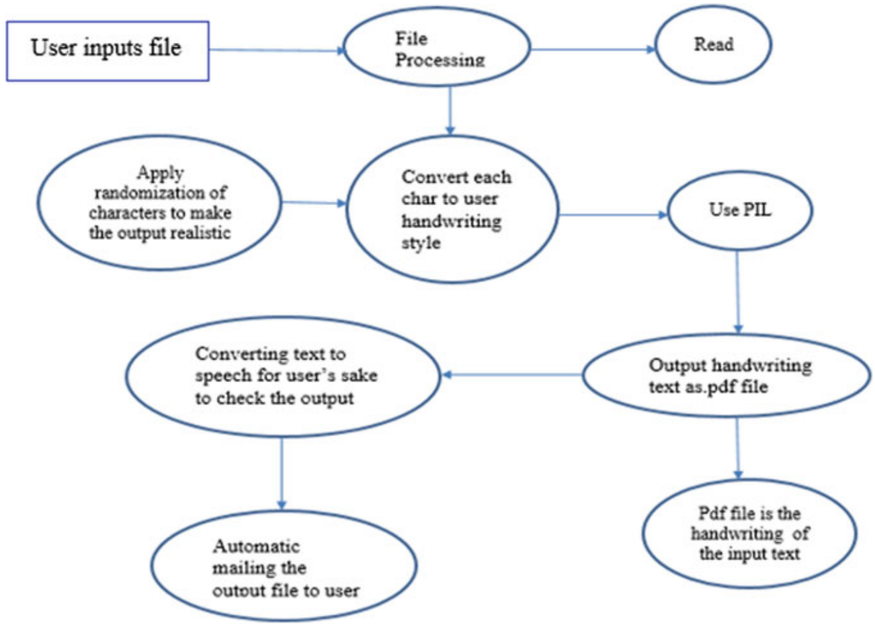

User inputs file
File Processing
Read
Apply randomization of characters to make the output realistic
Convert each char to user handwriting style
Use PIL
Converting text to speech for user's sake to check the output
Output handwriting text as.pdf file
Automatic mailing the outout file to user
Pdf file is the handwriting of the input text

Fig. 2 Architecture for second type of conversion

write the same style of 'a' every time. If the application takes only one handwriting style for each character, the output will be unrealistic when compared to written text by human hand. So, this project resolves it by including the randomization of characters; i.e., this application takes 3 to 4 handwriting styles of each alphanumeric character, and while converting to handwritten text, it picks one of the handwriting styles of the character randomly whenever it encounters that particular character. Let us take the text input as 'veins are blood vessels.' The output is as shown below in Fig. 3. Consider this as the first version of that sentence that is converted.

When this sentence is written by a human hand (not by the application) by about three times, the handwriting style of some characters may vary slightly. Even this project done the same to make the output handwritten style more realistic. See Fig. 4.

In the above output text, from Fig. 4, the letters 'v' and 's' in veins and vessels are a bit distinguished when observed. Considering both the versions of the same sentence, there are noticeable differences. These noticeable differences were implemented intentionally. Even a human cannot write the same 'v' or 's' each time. So, this project is able to make the output pragmatic.

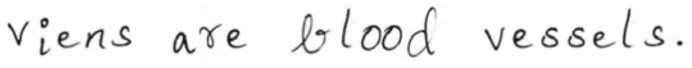
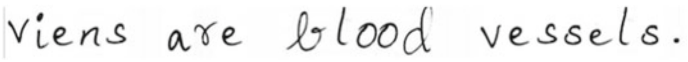


Fig. 3 Version-1 of given sentence



Viens are blood vessels.

Fig. 4 Version-2 of given sentence

3.5 .Txt Output File to .Mp3 File

Any application cannot be fully accurate. So, the user intervention is needed sometimes to validate the output. So, to make this job of the user easy, this project is capable of converting the output editable text obtained in the first type of conversion (from handwritten to typed text) to speech format to check that the obtained output text is recognized and extracted correctly as per the given input image. For this conversion, it used the python gTTS module. gTTS is a tool that converts the text entered, into audio which can be saved as an mp3 file the gTTS API provides the facility to convert text files into different languages. By listening to the audio file, the user can validate the output text with much fewer efforts.

3.6 Mailing the Output File to the User to Avoid Data Loss

This application adds an extra feature to mail the extracted output file to the user. Here the application takes the mail id of the user as an input, and it will send a mail to the user. In case of data loss in the local system or any other type of disaster or system failure where there is a chance of losing the file saved, so to avoid this destruction mailing of the output file is required. MIME Multipart is used for this implementation. Multipurpose Internet Mail Extensions (MIME) is an Internet standard used to support the transfer of single or multiple texts and also non-text attachments. SMTP library in python is used. The smtplib module in python is used for starting a client session object that can be used to send data on the Internet. It is required to give the sender and receiver's mail addresses for sending mail with the attachment. The attachment here is the output file together with the .mp3 file.

3.7 Dataset

In the conversion of the typed text to handwritten text, the application needs the handwriting styles of the user in form of images. The application needs at least one handwriting style of each English character and also for digits and special symbols. This is done manually by the user. If once the user gave all of his/her handwriting styles for every character, it is enough, and the user can run any number of inputs to be converted. It is also important that the names of these handwriting style images need to be given in a specific format. Let us take for the character 'a' the name of the handwriting style image should be 'a' and for 'A' the name should be 'aupper.' In this way, the images are used to get the output as handwritten text.

4 Results Analysis

The functions are presented to the users through an interactive and attractive user interface, one for each module. The output and end results could be observed as below. Initially, when the first module is made to run, a GUI as shown in Fig. 6 appears.

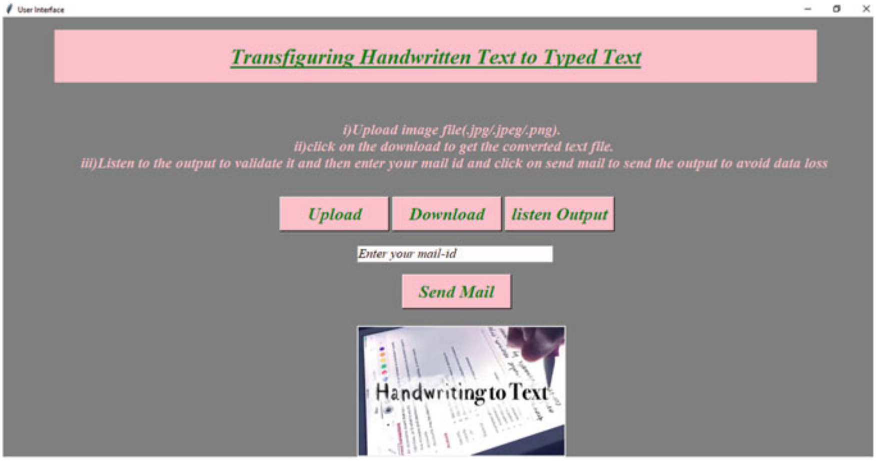
As the steps in the user interface suggest, the user is prompted to upload a file, and to do so, the user clicks on the upload button from the reference, Fig. 5. Once the file is uploaded, the user could go on and click on the download button as shown in Fig. 5.

In Fig. 6, another example, where a handwritten sentence is scanned and uploaded as an input to the software is recognized, and an equivalent text document is generated.

So, we tested it out for several formats including the typed text images. In Fig. 7, it could be seen that the given input is an image of a typed text, and the generated text file has the greatest accuracy, compared to the other inputs.

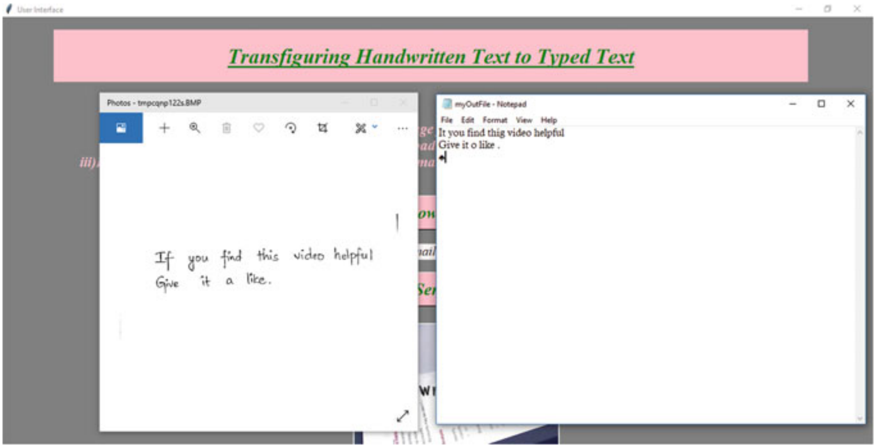
We have uploaded a scanned image of an Aadhar card prototype, and in the result window in Fig. 8, the text portion of the image is recognized and is generated in the resulting file. However, the images were ignored.

We have extended the use cases to a number plate as well. In Fig. 9, it could be seen that an image of a car number plate is uploaded, and the software successfully recognized the text and was able to generate a typed text of the same number plate as the resulting file.



i)Upload image file(.jpg/.jpeg/.png).
ii)click on the download to get the converted text file.
iii)Listen to the output to validate it and then enter your mail id and click on send mail to send the output to avoid data loss

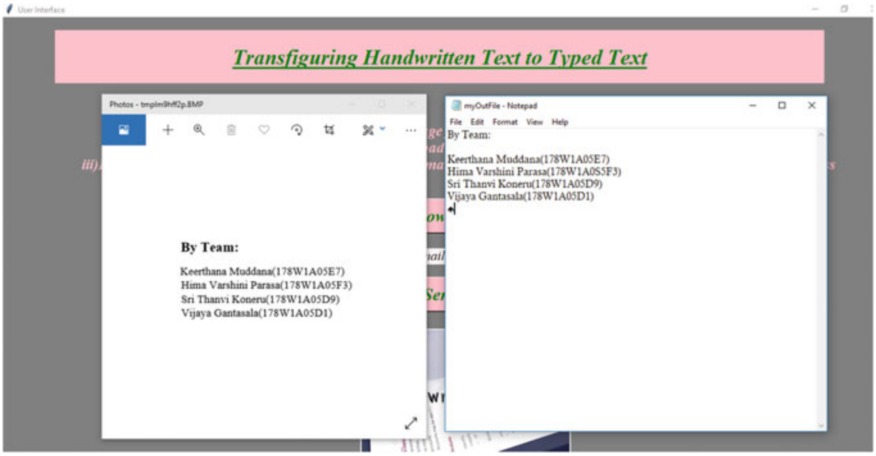
Fig. 5 UI of first transfiguration


User Interface
Transfiguring Handwritten Text to Typed Text
Photos - tmpcqnp122s.BMP
myOutFile - Notepad
File Edit Format View Help

If you find this video helpful
Give it a like.

```
It you find thig video helpful
Give it o like .
```

Fig. 6 Output of sample handwritten input 2


User Interface
Transfiguring Handwritten Text to Typed Text
Photos - tmplm9hff2p.BMP
myOutFile - Notepad
File Edit Format View Help

By Team:

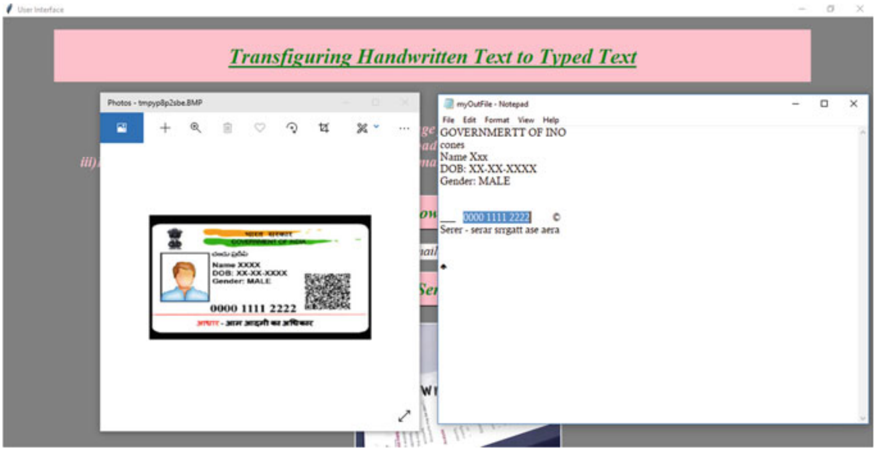
Keerthana Muddana(178W1A05E7)
Hima Varshini Parasa(178W1A05F3)
Sri Thanvi Koneru(178W1A05D9)
Vijaya Gantasala(178W1A05D1)

```
By Team:

Keerthana Muddana(178W1A05E7)
Hima Varshini Parasa(178W1A0S5F3)
Sri Thanvi Koneru(178W1A05D9)
Vijaya Gantasala(178W1A05D1)
```

Fig. 7 Output of input containing digits and characters

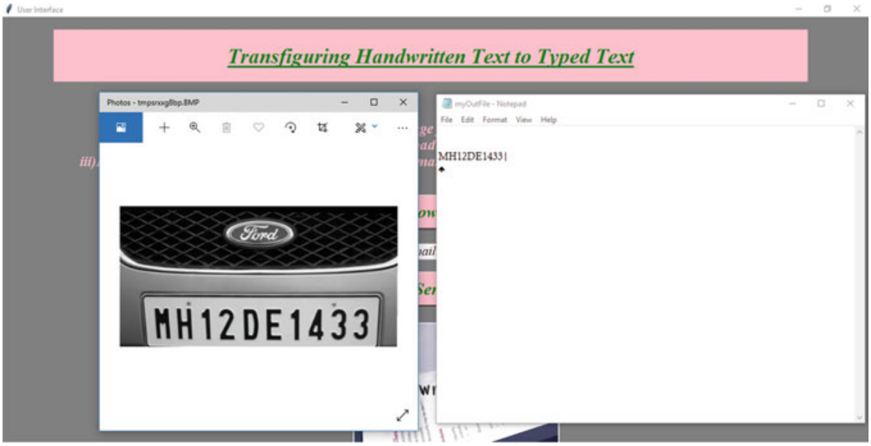
The final example that we have considered providing the system as an input is a postal card as seen in Fig. 10. The resulting file successfully produced the text from the postal card image through the entire text is in different fonts and styles.

In the second module, as the title from the user interface window from Fig. 11 suggests, it serves the purpose of Transfiguring Typewritten Text to Handwritten Text. Figure 11 represents the user interface screen. As mentioned in the first step, from Fig. 11, the user then clicks on the 'Select a File' button to upload the compatible file as the input. Then the 'Upload a file' label present on the user interface screen changes to 'File uploaded,' informing the user that necessary actions have taken place.



```
GOVERNMERTT OF INO
cones
Name Xxx
DOB: XX-XX-XXXX
Gender: MALE


___  0000 1111 2222      ©
Serer - serar srrgatt ase aera
```

Fig. 8 Recognizing and extracting Aadhar card details

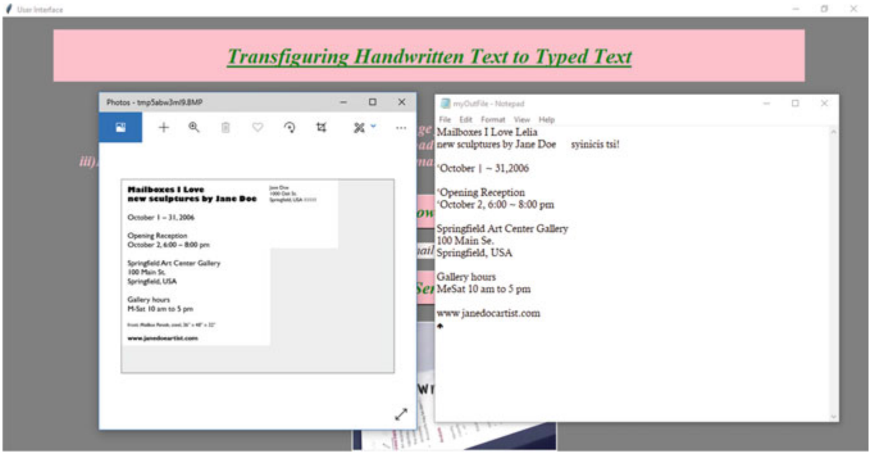

User Interface
Transfiguring Handwritten Text to Typed Text
Photos - tmpsnxg8bp.BMP
myOutFile - Notepad
File Edit Format View Help
Ford
MH12DE1433

```
MH12DE1433|
```

Fig. 9 Recognizing number plate information

Once the file is uploaded, the entire software mechanism is made to run when the user clicks on the ‘Convert to PDF’ button. The entire functioning is performed on the input file and the generated output file, i.e., a PDF of the handwritten style as shown in Fig. 12 is generated and uploaded in the same directory of the program directory. The generated file opens up in the default browser on the user’s local machine. This output could be evaluated by the user through the ‘Listen File’ button.

In Fig. 12, the text file that is uploaded consists of a single line and the generated PDF document consists of a single page with that single line transfigured into handwritten text. However, in Fig. 13, the text document itself consists of several lines.



Transfiguring Handwritten Text to Typed Text

Photos - tmp5abw3ml9.BMP

Mailboxes I Love
new sculptures by Jane Doe

October 1 – 31, 2006

Opening Reception
October 2, 6:00 – 8:00 pm

Springfield Art Center Gallery
100 Main St.
Springfield, USA

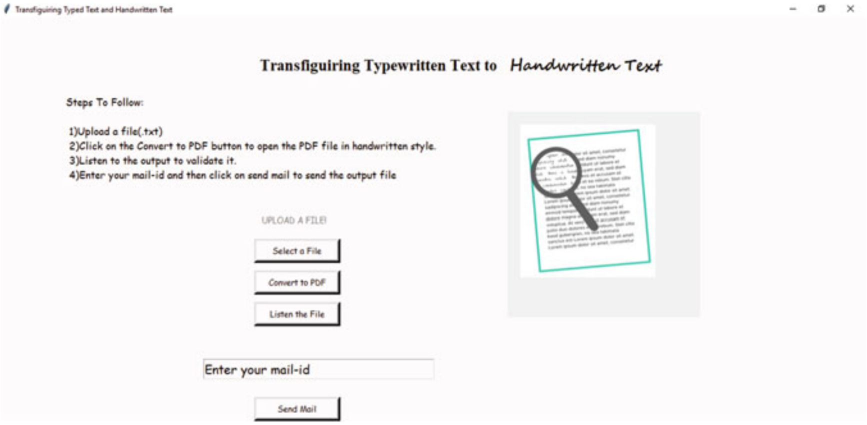
Gallery hours
M-Sat 10 am to 5 pm

www.janedoeartist.com

myOutFile - Notepad

```
Mailboxes I Love Lelia
new sculptures by Jane Doe     syinicis tsi!

'October | ~ 31,2006

'Opening Reception
'October 2, 6:00 ~ 8:00 pm

Springfield Art Center Gallery
100 Main Se.
Springfield, USA

Gallery hours
MeSat 10 am to 5 pm

www janedocartist.com
```

Fig. 10 Extracting postal card details



Transfiguiring Typed Text and Handwritten Text

Transfiguiring Typewritten Text to Handwritten Text

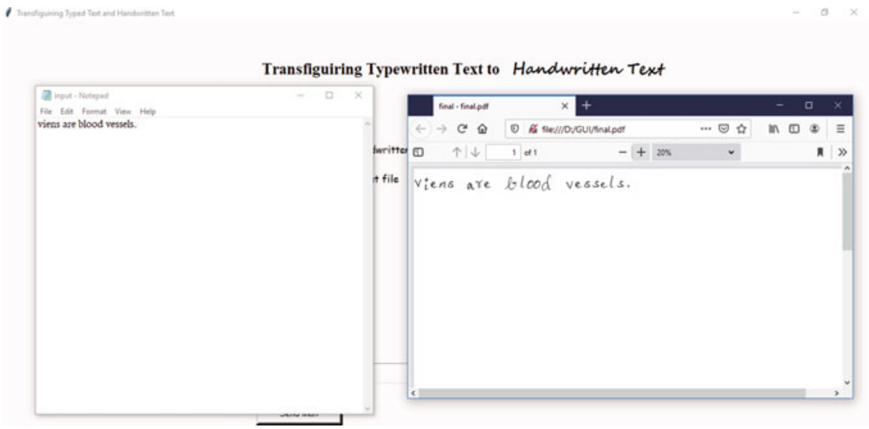
Steps To Follow:

1)Upload a file(.txt)
2)Click on the Convert to PDF button to open the PDF file in handwritten style.
3)Listen to the output to validate it.
4)Enter your mail-id and then click on send mail to send the output file

Fig. 11 UI for second type of conversion (Typed text to handwriting)

Since the software is storing the generated outputs in the project file directory on the local machine, we have extended our project to send the output file to the respective user. As a result, the generated outputs are safe from deletion and any future hardware failures. The results of this implementation could be observed in the following images. In Fig. 14, the mailing of the output file that has the converted handwritten text to typed text could be seen sent from the default email address to the email address given in the user interface.

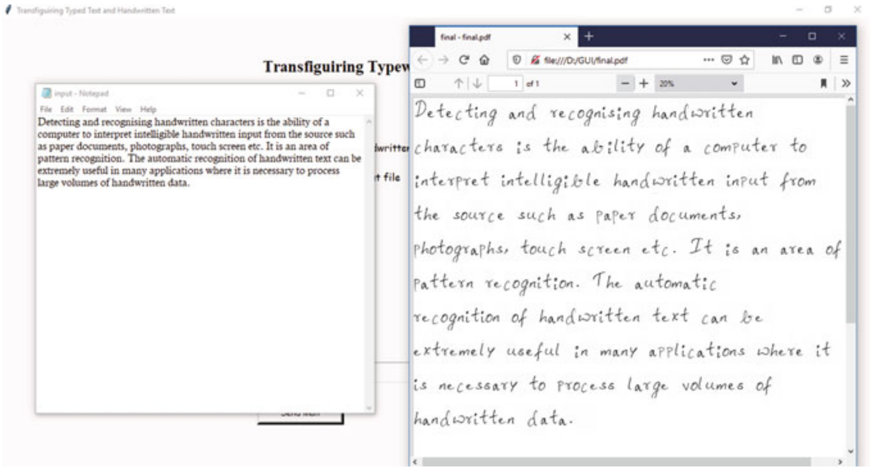
The attached file is the resultant file from the first module, i.e., the generated text document, and when this document is opened from the email that is received at the user's end, the screen in Fig. 15 appears, that is the opened text file from the received email.


Transfiguiring Typed Text and Handwritten Text
Transfiguiring Typewritten Text to Handwritten Text
input - Notepad
File Edit Format View Help
final - final.pdf
file:///D:/GUI/final.pdf
1 of 1
20%
dwritter
t file

viens are blood vessels.

viens are blood vessels.

Fig. 12 Converting given editable text to user's handwriting style



Detecting and recognising handwritten characters is the ability of a
computer to interpret intelligible handwritten input from the source such
as paper documents, photographs, touch screen etc. It is an area of
pattern recognition. The automatic recognition of handwritten text can be
extremely useful in many applications where it is necessary to process
large volumes of handwritten data.

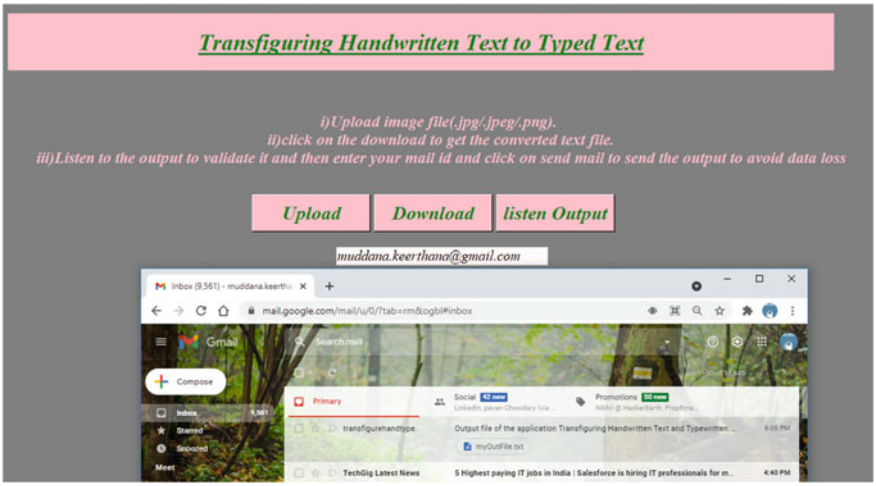
Detecting and recognising handwritten
characters is the ability of a computer to
interpret intelligible handwritten input from
the source such as paper documents,
photographs, touch screen etc. It is an area of
pattern recognition. The automatic
recognition of handwritten text can be
extremely useful in many applications where it
is necessary to process large volumes of
handwritten data.

Fig. 13 Transforming typed text to handwriting style

When the email is opened, the expected screen is represented in Fig. 16, which is the view of the mail sent with the attachment 'final.pdf.'

5 Conclusion

The proposed system has come up with accurate and efficient approaches to transfigure handwritten text and typewritten text with easy-to-use and attractive graphical user interfaces for each module for effective user interaction. This application also



Transfiguring Handwritten Text to Typed Text

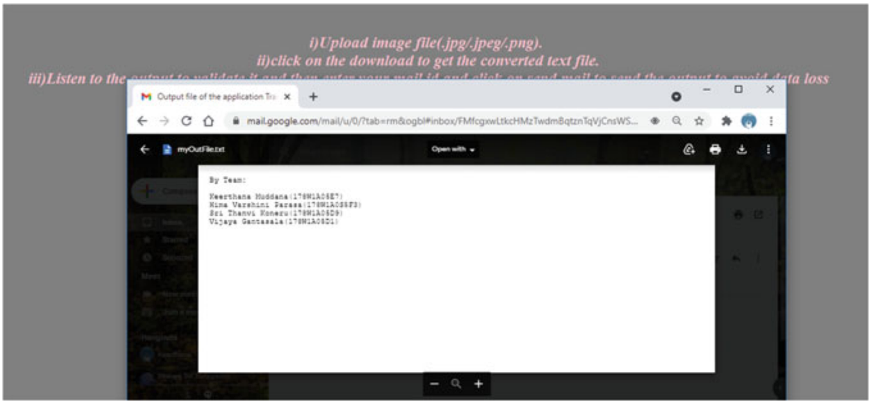
i)Upload image file(.jpg/.jpeg/.png).
ii)click on the download to get the converted text file.
iii)Listen to the output to validate it and then enter your mail id and click on send mail to send the output to avoid data loss

muddana.keerthana@gmail.com

transfigurehandtype. Output file of the application Transfiguring Handwritten Text and Typewritten ... 6:06 PM
myOutFile.txt

TechGig Latest News 5 Highest paying IT jobs in India | Salesforce is hiring IT professionals for m... 4:40 PM

Fig. 14 Mailing the output file that has the converted handwritten text to typed text

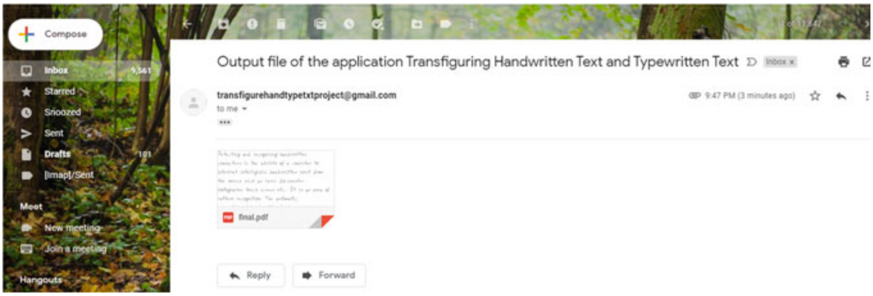


i)Upload image file(.jpg/.jpeg/.png).
ii)click on the download to get the converted text file.
iii)Listen to the

Output file of the application Tra
mail.google.com/mail/u/0/?tab=rm&ogbl#inbox/FMfcgxwLtkcHMzTwdmBqtznTqVjCnsWS...
myOutFile.txt
Open with

```
By Team:

Keerthana Muddana(178W1A05E7)
Hima Varshini Parasa(178W1A05F3)
Sri Thanvi Koneru(178W1A05D9)
Vijaya Gantasala(178W1A05D1)
```

Fig. 15 Output file containing typed text sent via mail



Output file of the application Transfiguring Handwritten Text and Typewritten Text

transfigurehandtypetxtproject@gmail.com
to me
9:47 PM (3 minutes ago)

final.pdf

Fig. 16 View of the mail sent with the attachment

converted the output file to corresponding speech format for easy validation of the converted text, and the output file is mailed to the user to avoid data loss in the future. The results obtained here show the reduction in manpower and time. This paper is only limited in converting English text, and in future, this is extended to implementing and upgrading the software built to recognize and interpret languages other than English, preferably Indian languages and also converting them to corresponding language's audio format.

References

1. Rohan Vaidya,Darshan Trivedi and Sagar , "Handwritten character recognition using deep learning",IEEE Xplore,2018 Compliant - Part Number: CFP18BAC-ART.
2. Suman Avdesh Yadav, Smitha Sharma and Shipra Ravi Kumar, "A Robust Approach For Offline English Character Recognition", International conference on futuristic trend in computational analysis and knowledge management (ABLAZE 2015), 2015.
3. Jamshed Memom, Maira Sami, Rizwan Ahmed Khan and Mueen Uddin, "Handwritten Optical Character Recognition(OCR)", IEEE Access, Digital Object Identifier, https://doi.org/10.1109/ACCESS.2020.3012542.
4. Manoj Sonkusare and Narendra Sahu, "A Survey On Handwritten Character Recognition (Hcr) Techniques For English Alphabets", Advances in Vision Computing: An International Journal (AVC), vol. 3,no.1 , March 2016.
5. Tsehay Admassu Assegie and Pramod Sekharan Nair, Handwritten digits recognition with decision tree classification: a machine learning approach. International Journal of Electrical and Computer Engineering (IJECE) **9**(5), 4446–4451 (2019)
6. Senka Drobac1 and Krister Linden," Optical character recognition with neural networks and post-correction with finite state methods", International Journal on Document Analysis and Recognition (IJDAR) (2020) 23:279–295.
7. Anshul Gupta; Manisha Srivastava and Chitralekha Mahanta," Offline handwritten character recognition using neural network", IEEE Xplore , 01 March 2012. CD**:** 978–1–4577–2057–4.
8. R. Mandal, N. Manna, Handwritten english character segmentation by baseline pixel burst method (BPBM). Adv. Model. Anal. B **57**(1), 31–46 (2014)
9. Shrinivas R. Zanwar, Ulhas B. Shinde, Abhilasha S. Narote and Sandipann P. Narote, "A Comprehensive survey on soft computing based OCR techniques", International Journal Of Scientific & Technology Research(IJSTR)-Volume 8, ISSUE 12, December 2019,ISSN 2277–8616.
10. J. Pradeep, E.Srinivasan and S.Himavathi, "Diagonal Based Feature Extraction For Handwritten Alphabets Recognition System Using Neural Network",International Journal of Computer Science & Information Technology (IJCSIT), vol. 3, no. 3,Feb 2011.
11. Javed, S.T., Hussain, S.: Segmentation based urdu nastalique OCR. In: Iberoamerican Congress on Pattern Recognition. Springer (2013).
12. Dhananjaya M S, Sushma R and Niranjana Krupa B, " Kannada Text to Speech Conversion: A Novel Approach", International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2016 IEEE.
13. Hsin-Chia Fu and Yeong Yuh Xu, "Multilinguistic Handwritten Character Recognitionby Bayesian Decision-Based Neural Networks", IEEE Transactions on Signal Processing, Volume 46, NO. 10, October 1998.
14. Mudunuri Prashanth Varma, Shubhro Jyoti Hore, Uday.C, S.Omnath Reddy and Vinay Jha Pillai, "Optimized handwritten character recognition using ANN)", International Journal Of Scientific & Technology Research(IJSTR)-Volume 9, ISSUE 01, January 2020,ISSN 2277–8616.

15. P. Banumathi , Dr. G. M. Nasira .”Handwritten Tamil Character Recognition using Artificial Neural Networks” in 2011 International Conference on Process Automation, Control and Computing

Smart Home with Condition Monitoring

P. Narendran, Vyshnavi Reddy, Sreelekshmi Saju, L. U. Suriya, and V. Ravi Kumar Pandi

Abstract This article presents the optimal idea of Smart homes with enhanced condition monitoring systems to establish simpler communication, domestic tasks, and excellent security for elderly population. The proposed smart home with condition monitoring capabilities has been developed by using a single Arduino controller using Proteus platform for monitoring and controlling many interconnected appliances such as lights, temperature sensors, breath-rate, and motion detectors. The proposed model is further enhanced by integrating a smartphone application, which displays the detected values from various sensors in the mobile phone screen. This research work has proposed a system for the smart home automation by using Arduino with condition monitoring, and it is done by integrating motion sensors, and also, a smartphone application has been developed.

Index Terms Smart home automation · Condition monitoring · Arduino · Proteus · PIR sensor · Temperature sensor · Heart rate detecting sensor · Ultrasonic sensor smartphone android application

1 Introduction

Over the last two decades, most of the countries' life expectancy has grown considerably. This expectancy is accompanied by advances in medical technology and diagnostic technologies, as well as a greater awareness for personal, environmental, hygiene, and nutrition. However, increased life expectancy and declining birth rates are expected to lead to a larger aging population in the near future. With the advances in medical system helping all the people to live longer lives, the number of senior citizens is expected to double by 2050 and will be need of radical societal change, stats from the report released by the World Health Organization for the international day

P. Narendran (✉) · V. Reddy · S. Saju · L. U. Suriya · V. Ravi Kumar Pandi
Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India

V. Ravi Kumar Pandi
e-mail: ravikumarpandiv@am.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_50

of older persons(1 October). The main theme for 2016's International Day of Older Persons was "Take a Stand Against Ageism." Research shows that concentrating more positively about aging can increase life expectancy by 7.5 years. [1].

Adults prefer to live independently and control themselves in their own homes, which promotes feelings of ability and that leads them to leave their parents alone or need a medical person or a caretaker to look after their parents on regular basis [2]. In economical view, the expenses of living at house with medical monitoring system, and intelligence equipment is overpriced but even more beneficial than joining medical centers. This problem leads to the solution of implementing smart home systems with condition monitoring system, and healthcare team will reduce the cost of personalized assistance and elders expectancy at home. This smart home with condition monitoring system is very inexpensive and can give the users the ability to control any electronic device without motion work in the way the user to control all the electronic devices using the smartphone application.

The primary objective is to enable the senior citizens to receive uninterrupted healthcare services while residing in their comfortable home environment. The proposed research work mainly includes a breathing meter to ensure the proper breathing rate, sleep hours, diet intake, water supplement, and also an oxygen meter is included to check the level of oxygen content in the blood under software-based system. This research work also includes a mobile application-based prototype and smart home with condition monitoring system and sensors [3]. The advent of smart home with condition monitoring proposes an improved quality and healthy living by leveraging healthcare support services for elderly and disabled population [4].

Applications of Arduino, machine learning, artificial intelligence, IoT, and raspberry pi are not restricted to perform specific tasks, those are playing significant roles in many of the fields and also used in all places where automation is introduced like education, industries, transportation, and even in finance [5, 6]. Tracking and taking care of elder people in the smart home using condition monitoring is presented in various researches. In those ways, some of the related works are Authors in paper [7] detailed about the technologies to assist senior citizens to live well at home. This paper discussed the speedy population aging which is occurring worldwide, there is raising interest in "smart home" technologies that can assist pensioner to continue living at home with safety and self-rule. Authors in paper [8] explain about the Design and Implementation of a Solar Integration in Electric Wheelchair. They have conducted an extensive search on relevant databases. Authors in this paper [9] have addressed the smart house system for the golden-ager and the disabled. This has proposed the solution can provide an improved and better approach to healthcare management. Authors of paper [10] have presented the home computerization for the elderly and disabled with Internet of Things. In terms of elder care, privacy and security are both concerned with safety [11–13].

2 Problem Formulation

The problems of the elderly. Though aging is the natural stage of human life, it brings with it innumerable problems for the people who have grown old. On detailed analysis about major problems facing by elderly people are many like economic problems, psychological problems, frequent health issues, and household problems. In this, major issue is health problem [14]. The National council on aging states, about 92 percent of old-timer have at least one chronic disease and 77 percent are at least with two. Heart disease, stroke, cancer, and diabetes are among the most common and costly chronic health conditions causing two-thirds of deaths each year [15]. The National center for chronic disease prevention and health promotion recommends meeting physicians for an annual checkup, maintaining healthy diet and keeping physical fitness, routine to help, manage or prevent chronic diseases. So to prevent all this, this research work has integrated with different sensing methods and enabled it to make ease the life of older people at their home step. So this Smart home with condition monitoring technology for aging people, have the context in their house with smart devices in a huge number is best to our future way of life [16, 17].

3 Proposed System

In an existing smart home, either the home is automated for reducing the work or it is done for only security reasons, which only makes the work easier and secure. The complete smart home had to take care of all the features necessary in the home to be automated. In medical cases, either the doctor or the nurse had to come home for a medical check in another case patient have to move to hospitals for check-ups. In routine cases, it is a huge task for the working children to take care of their old parents. Our system consists of a condition monitoring system with home automation features specially made for senior citizens and home quarantined people to interact with the physician via smartphone android application. This system has all the home automation features and continuously monitors people's health and collect, store and carry the data, and also be a medium between the doctors and patients, and further, it will be sent to their consultant doctor through an Android mobile application.

Figure 1 shows the exact idea in flowchart format of this proposed system that the data will be collected from all the medical sensors, and cameras then stores in the cloud. Those data will be used to transfer and access by the care-takers in case of emergency period for medical history factors and also for future use.

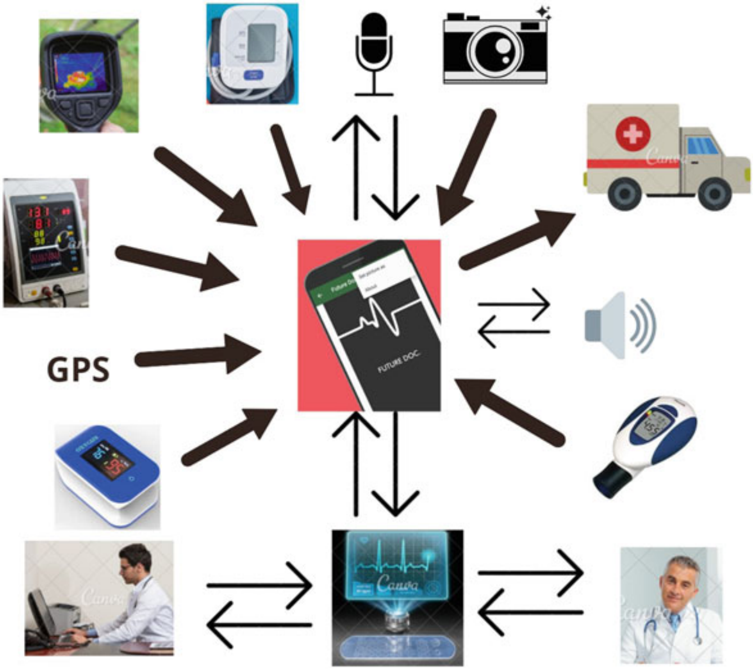


Fig. 1 Flowchart of SHCM controller

3.1 Home Automation

Figure 2 gives the complete knowledge on smart home which allows to tap into high-tech functionality. As technology development continues to expand, as well as the possibilities for consumer to make their life easier, enjoyable, and being able to keep all the technology connected through one interface is a massive step toward technology and home management. The smart home automation for senior citizens consists of automatic light, automatic water-tank filling, emergency button, and doorbell using the Arduino, such that it would be better for the senior citizens and in making their surroundings flexible. Depending on how the elders use of smart home technology, it leads to make more space and energy-efficient home automation system. In the way, automatic lights are controlled by motion detecting sensors in this not only lights, also includes fans and some electronic gadgets, automatic water-tank filling using ultrasonic sensor, emergency call by button system, and light doorbell in contact of blinking led lights. The most important in this fast world TIME can be precisely saved with all those smart automation.

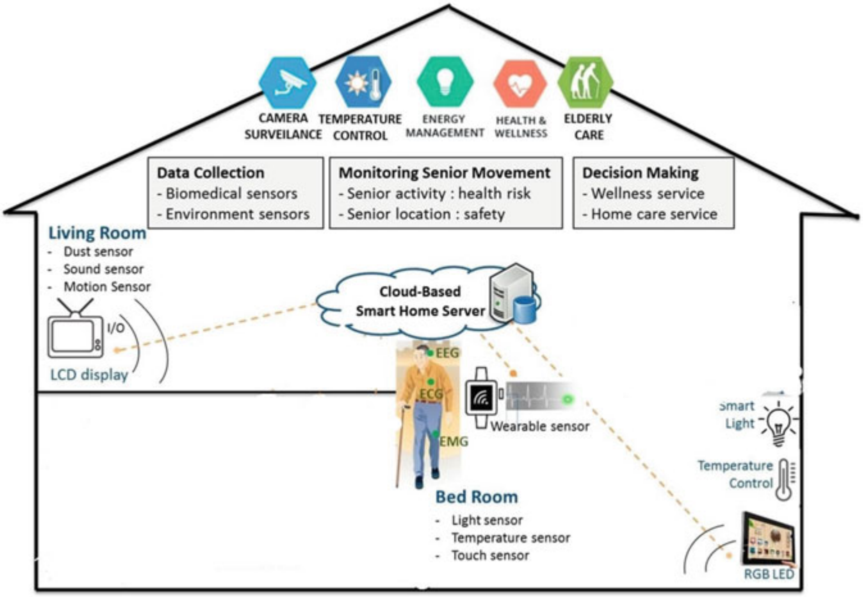


Fig. 2 Model of home automation system

3.2 *Condition Monitoring*

The health monitoring system where the person's health condition is detected and helps them to reminder of their intake medicine, diet, water supplements and regulates the sleep hours. As part of this, created an app where the individual person manually enters the parameters like time and date of intakes of medicine food, drinking, and sleeping hours, and these details can be monitor by doctor which is featured in Fig. 3. If case of emergency, doctor can contact the patient with that the vital first aid will be given as early as possible.

3.3 *Android Application*

The idea of home automation derives from the ideal of doing aspects inside a home to be easily controllable and manageable as well as they can automatically work as the resident's demand. So, this research work presents the design of mobile application as a user interface that provides comfort for the patients or quarantine people. And also is to be able to control the home appliance remotely. This application has featured that the patients and doctors can interact between each other under Android studio platform which also includes BMI calculator, sleep hours monitors, medicine remembrance, etc. This smartphone android application contains the detailed information about the patients and doctors in respective sights.

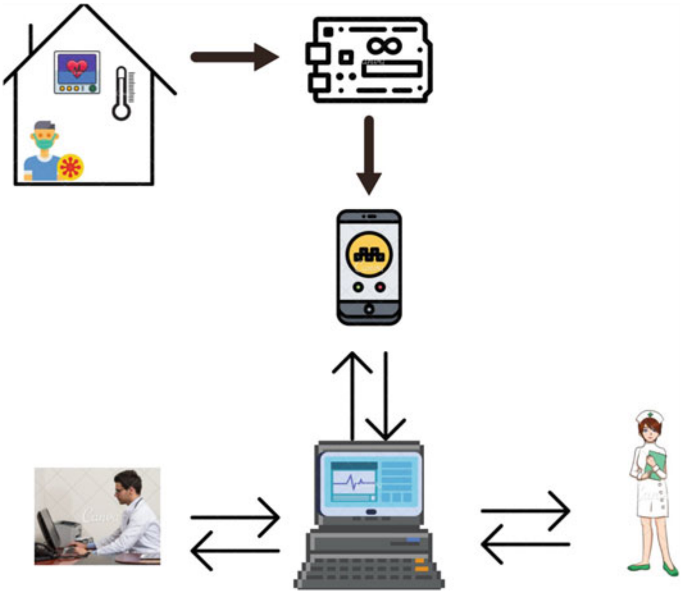


Fig. 3 Flow diagram of condition monitoring system

4 Result

This research work was to set one's site on house with automation and health monitoring using smart phone. This has been accomplished by building implanted system which depends on sensors using Proteus simulator to transmit the reading of important signs to smartphone via android application and stores in the cloud. This research work clearly suggests that older people or their care takers also include physicians to have positive attitude toward smart home technology devices. This smart home consists of the same technology and equipment as home automation for security, entertainment and energy conservation but tailors it toward older adults, people with disabilities and quarantine people. Every smart home varies with technologies but unites with Android phone and here we use the smartphone applications.

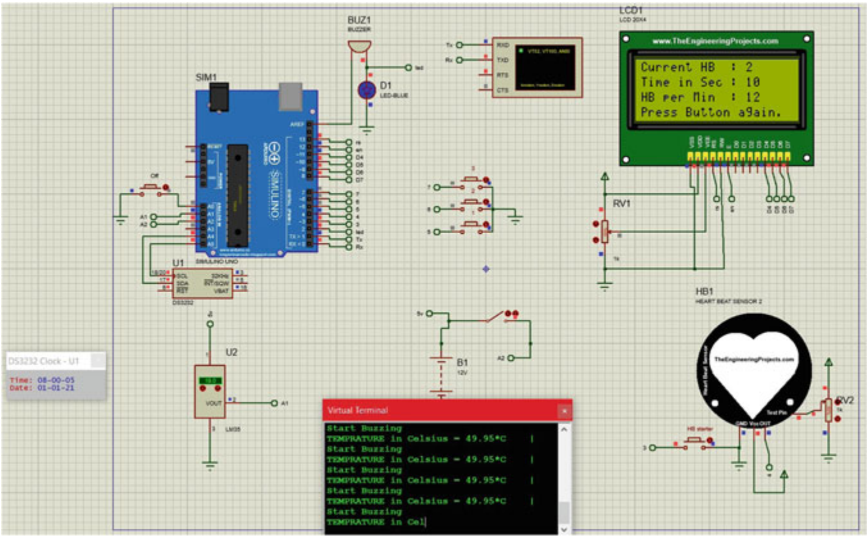
(a) *Sensor Validation*: In Fig. 11, it is found that all the sensors of monitoring medic kit are working, and specific results are displayed in LCD display and virtual display. This kit consists of temperature sensor, heart beat sensor, and medicine reminder kit, where the LM35 sensor compute integrated circuit temperature with an output voltage proportional to the Centigrade temperature, heart beat reads by the change the value of heart beat rate from the variable resistor connected with heart beat sensor in the way, when we press the button attached to the circuit, Arduino starts counting the heart beat ratings and displays on the LCD. It will start counting the heart beat rate as well as count the time in seconds. After 10 s, it will be multiplied by six with the current heart rate and will give the heart beat per minute, medicine reminder kit works is powered using 5 V supply. RTC is used as a clock. This system is simulated only for a single medic kit. The time slot is convertible in the program and can be triggered accordingly. Here it is fixed in three durations (i.e., 8am, 2 pm, and 8 pm.) We have divided time slots into three modes. Mode 1 selects to take medicine once/day at 8am when the user presses the first push button. Mode 2 twice/day at 8am and 8 pm when the user presses the second push button. Mode 3 thrice/day at 8am, 2 pm, and 8 pm if the user presses the third push button, then comes a button to snooze.

In the following figures, it is found that emergency circuit system, communication system, water-level indicator with automatic water-tank filler, smart door bell system, and PIR sensors. The validation of those system can be identified by the indication of LED and LCD display.

(b) *Medic Kit*: Fig. 4: Helps all the people to regularly check their health with their doctors from their home via smartphone.

(c) *Communication System (Emergency button and door)*:Fig. 5: Easy for communication purpose especially to the peoples who are physically challenged and paralyzed patients. It works in the way, Sw1 is the button outside the room, Sw2 is the communication button, and Sw3 is the emergency button and motor indicates the door. Here, Sw2 is in ON state the LED blink slowly, and Sw3 is in ON state the LED blinks faster and the motor rotates.

(d) *Water:* Fig. 6: Helps in water management and older people in filling their water tank automatically. Water-level indicator indicates the water level by



```
Current HB   : 2
Time in Sec  : 10
HB per Min   : 12
Press Button a9ain.
```

```
Time: 08-00-05
Date: 01-01-21
```

```
Start Buzzing
TEMPRATURE in Celsius = 49.95*C    |
Start Buzzing
TEMPRATURE in Celsius = 49.95*C    |
Start Buzzing
TEMPRATURE in Celsius = 49.95*C    |
Start Buzzing
TEMPRATURE in Celsius = 49.95*C    |
Start Buzzing
TEMPRATURE in Cel
```

Fig. 4 Medic kit

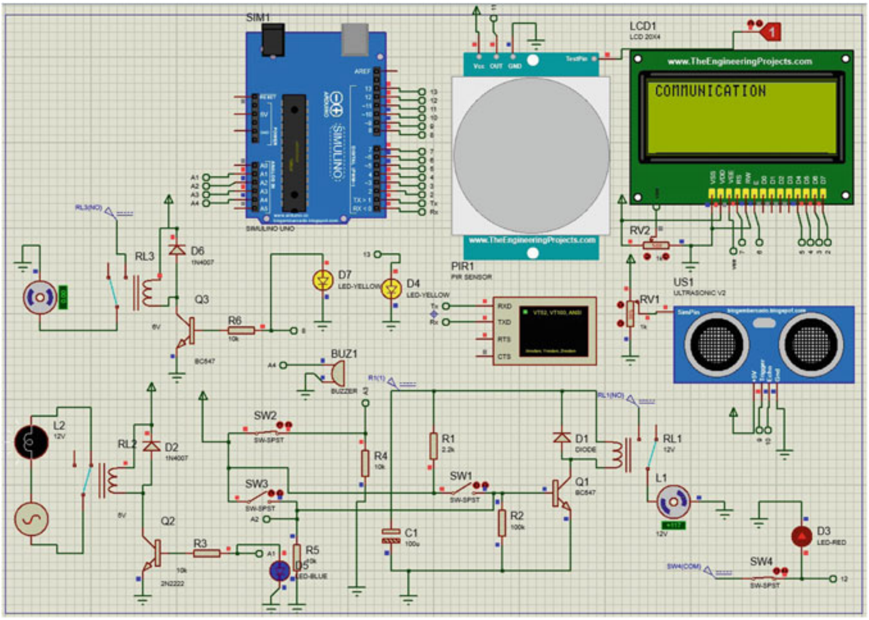


Fig. 5 Communication system

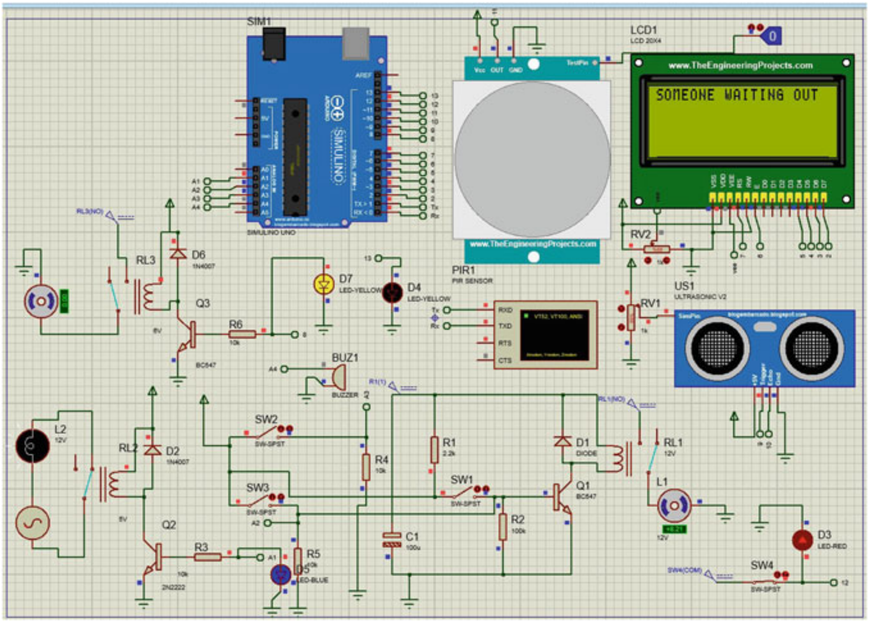


Fig. 6 Automatic water-tank and door system

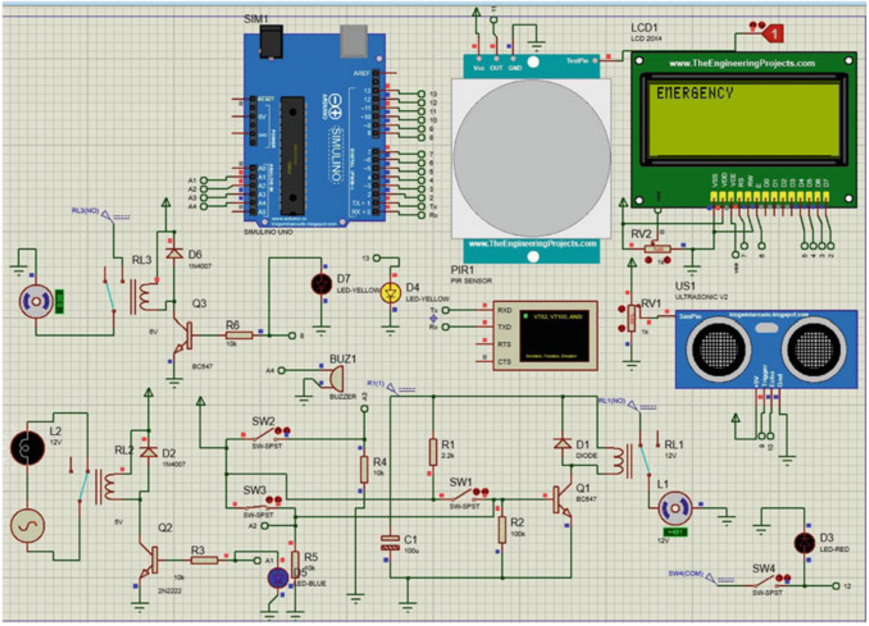


Fig. 7 Emergency button and light system

using ultrasonic sensors. The concept used in this water controller project where the water motor pump is automatically turned on when water level in the tank becomes low using potentiometer to control ultrasonic sensor when the distance is less indicating water tank is full and motor is OFF.

(e) *Light:* Fig. 7: Helps the special doorbell system with light connected to it for blinking when the bell rings. When the push button is pressed the led blinks indicating doorbell rings. The led pleased outside the door is used to indicate the doorbell is pressed when the buzzer is off and automatic light ON and OFF implemented through the application of PIR sensor which is detecting the motion and switch ON or OFF; i.e., we can say ALL IN ONE.

(f) *Android application:* This smartphone android application includes the features which have been done in Proteus simulation platform using Arduino software. The user interface of smart-mobile app development includes nine stages.

First stage includes the login page with patients and doctors name and includes related details. Figure 8 shows the opening page is the login page which has the “Sign Up” option. Once signed up with the credentials given by the user.

Figure 8 also indicates the main page that opens when the user login includes four selection tabs. Prior one is the doctor’s tab where patients can use this to find specialized doctors, second the patient’ tab where doctors can use it to find their patients list, third is dashboard and finally symptoms tab which helps users to know the symptoms of some common diseases.

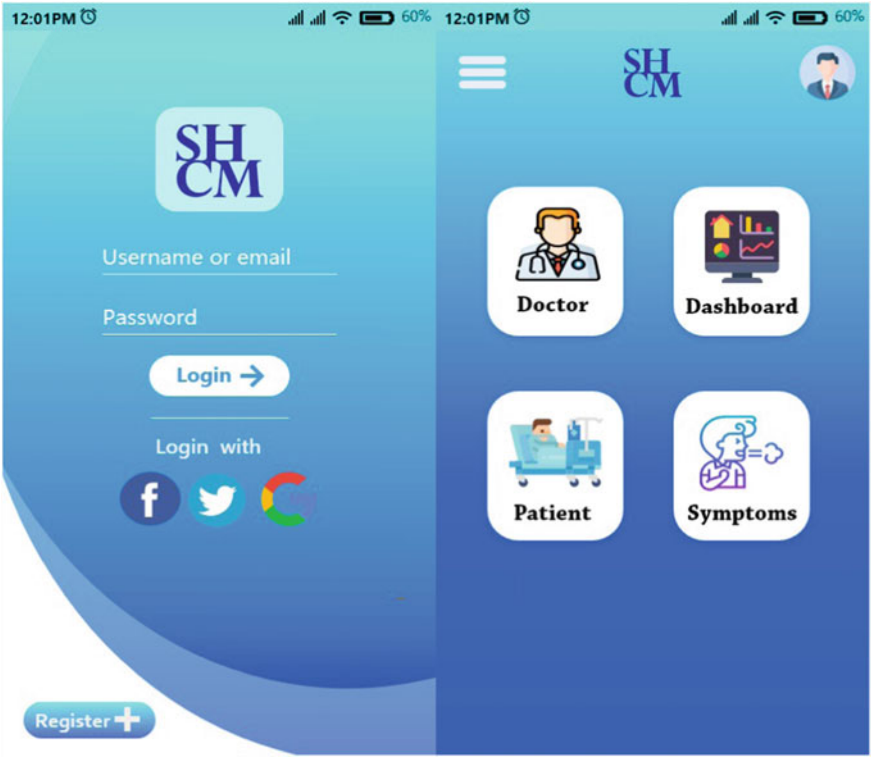


Fig. 8 Login page

Once login page changes, next part enters into Fig. 9 menu alignment tab. In this page, each individual can have an organized details about them, which is also editable.

Figure 10 shows the list of doctors and patients that is displayed when the patient's tab and doctor's tab are clicked, respectively.

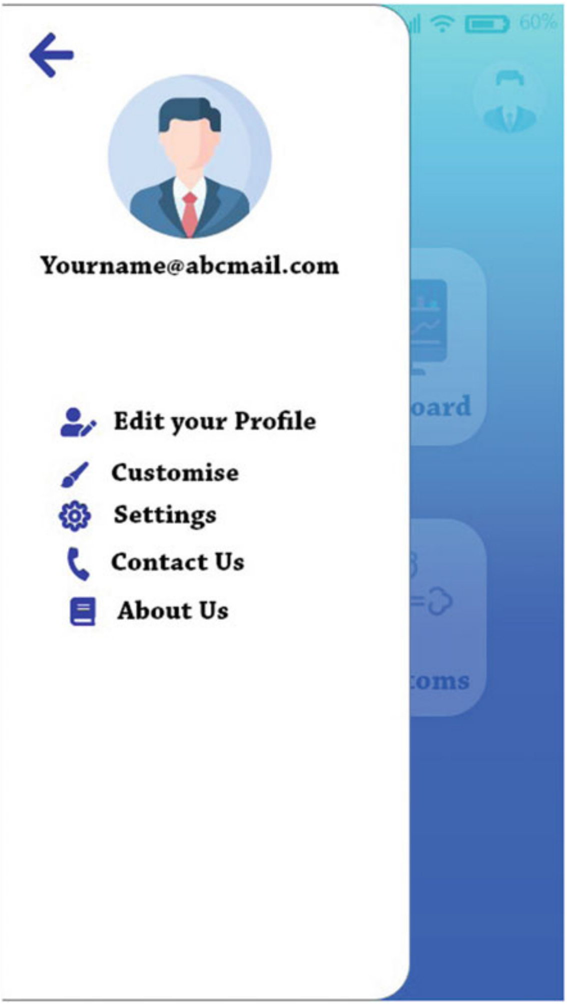
Figure 11 exhibits the fourth stage. This stage has the doctor's tab that contains credentials and details of an individual physicians from various specializations and in 11, doctors also can check their patient's list and details.

Figure 11 confers the patient's tab contains the medical history and personal details of the individual and activity feed, and patients can upload their medical data manually in dashboard and also can look for the doctors.

Figure 12 shows information about personalized monitoring system, and also every individual can regulate their daily schedule for a healthy life.

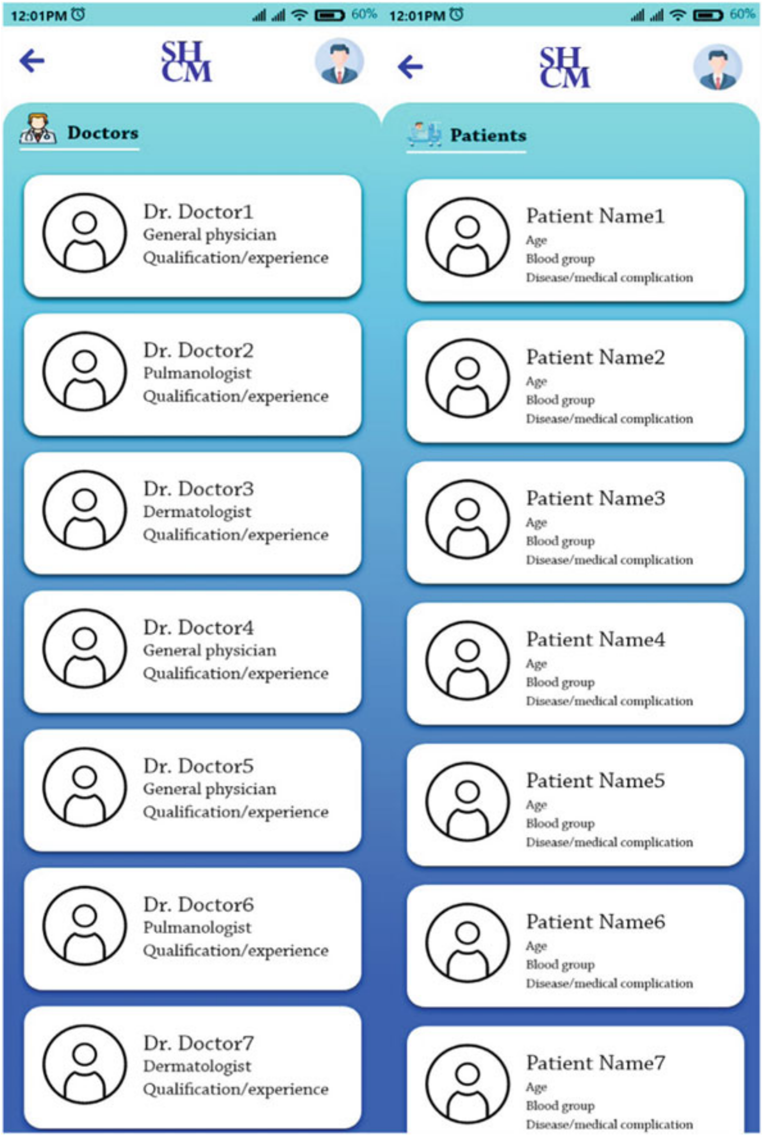
Final stage includes the symptoms pages which is featured in Fig. 12 which is used to help the user in finding the exact symptoms of some common severe diseases as well as the detailed symptoms theory provided for every disease in listed form, and for more pieces of information, Google link is also provided.

Fig. 9 Menu alignment page



5 Future Works

This smart home with condition monitoring project can be improved by developing smartphone mobile application by using Arduino studio software, and it can be integrated into a circuit. It can also be performed in the smartphone application and Blynk smartphone application in order to obtain data from the implied body heat temperature detector, light doorbell, heartbeat rate detector, emergency button and door, automatic light, medical remembrance, and water-level indicator in both auto home side and Medic kit side and share to the users take care and physician.



Doctors

Dr. Doctor1
General physician
Qualification/experience

Dr. Doctor2
Pulmanologist
Qualification/experience

Dr. Doctor3
Dermatologist
Qualification/experience

Dr. Doctor4
General physician
Qualification/experience

Dr. Doctor5
General physician
Qualification/experience

Dr. Doctor6
Pulmanologist
Qualification/experience

Dr. Doctor7
Dermatologist
Qualification/experience

Patients

Patient Name1
Age
Blood group
Disease/medical complication

Patient Name2
Age
Blood group
Disease/medical complication

Patient Name3
Age
Blood group
Disease/medical complication

Patient Name4
Age
Blood group
Disease/medical complication

Patient Name5
Age
Blood group
Disease/medical complication

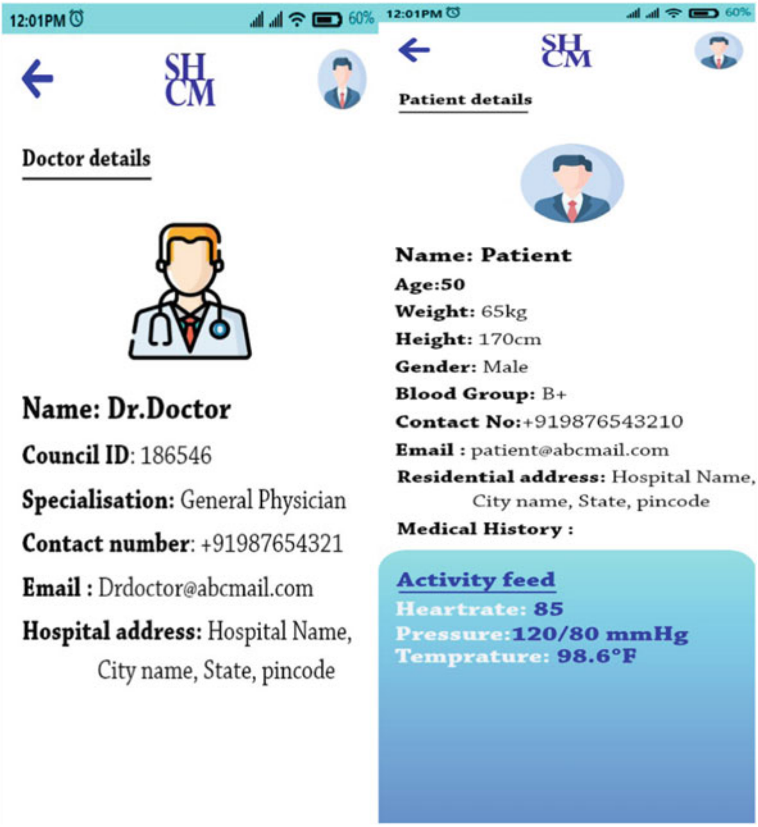
Patient Name6
Age
Blood group
Disease/medical complication

Patient Name7
Age
Blood group
Disease/medical complication

Fig. 10 List of doctors and patients tab

6 Conclusion

Smart home with condition monitoring is a fulfilled system that is expected to lead healthcare, safety, and well-being services to the user's doorstep with the help of modern technologies such as environmental and medical sensors, actuators, high



Doctor details

Name: Dr.Doctor

Council ID: 186546

Specialisation: General Physician

Contact number: +91987654321

Email : Drdoctor@abcmail.com

Hospital address: Hospital Name,
City name, State, pincode

Patient details

Name: Patient

Age:50

Weight: 65kg

Height: 170cm

Gender: Male

Blood Group: B+

Contact No:+919876543210

Email : patient@abcmail.com

Residential address: Hospital Name,
City name, State, pincode

Medical History :

Activity feed

Heartrate: 85

Pressure:120/80 mmHg

Temprature: 98.6°F

Fig. 11 Doctors and patients details page

performance computing processors, and wireless communications platforms. Smart home with condition monitoring has been successfully combined with body heat temperature detector, light doorbell, heartbeat rate detector, emergency button and door, automatic light, medical remembrance, and water-level indicator in both auto home side and medic kit side. These has been integrated into a circuit and also performed in smartphone application. Smart home with condition has also included with smart phone application which can be very useful to mature patients to save and share their medical information to their closed ones and doctors in aspect of such emergency condition. Smart home with condition monitoring has been successfully implemented along with body heat temperature detector, light, doorbell, heartbeat rate detector, medical remembrance, emergency button and door, automatic light, and water-level indicator in both home and medic kit side. These components are then integrated into a circuit, and it is also interconnected to have IoT-based communications in order to establish a better monitoring through smartphone application.

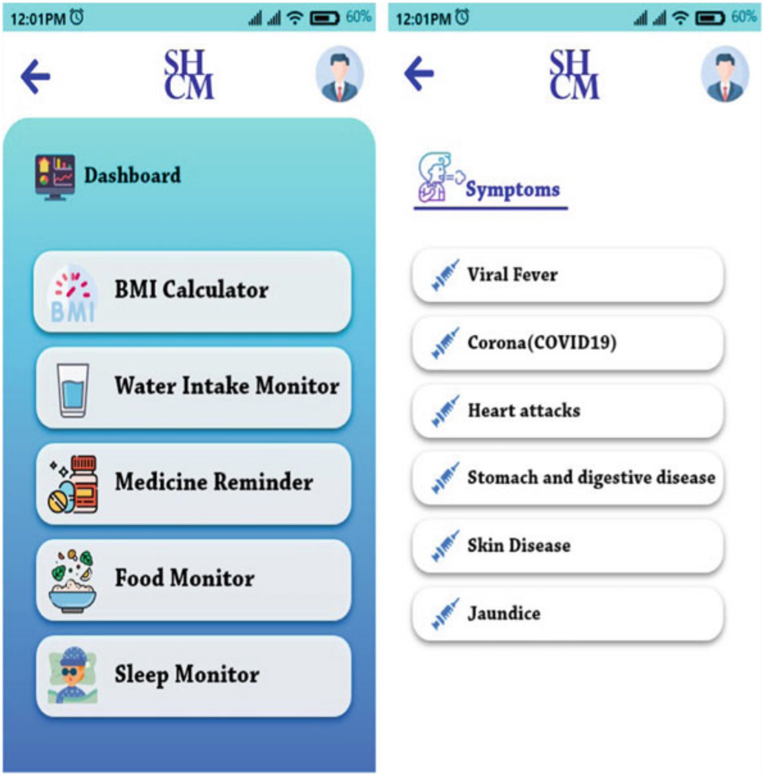


Fig. 12 Dashboard and symptoms page

References

1. Han, Daeman, Lim, Jaehyun, Smart home energy management system using zigbee, in *IEEE Transactions on Consumer Electronics* (2010), pp. 1403–1410
2. R.S. Ransing, M. Rajput, Smart home for elderly care, based on wireless sensor network, in *2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE)* (2015), pp 1–5
3. E. Ruiz, R. Avelar, X. Wang, Poster: protecting remote controlling apps of smart-home-oriented iot devices, in *2018 IEEE/ACM 40th International Conference on Software Engineering: Companion (ICSECompanion)* (2018), pp. 212–213
4. D. Stefanov, Z. Bien, W.-C. Bang, The smart house for older persons and persons with physical disabilities: structure, technology arrangements, and perspectives. IEEE Trans. Neural Syst. Rehabil. Eng. **12**, 228–250 (2004)
5. M. Elkhodr, S. Shahrestani, H. Cheung, A smart home application based on the internet of things management platform, in *2015 IEEE International Conference on Data Science and Data Intensive Systems* (2015), pp. 491–496
6. B. Ghazal, K. Al-Khatib, Smart home automation system for elderly, and handicapped people using xbee, in *International Journal for Smart Home* (2015), pp 203–210

7. A. Aswathy, G. Sukumar, M. Kumar, Akshay, Asha, V.R. Pandi, Solar powered intelligent electric wheel chair with health monitoring system, in *2017 International Conference on Technological Advancements in Power and Energy (TAP Energy)* (2017), pp. 1–5
8. Aromal, Gokulnath, Amrithesh, Arun, A. Varma, V. R. Pandi, Design and implementation of a solar integration in electric wheelchair, in *2018 4th International Conference for Convergence in Technology(I2CT)* (2018), pp. 1–6
9. Chan, Hariton, Ringeard, Campo, Smart house automation system for the elderly and the disabled, in *1995IEEE International Conference on System,Man and Cybernetics.Intelligent Systems for the 21st century*, vol. 2 (1995), pp. 1586–1589
10. S. Dhiman, S. Gupta, A. Chopra, P.C. Vashist, Home computerization for the elderly and disabled with internet of things, in *2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC)* (2019), pp. 570–573
11. S. Bulusu, M. Krosuri, R. Koripella, N. Sampath, Smart and secure home automation using internet of things enabling technologies. J. Comput. Theo. Nanosci. 390–395 (2020)
12. D. Pal, T. Triyason, S. Funikul, Smart homes and quality of life for the elderly: a systematic review, in *2017 IEEE International Symposium on Multimedia (ISM)* (2017), pp. 413–419
13. D.P. Suja, Purushothaman, Development of smart home using gesture recognition for disabled and elderly. J. Comput. Theo. Nanosci. **17**, 77–181 (2020)
14. Majumder, Aghayi, Noferesti, Memarzadeh-Tehran, H. Mondal, Pang, Deen, Smart homes for elderly healthcare-recent advances and research challenges. Sensors (Basel, Switzerland) (2017)
15. P. Suesaowaluk, Home automation system based mobile application, in *2020 2nd World Symposium on Artificial Intelligence (WSAI)*(2020), pp 97–102
16. S. Bajpai, D. Radha, Smart phone as a controlling device for smart home using speech recognition, in *International Conference on Communication and Signal Processing—ICCSP' 19* (2019)
17. S.K. Sooraj, E. Sundaravel, B. Shreesh, K. Sireesha, Iot smart home assistant for physically challenged and elderly people, in *International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India* (2020)

Developing a Smart System for Reducing Traffic Congestion

Nawsheen Tarannum Promy, Proteeti Prova Rawshan, Md. Mahfujul Islam, and Muhmmad Nazrul Islam

Abstract Traffic congestion is a common challenge in the developing countries like Bangladesh. This is the quotidian scenario in most of the metropolitan cities of the world. Because of heavy traffic, people lose their valuable time from their busy schedule. As the number of road users is continuously increasing and resources are limited, developing an intelligent automated traffic control system is essential. Therefore, the need arises for simulating and optimizing traffic control systems to accommodate this increasing demand in a better way. In this paper, an automated traffic control system has been proposed. To achieve this objective, firstly, a research study was conducted through semi-structured interviews and observation to reveal the requirements for developing an automated traffic control and management system for Dhaka City. Secondly, a prototypical automated traffic management system was developed by considering the revealed requirements. Thirdly, the proposed system was evaluated with seven participants in a laboratory environment and found that the proposed system is effective and efficient to control the traffic congestion in Dhaka City.

Keywords Smart system · Traffic management · Speed control · Traffic jam · Bangladesh

1 Introduction

Dhaka, the capital of Bangladesh, is one of the most populous, jam-packed, and busiest cities globally [1]. Approximately 20 million people live in this city [2] while about 160 thousand motor vehicles are registered in Bangladesh [3]. Due to a sloppy traffic management system, lack of strategic planning and proper traffic infrastructure

N. T. Promy · P. P. Rawshan · Md. M. Islam · M. N. Islam (✉)
Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka 1216, Bangladesh
e-mail: nazrul@cse.mist.ac.bd

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_51

facilities, high traffic jams are noticed inside the city; thus people miss their essential work or cannot even reach their respective destination in due time [4].

One of the major causes of this traffic is breaking the traffic rules. Narrow roads, unnecessary overtaking, illegal parking, riding mostly at the busy hours (8.00 am - 11.00 am and 4.00 pm - 8.00 pm), etc., are the prime reasons for this traffic jam in the cities. Irresponsible drivers always violate the traffic rules like changing lanes, breaking the speed limit, overtaking from the wrong side. For instance, according to Dhaka Metropolitan Police (DMP), on March 21st, 2020, the number of the total case lodged against 370 vehicle owners due to driving in the wrong direction [5]. Again, they tend not to follow the parking restriction and park cars on the main road, making the streets narrow.

Another problem arises due to the lights on the main road. Sometimes due to fog [6] or heavy rain, it becomes quite challenging to see the road correctly, despite this problem the traffic control system does not allow to switch on the light in that weather. In that dim daylight for safety purpose, people start driving their cars slowly. That may also create traffic in gloomy weather [7].

Therefore, this paper aims to design and develop a smart and self-control traffic system to abridge the traffic congestion in Dhaka City. To accomplish the objective, the existing system has been discussed and a requirement elicitation study was conducted to figure out the required features that should be incorporated to develop a system for Dhaka City residents and introduced a smart system to address the needed features.

The rest of the paper is organized as follows. The related studies are briefly discussed in Sect. 2. Section 3 discusses the requirement elicitation study. Section 4 illustrates the conceptual framework of the proposed system. Section 5 briefly discusses the features of the proposed system. Section 6 delineates sample cases that evaluate the system. The future expansion of the project with its limitations, followed by concluding remarks, is discussed in Sect. 7.

2 Literature Review

This section briefly introduces the related work. To find out the related literature, the major scholarly databases were searched including ACM digital library, Google Scholar, IEEE explorer, Researchgate, and ScienceDirect, using suitable search strings.

RFID tags were used for resolving numerous traffic issues. For example, RFID was used to detect the frequency of transportation in a specific time-period and traffic congestion and to control the traffic signals accordingly in [8]. But this system can track congestion only for one-way roads, not for roundabouts. Again, RFID-based parking-lot management system was proposed by Parkhi et al. [9] where payment can be made by crosschecking the check-in and check-out time. But the system allows to pay after checking out from the parking lot, then later paying the fines for further checking-in, which does not help the scarcity of parking necessity; hence, anyone can take advantage of the system. In another study, Sahu et al. [10] proposed a GSM-based parking system where the system interacts with the user through SMS-based authentication to enhance the security.

For speed controlling and avoiding overtaking problems, Barcel et al. [11] introduced a system where "superblock" zones are deployed in the most highly dense traffic areas of the city. Whenever vehicles enter the superblock, they have to maintain an imposed speed limit to avoid these problems. But the amount of space and resource, it requires to deploy may not be possible in the context of Dhaka. Therefore, the whole city would require to be re-planned accordingly. The effect of speed in traffic safety is studied by Nilsson and Goran [12] and further considered by Vaughn and David [13], Thakker and Kaushal [14], Fujii and Yutaka [15]. They proposed solutions where the speed will be obtained via GPS, satellite, or transmitter and will be compared with the maximum speed limit. Under the moto-enforcement digitalization program, Bangladesh Road Transport Authority has installed RFID tags in the cars and monitors the speed with RFID stations [16]. But none of these introduced systems penalizes the driver, and the vehicle may get away with breaking the law.

Building queues and delayed traffic flow are evident in traffic junctions. Kimber and Hollis [17] tried to figure out the effect of signals and roundabouts on the length of the queue and average traffic flow. As outcome, they found that roundabouts seemed a better alternative as vehicle speed was reduced and congestion was lesser [18]. Again, Troutbeck [19] studied that the behavior of vehicles changes with increasing circulating flow, while Maycock & Hall [20] discussed the accidents and safety for the pedestrians at four-arm roundabouts. None of these studies discussed dividers and bars, which hampers traffic flow and the unsettling problems.

Using the real-time data collected from California, Lopez et al. [21] came up with a hybrid model that could anticipate congestion of a 9 km long stretch of California. Devi et al. [22] introduced a machine learning-based traffic congestion prediction for a IoT-based smart city that will notify the vehicles before entering into a congested path. Lv et al. [23] proposed a deep learning-based traffic flow prediction where a stacked autoencoder model was used to absorb generic traffic flow features. Again, Soomro et al. [24] suggested a system that would combine both IoT and AI to reduce the traffic congestion in a smart city. An innovative traffic volume and vehicle classification monitoring system were proposed by Huang et al. [25]. They proposed an algorithm, and with their experiments, they have proved that accelerometer is sufficient enough for accurate monitoring and classification.

As the organization of a proficient traffic system needs a well-distributed power supply, many energy-consuming insights have been introduced along with planning the traffic system of a city [26, 27]. But neither of these systems suggest using this power for maintaining traffic nor for street lights.

In sum, there are numerous existing ways of implementing a smart traffic control system, but most of them focus on solving one aspect, ignoring other issues. There is almost no study that is focused on the context of Dhaka City. Thus, this research focused on developing a self-controlled traffic management system to ensure safety, saves energy, and keeps track of every penalty.

3 Requirement Elicitation Study

Understanding the user requirements is foremost important task for developing any useful software system [28–30]. Thus, the main motive of the requirement elicitation study was to identify the requirements for developing a traffic control and management system that can help a large number of residents of Dhaka City to move around the city systematically. The requirement elicitation study was conducted through semi-structured interviews and observation.

3.1 Participant's Profile

The participants include passengers, teenagers, school and college-going students, passerby, job holders, rickshaw pullers, car and bus drivers, and traffic police. A total of 30 participants (18 male and 12 female) were interviewed who lived in different areas in Dhaka City. The participants' average age was 31 and ranged between 16 and 65. Among the participants, 12 of them had their cars, and the rest traveled by public transports or ride-sharing applications. Among those 12 participants (car owners), 9 had personal drivers for their vehicle, and the rest drove their car.

3.2 Study Procedure

At first, the participants were briefed about the interview's purpose and take their written consent to maintain anonymity and research ethics. During interviews, participants were asked several questions regarding the problems, leading to massive congestion inside the city. Apart from this, other questions like how they regularly travel, which type of vehicle they prefer, how often they have faced difficulties in finding parking for their vehicle (if they have), how often they have observed accidents due to speeding and low light, where they have faced most traffic in the roads were asked during the interview sessions. The interview responses were recorded and later transcribed for analysis. Apart from this, to understand the traffic phenomena properly, we have visited some overpopulated areas of Dhaka City to observe the possible reasons for a traffic jam.

3.3 Study Findings

Based on interview responses and observation, the following inconveniences caused by the traffic system have been found:

a) *Vehicle entry access control:* In the present traffic system, there is no hard and fast rule of entry time of heavy vehicles. Therefore, buses or trucks can quickly enter the metropolitan area in busy hours (8.00 am - 11.00 am and 4.00 pm - 8.00 pm). This creates enormous traffic congestion in Dhaka City.
b) *Rule breaking tendency of the driver:* In Dhaka City, most drivers try to overtake other vehicles without considering the rules. They do not follow the speed limit signs on the road. The public buses which carry passengers stop their buses from taking extra passengers in the middle of the road, which is one of the primary reasons for traffic congestion in Dhaka.
c) *Ignorant drivers and inadequacy of parking:* Drivers also do not follow the rules of parking. They park their cars beside the road, making the streets narrow, and other vehicles cannot move properly.
d) *Excessive congestion in four-way junction roads:* In four-way junction roads, drivers often try to take the right turns, which makes a complicated and jam-packed situation on the streets. In many busy streets and highways, there are always signs not to take the right turns. Irresponsible and ignorant drivers often break these rules because they want to reach early to their destination. But instead of arriving early, they create massive traffic in the middle of the road.
e) *Accidents due to dim lights and foggy weather:* During the rainy season, there is often a very dim light, and in winter, there is colossal fog in the morning. But there is no automated system of switching on the light to make a clear vision for the drivers. Due to dim light and foggy weather, drivers can not see other vehicles coming from the opposite sides of the road.

4 The Conceptual Framework

The conceptual framework of the proposed system to address the revealed requirements is shown in Fig. 1. The conceptual framework consists of five features:

a) A toll booth will be (i.e., Gate RFID Receiver) at the entry point of the city. The system will allow only the registered vehicles to enter the city. If any vehicle is not registered for that metropolitan area, it will not be allowed to enter. The unregistered vehicle will get this opportunity after registering. After registering, every vehicle will get a smart card (i.e., RFID Card). They can recharge their card anytime from any toll booth. This is necessary to check the vehicles if they are violating any rules. If they break the rules, a certain amount of money will be cut from the smart card.
b) Parking charge will be collected automatically by the proposed system. This system will be automatized via an Android application. In every parking lot, there will be a toll booth. Once the car enters the parking booth, a certain amount of money will be cut from its smart card.

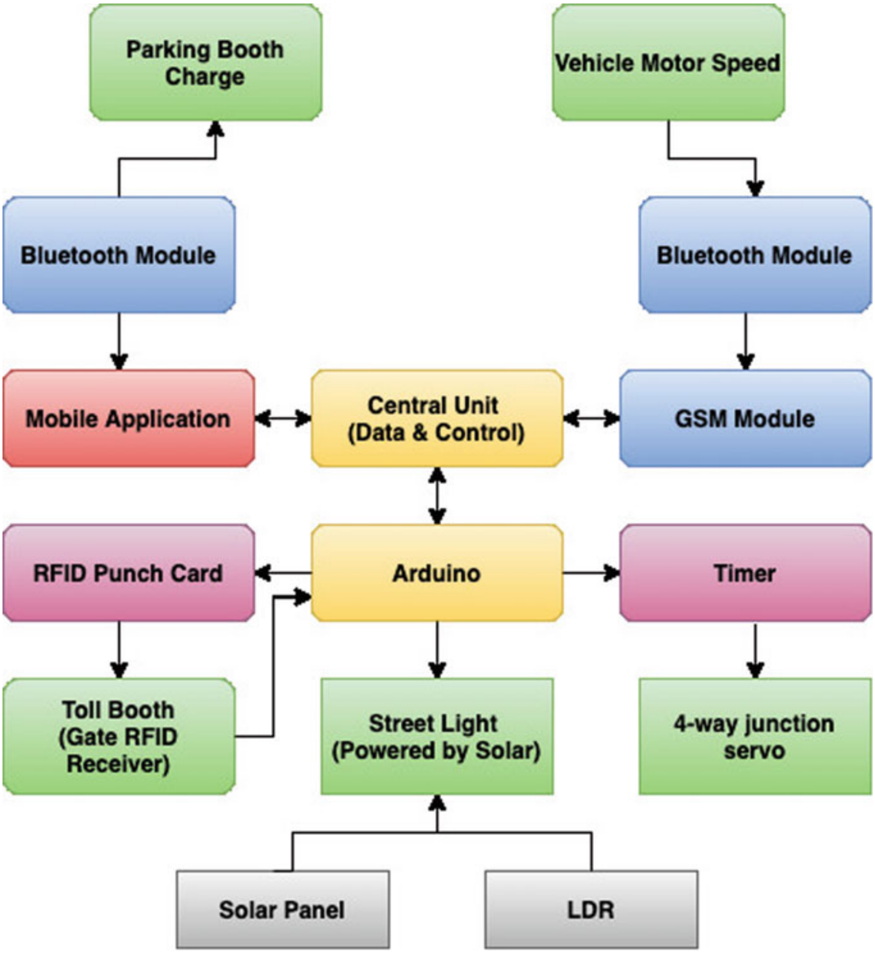


Fig. 1 Conceptual framework of the proposed system.

c) Speed limit will be checked by the system via checking the motor speed with the standard speed limit stored in the central unit. If any car violates the rule of the speed limit, it will be penalized.

d) A pair of automatic bar will be introduced to control the convolution in the middle of a four-way junction. It will make the four-way junction into two ways so that no vehicles can take a right turn by violating the traffic rules.

e) Solar energy will be used in the toll booths and the street lights. The street light will be automatically switched on if it finds the low intensity of daylight via LDR and solar.

All these hardwares are controlled with Arduino. The outputs are checked and matched against the data stored in the server.

5 Development of the Proposed System

The proposed automated traffic control and management system were developed into two parts: a prototypical version of the hardware part and a mobile application. The required hardware components are shown in Table 1. The mobile application is built with Android Studio while user interfaces are made using basic XML Markup, and PHP was used to connect the database to the MySQL server.

5.1 The Mobile Application

An Android mobile application was built for the proposed parking system. The application has been developed in an Android development platform called Android Studio.

The programming language Java was used to develop the application as it is well-supported by the Android Studio. The UI of the mobile application is made of the Extensible Markup Language (XML), as the format is both human- and machine-readable. The UI design for parking using the proposed application is shown in Fig. 2.

MySQL has been used as the database in the proposed system, the most popular and widely used database. After the registration process of a person, their identity is tracked by a UNIQUE KEY in the MySQL server. In the MySQL server, several rows have been allocated for different owners to have their different information. An open-source server has been configured to store the information of the owners of the cars, their arrival time in the city, last departure time, records of crossing the speed limits, parking charge, etc. PHP is the server-side scripting language that is used to connect the MySQL database server to Android Studio.

Table 1 Required components of the system

Functionalities	Components
Entire system control	Arduino Mega
Vehicle prototyping	Arduino Uno
SMS notification for penalization	GSM
i) Vehicle authentication and toll booth	RFID card & reader
Collecting speed limit from vehicle to central unit	Bluetooth module
Automated bar control for four-way junction	Servo motors
i) Street lights, toll Booth and bar control	Solar panel, LED, LDR

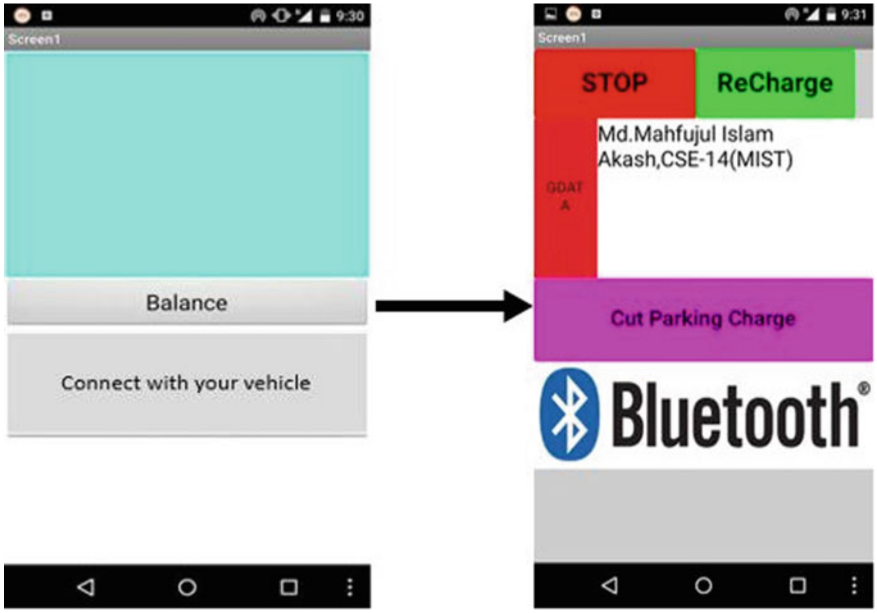


Fig. 2 User interfaces of the mobile application

5.2 Hardware Implementation

The prototypical implementation of the hardware system is shown in Fig. 3. At the entrance of the city, we have established a toll booth. The flowchart of the city entrance system is shown in Fig. 4. Here, every car needs to pay a fixed amount of money to get permission to travel inside the city. Initially, we have considered 1000 TAKA as a fixed amount to enter into the city. A smart RFID card is given to every vehicle owner, which they punch before entering the city. If any vehicle isn't registered earlier, that vehicle will be registered with the smart punch card right away.

The parking system of our proposed system is shown in Fig. 5. At the entrance of the parking lot, there will be a toll booth. If any person wants to park a car there, the toll collector will connect the mobile application via Bluetooth. A certain amount of money will be charged from the smart card, and the amount will be shown on the phone of the vehicle owner. A counter is set, which will cross-check how much time a car spends inside the parking lot and will charge accordingly. In the prototype, we have pondered that we will charge 30 TAKA to the vehicle owner in front of the entrance and then for every hour of parking 40 TAKA will be charged from the smart card. It is also checked if the vehicle owner has sufficient balance to park his vehicle for minimum 8 h in the parking lot.

For every roundabout or four-way junction, bars are added for every pair of roads. When the first pair of opposite lanes are open, a bar will come down in parallel to these

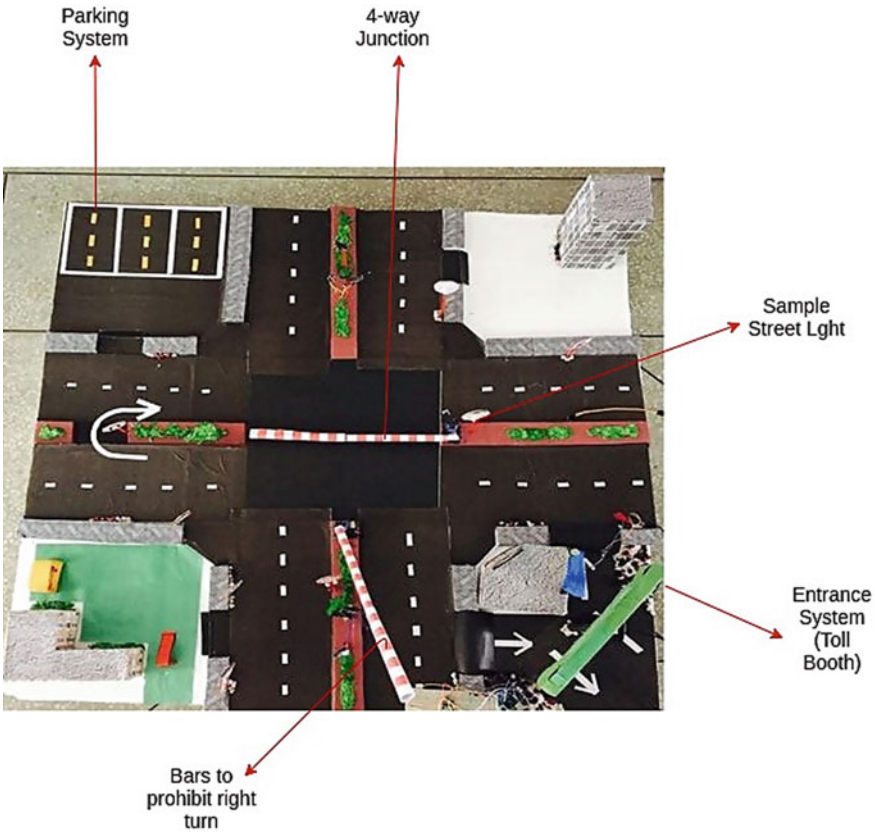


Fig. 3 Prototype of the implemented system

opposite lines to prohibit the right turn because this right turn creates unnecessary traffic in the midpoint, and the vice-versa is applied in the other pairs of opposite lanes (see Fig. 6). Here we meditate the threshold of the time is 120 s for each pair of lanes in a four-way junction.

The flowchart, as shown in Fig. 7, shows how the proposed system controls the speed limit. If the drivers break the speed limit for any reason, they will be penalized by a fixed amount of money. A counter will be checked how many times they have broken this rule. From the wheel of the vehicle, we will calculate the revolution per minute (RPM) of the wheel. A threshold RPM will be set. If any wheel exceeds that limit, this means the vehicle is crossing its speed. With the help of the Bluetooth module, a message will be transmitted to the GSM module in the central unit. This will keep track of the number of breaking the rules, will cut a certain amount of money from the smart card of the owner of that vehicle, and a message will be sent to the owner's mobile to notify him. In our prototype, we have contemplated 200 rpm as the speed limit. If any vehicle owner crosses the threshold of 200 rpm, he/she

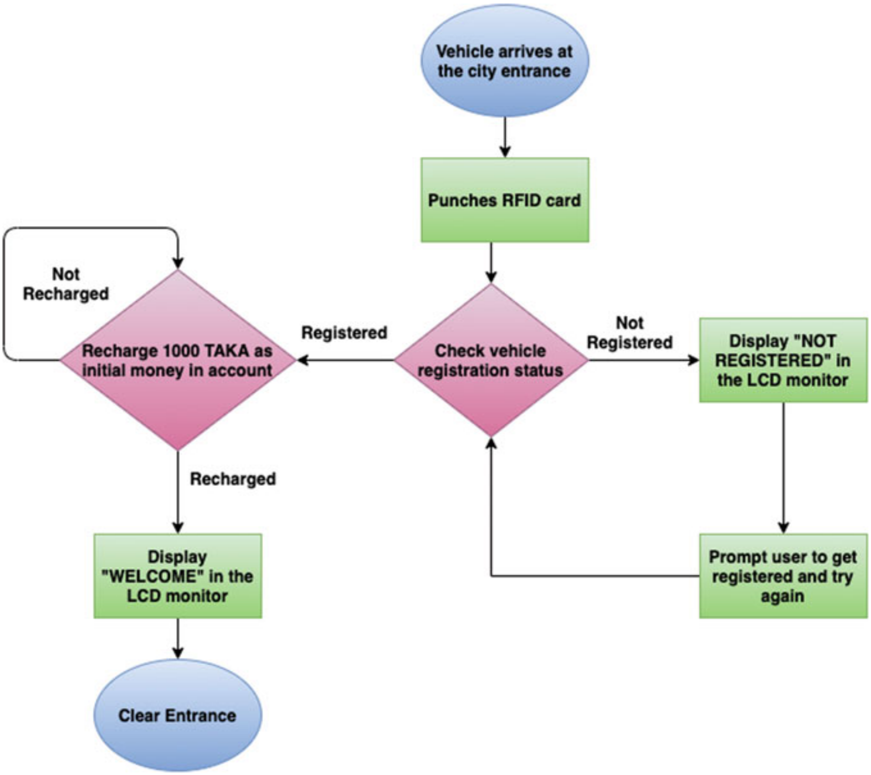

Vehicle arrives at the city entrance
Punches RFID card
Check vehicle registration status
Not Registered
Display "NOT REGISTERED" in the LCD monitor
Prompt user to get registered and try again
Registered
Recharge 1000 TAKA as initial money in account
Not Recharged
Recharged
Display "WELCOME" in the LCD monitor
Clear Entrance

Fig. 4 Flowchart of city entrance system

will be charged 50 TAKA for breaking the speed limit. Simultaneously, an SMS will be sent to the owner's mobile number that he/she has broken the speed limit and penalized.

The whole system proposes a smart traffic controlling system along with decreasing the pressure on the traditional way of consuming power by using the solar panel and using the consumed energy for the streetlights and also to supply this electricity to the underground parking area and toll collect booth.

6 Evaluating the System

A lightweighted evaluation study was conducted following the approach adopted in [31] at the Software Engineering Lab of the authors' institute, replicated with 57 participants (fifty students and seven faculty members). Initially, participants were briefed about the study's purpose and then demonstrate the system for 3–4 minutes. After that, participants were asked to perform each function. A set of test cases were

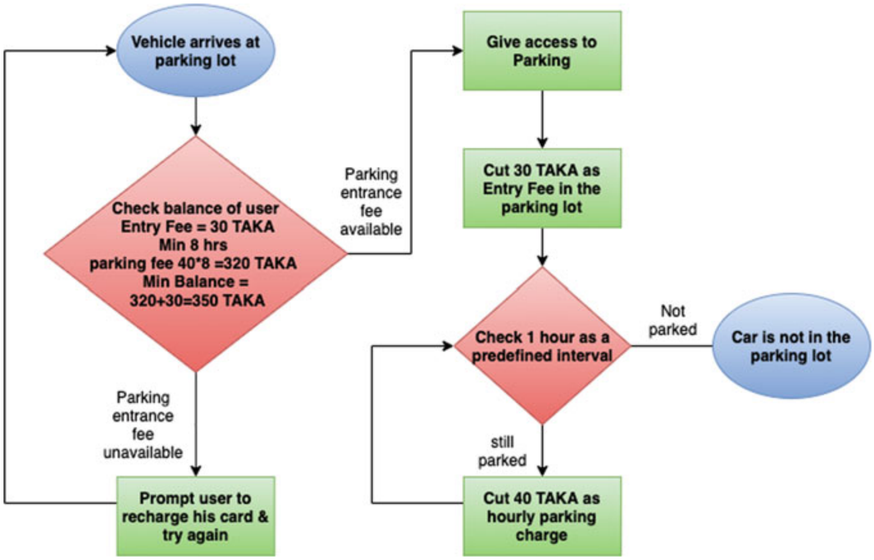

Vehicle arrives at parking lot
Check balance of user Entry Fee = 30 TAKA Min 8 hrs parking fee 40*8 =320 TAKA Min Balance = 320+30=350 TAKA
Parking entrance fee available
Give access to Parking
Cut 30 TAKA as Entry Fee in the parking lot
Check 1 hour as a predefined interval
Not parked
Car is not in the parking lot
still parked
Cut 40 TAKA as hourly parking charge
Parking entrance fee unavailable
Prompt user to recharge his card & try again

Fig. 5 Flowchart of parking system

prepared to evaluate the system functionality, as given in Table 2. The outcomes of this study are also shown in Table 2.

The result showed that the maximum number of participants completed the maximum number of tasks at their first attempt. For example, entrance check-in and bar control test were completed by 80% within their first attempt. However, the speed test had more failures because of not maintaining the fixed RPM all the time. Again, solar energy works comparatively better than other features even though it sometimes failed due to the lack of charging.

7 Discussions and Conclusion

This research proposes a conceptual framework and develops a prototypical version of smart traffic control and management system for one of the most densely populated cities (i.e., Dhaka, the capital of Bangladesh). The evaluation study showed that the proposed system is effective and efficient for traffic control and management. This system automates the process of traffic control system, hence save time, cost, and effort. The proposed system will help make general people more conscious about the traffic rules and minimize the traffic jam in the four-way junction points utilizing solar energy.

This research has some limitations as well. The proposed system cannot detect a vehicle on the road and could not find vehicles that change lanes, overtake unneces-

Table 2 Evaluation table of the proposed system

Functionality	Test case	Success rate (n = 57) (%)	Observation/Findings
Entrance Check-in	Upon vehicle entry, smart card is swiped at the booth. If the user is registered, “Allowed” is shown. If not, it is displayed “Not Registered”.	98.2	(i) Unregistered vehicles must not be allowed to enter into the city. Once they become a registered member, they will be authorized to enter. (ii) To enter the city, one must have to have a smart card.
Parking Charge	When a vehicle enters the parking lot, a toll manager will start a counter on the application for cutting the parking charge. User is notified via SMS is there is not sufficient money.	82.4	(i) If the current amount is no sufficient that for which the minimum parking charge cannot be cut, it sends a message to the user to recharge credits. (ii) User needs to be notified about how much money is needed to be kept as minimum.
Speed Test	To check the speed limit of a car, a threshold rpm is fixed. If any car breaks this, it will be counted as a rule violation. If the speed is violated, an SMS is sent to warn the driver.	71.9	(i) Sends the user messages when the highway is clear, and user can increase speed without any disruption to any citizen. (ii) Also, not always a fixed RPM can be maintained, even if that is under the speed limit.
Solar Energy Test	Solar is used to give power to our street lights. For this, on daylight, solar panel is charged and is used to light street lights at night.	91.2	(i)The prototype is taken into a dark area where the light intensity is low. After some time, street lights turn on automatically. (ii) If the solar panels are not sufficiently charged on daylight, the streetlights won’t light for much longer.
Bar Control Test	When the first pair of two roads is green in the signal, then the bar comes down between the roads.	100	In our prototype, we have given 5 s delay of changing green into red or red into the green signal. It worked properly.

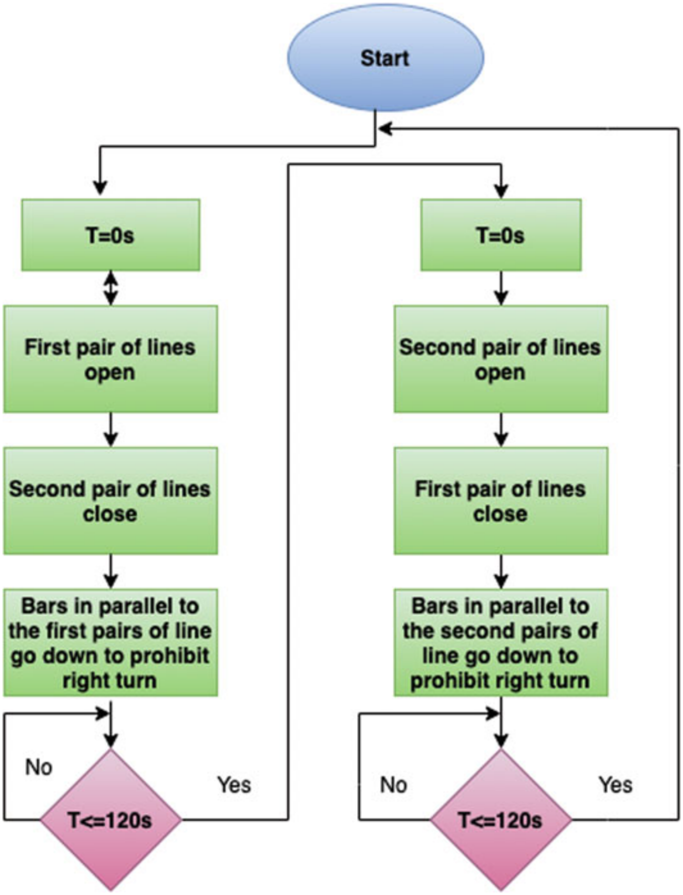

Start
T=0s
First pair of lines open
Second pair of lines close
Bars in parallel to the first pairs of line go down to prohibit right turn
No
T<=120s
Yes
T=0s
Second pair of lines open
First pair of lines close
Bars in parallel to the second pairs of line go down to prohibit right turn
No
T<=120s
Yes

Fig. 6 Flowchart of four-way traffic system

sarily, and back gears on a busy highway. Again, the parking system needed to be more space-optimized and efficient. Finally, only the prototype of the system was developed.

Future work will emphasize working on these limitations to make the system more efficient. The detection of motion using a PIR sensor will help us to use streetlights more efficiently. To make the parking system more efficient, a parking application will be introduced for finding a parking spot in the nearest location quickly.

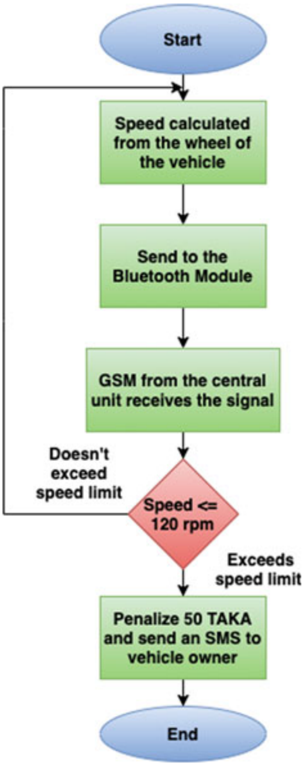

Start
Speed calculated from the wheel of the vehicle
Send to the Bluetooth Module
GSM from the central unit receives the signal
Doesn't exceed speed limit
Speed <= 120 rpm
Exceeds speed limit
Penalize 50 TAKA and send an SMS to vehicle owner
End

Fig. 7 Flowchart of speed limit

References

1. K.M. Munim, I. Islam, M. Sarker, M.N. Islam, Towards developing an intelligent automated water pumping system for dhaka city in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (2019), pp. 1–5. IEEE
2. Dhaka: Bangladesh metro area population 1950-2020. https://www.macrotrends.net/cities/20119/dhaka/population. [Online; Last accessed 31-January-2021]
3. Bangladesh motor vehicle registered: Dhaka: Total. https://www.ceicdata.com/en/bangladesh/motor-vehicle-registered/motor-vehicle-registered-dhaka-total. [Online; Last accessed 31-January-2021]
4. A. Quium, S. Hoque, The completeness and vulnerability of road network in bangladesh. Eng. Concerns Flood, 59–75 (2002)
5. Daily report on case lodged under mvo. https://dmp.gov.bd/daily-report-on-case-lodged-under-mvo/. [Online; Last accessed 4-July-2020]
6. Traffic accident in heavy fog kills three. https://www.dhakatribune.com/bangladesh/nation/2018/01/16/accident-fog-kills-three/. [Online; Last accessed: 9-June-2020]
7. Storm causes waterlogging, traffic congestion in Dhaka. https://bdnews24.com/bangladesh/2018/04/29/storm-and-rain-causes-waterlogging-traffic-congestion-in-dhaka. [Online; Last accessed: 9-June-2020]

8. R. Sundar, S. Hebbar, V. Golla, Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection. IEEE Sensors J. **15**(2), 1109–1113 (2015)
9. P. Parkhi, S. Thakur, Rfid-based parking management system. Int. J. Adv. Res. Comput. Commun. Eng. (2014)
10. V.G. Sahu, V. Gulhane, N. Shelokar, A web based centralized vehicle parking system using gsm security. IJAIEM **2**(4) (2013)
11. J.L. BarcelóJ, D.G. Ferrer, R. Grau, *Microscopic Traffic Simulation for ATT Systems Analysis a Parallel Computing Version* (Nov 13 2012), uS Patent D670, 583
12. G. Nilsson, Traffic safety dimensions and the power model to describe the effect of speed on safety. Bulletin-Lunds Tekniska Högskola, Inst för Teknik och Samhälle, Lunds Universitet **221** (2004)
13. D. Vaughn, *Vehicle Speed Control Based on GPS/Map Matching of Posted Speeds* (Jan 16 1996), uS Patent 5,485,161
14. K. Thakker, *Wireless Intelligent Vehicle Speed Control or Monitoring System and Method* (Jun 12 2001), uS Patent 6,246,948
15. Y. Fujii, *Vehicle Speed Control System* (May 24 1994), uS Patent 5,315,295
16. BRTA Annual Report. http://www.brta.gov.bd/site/page/4e043627-80ef-4f0f-a1df-320a870c1968. [Online; Last accessed: 9-June-2018]
17. R. Kimber, E.M. Hollis, Traffic queues and delays at road junctions. Tech. rep. (1979)
18. M.E. Fouladvand, Z. Sadjadi, M.R. Shaebani, Characteristics of vehicular traffic flow at a roundabout. Phys. Rev. E **70**(4), 046132 (2004)
19. R. Troutbeck, Traffic interactions at roundabouts, in *Australian Road Research Board (ARRB) Conference, 15th, 1990, Darwin, Northern Territory*, vol. 15 (1990)
20. G. Maycock, R. Hall, Accidents at 4-arm roundabouts. Tech. Rep. (1984)
21. P. Lopez-Garcia, E. Onieva, E. Osaba, A.D. Masegosa, A. Perallos, A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy. IEEE Trans. Intelligent Transp. Syst. **17**(2), 557–569 (2015)
22. S. Devi, T. Neetha, Machine learning based traffic congestion prediction in a iot based smart city. Int. Res. J. Eng. Technol **4**, 3442–3445 (2017)
23. Y. Lv, Y. Duan, W. Kang, Z. Li, F.Y. Wang, Traffic flow prediction with big data: a deep learning approach. IEEE Trans. Intelligent Transp. Syst. **16**(2), 865–873 (2014)
24. S. Soomro, M.H. Miraz, A. Prasanth, M. Abdullah, *Artificial Intelligence Enabled iot: Traffic Congestion Reduction in Smart Cities* (2018)
25. Y. Huang, L. Wang, Y. Hou, W. Zhang, Y. Zhang, A prototype iot based wireless sensor network for traffic information monitoring. Int. J. Pavement Res. Technol. **11**(2), 146–152 (2018)
26. J. Byrne, J. Taminiau, L. Kurdgelashvili, K.N. Kim, A review of the solar city concept and methods to assess rooftop solar electric potential, with an illustrative application to the city of seoul. Renew. Sustain. Energy Rev. **41**, 830–844 (2015)
27. RD: *Renewable Energy Policies in the Gulf Countries: A Case Study of the Carbon-neutral "Masdar city" in Abu dhabi*. Elsevier (2009)
28. S. Hoque, S.S. Sharmee, M.N. Islam,D. Shahrin, F. Kabir, Ponno aalap: an interactive web portal for improving consumer experience, in *2020 IEEE Region 10 Symposium (TENSYMP)*, IEEE (2020), pp. 1770–1774
29. M.N. Islam, S.R. Khan, N.N. Islam, M. Rezwan-A-Rownok, S.R. Zaman, A mobile application for mental health care during covid-19 pandemic: Development and usability evaluation with system usability scale, in *International Conference on Computational Intelligence in Information System*. Springer, Berlin (2021), pp. 33–42
30. M.R. Jahan, F.I. Aziz, M.B.I. Ema, A.B. Islam, M.N. Islam, A wearable system for path finding to assist elderly people in an indoor environment, in *Proceedings of the XX International Conference on Human Computer Interaction* (2019), pp. 1–7
31. T. Hossain, M.S.U.A. Sabbir, A. Mariam, T.T. Inan, M.N. Islam, K. Mahbub, M.T. Sazid, Towards developing an intelligent wheelchair for people with congenital disabilities and mobility impairment, in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (2019), pp. 1–7. IEEE

Meta-Heuristics–Fireflies Correlated Algorithm for Optimized Routing in Manets

M. Ilango, A. V. Senthil Kumar, and Binod Kumar

Abstract The transmission of data from a source to a destination node is examined in this article. Existing binomial logistic regression (BLR-OR) routing algorithm is employed to effectively higher network lifetime with lower delay. However, regression coefficient value of a single mobile node in MANETs takes more time to compute the distance of neighbor node, so that packets fail to reach the destination and it was expired. Meta-heuristics algorithm is introduced to overcome the existing problem. During the transmission, packets from source to neighbor node using correlation coefficient value has been computed based on energy, bandwidth and delay time. Optimized routing path of a network represents the attractiveness factors of firefly algorithm. Data packets are sent to neighbor node; it saves energy, bandwidth and improves the packet delivery ratio.

Keywords MAENTs · Meta-heuristics · Bandwidth · Energy · Packet delivery ratio · Delay time · Correlation · Routing

1 Introduction

Mobile Ad hoc networks are wireless networks that connects collection of mobile nodes. Each node will travel independently in either direction, or as a result, may always shift its lines to other devices. Since the nodes act as routers in MANET, the data transition between nodes is handled by other nodes.

MANETS is a self-creating, self-organizing and autonomous mobile node system connected by wireless connections without a static infrastructure. The AODV (Ad hoc on-demand distance vector) routing protocol is designed for ad hoc networks' mobile

M. Ilango (✉)
Hindusthan College of Arts and Science, Coimbatore, India
e-mail: ilango_kn@yahoo.co.in

A. V. Senthil Kumar
Department of MCA, Hindusthan College of Arts and Science, Coimbatore, India

B. Kumar
Department of MCA, Rajarshi Shahu College of Engineering, Pimpri-Chinchwad, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_52

nodes. It has a low head-to-head processing memory and low network utilization, and it determines unicast routes to destinations inside the ad hoc network. Using destination sequence numbers to maintain loop independence at all times, it prevents issues associated with the classic distance vector protocol. The AODV algorithm allows for complex self-starting, multi-hop routing between mobile nodes that want to establish and manage an ad hoc network.

AODV makes it easy for mobile nodes to obtain new routes and reduces the need for nodes to handle routes to destinations that are not in active communication. AODV enables mobile nodes to adapt quickly to network outages and changes in network topology. Meta-heuristics describes a statistical approach that optimizes the problem to improve the solution of the problem in MANETs.

The network of correlations is stable and reliable not just for image processing, but also for the self-organization of the network [6, 7]. MH-FAORM method is to compare various parameters in order to find an optimal path between multiple paths in the network, minimize residual error in the parameter and delay of the packet. Co-efficient values have an ideal route in the network for excellent predictions.

As demonstrated in Fig. 1, the data packet transmission of meta-heuristics firefly technique is carried out through the intermediate nodes.

Meta-heuristics in the firefly algorithm were designed to help you find the best structured route in MANETs. Fireflies are winged beetles or insects that produce light and blinking at night. The light has no infrared or an ultraviolet frequency which is chemically produced from the lower abdomen is called bioluminescence. They use the flash light especially to attract mates or prey. The flashlight was often used as an alarm to keep fireflies away from possible predators.

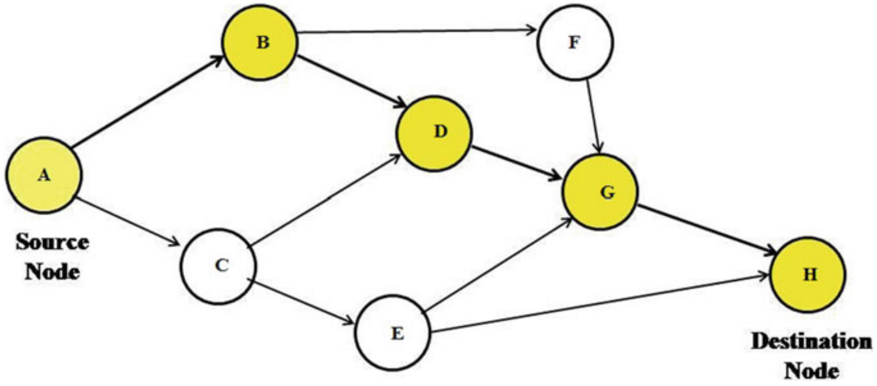


Fig. 1 Firefly MANETs structure

2 Related Works

Jung et al. [8] developed the CLAR protocol to provide a more efficient load transfer metric, as well as a compilation of the light load path and a reduction of the congested node in the high node network. However, high node traffic in the network is managed by the CLAR protocol. Islam and Shaikh [18] have developed a cache replacement technique for MANET, which could be tested by different routing protocols. The memory space of the packets has been limited. Zhang [2] developed cluster-based routing algorithms with high performance aggregation of node data. However, the association of spatial data to minimize energy efficiency was not adequate.

Kim's [21] development of a meta-heuristic approach to adaptive multi-path routing algorithm was a valuable tool for seeking an efficient solution to the problem of ad hoc mobile network routing conflict. Algarni and Almogrem [1] proposed an energy-efficient MANET multicasting protocol based on a busy tone approach to reduce the overhead power caused by control packets during error recovery. The correlated data routing in UWSN, Xu and Liu [16] was identified a multi-population firefly algorithm, optimization of the routing path, taking into account data similarities and sampling rates in the various nodes. However, data packets cross from source to destination with default values such as the correlation coefficient, packet size ad TTL. It was then difficult to produce the optimized path.

The suggested spatial time correlation routing strategy, Sushma and Manjula Devi, was planned to minimize energy requirements in wireless sensor networks and eliminate data packet redundancy [22]. Chaubey et al. [17] have developed a performance study of the TSDRP and AODV protocols for the exploration of stable routing in MANET. However, the size of the mobile node network is difficult to manage. In a data aggregation wireless sensor network, Zeydan et al. [5] suggested energy efficient routing data correlation and specified a routing route to the sink node for energy maximization problems. It does, however, discuss the problem of efficient energy minimization. Approach can provide applicable cost indicators such as minimizing wait time and optimizing network life and throughput.

Mahmood et al. [23] suggested a hierarchical AODV routing protocol to minimize network bandwidth burden and easily detect connection breakage. Devika and Sudha [4] have been designing a fisheye state routing protocol that has proven to be the best for many parameters, such as packet distribution ratio, energy consumption and throughput. However, the life of network was lower. Heuristic approach identified the multiple redundant pathways between source and destination. However, the correlation between redundant routes guarantees the secure transmission of optimum route data in MANET [19].

Reverse AODV routing protocol is intended to reduce overhead routing in a network. However, a reliable factor value was used to evaluate the secondary direction based on the descending order [20]. According to Ciullo et al. [3], correlated node motions have a significant impact on throughput, and latency leads to increased production when nodes travel independently. However, the separate node is part of the destination's super cluster, which adds to the wait. The PLEER method is designed with probabilistic re-broadcasting of energy-efficient routing based on a neighbor node of coverage [10]. To find the best route between all available data packet transmission routes, the deterministic multicast routing mechanism estimates the reliability and resources of mobile nodes [9].

The stepwise regression and binomial logistic regression technique is designed to identify the optimal network link based on the calculation of the regression coefficient value [11, 15]. It reduces data transmission energy and bandwidth consumption in MANETs [13]. The shortest and multi-path technique is designed to select the best and shortest path between nodes to reach from source to destination over all paths [12, 14].

3 Methodology

3.1 Problem Identification and Definition

The suggested MH-FAORM approaches are meant to mitigate current issues. The goals of the MH-FAORM are

1. Identifying a neighbor node in MANET is required to fix the broken problem.
2. Reduce overhead and wait time using TTL (time to live).
3. Identify a categorical node using a meta-heuristics algorithm with a correlation coefficient.

BLR-OR technique has been introduced in the current method, certain disadvantages are noted, such as

1. Non-recently used routing energy has expired.
2. If the routing is disabled, the neighbor node cannot be alerted.

3.2 Meta-Heuristics—Fireflies Correlated Algorithm for Optimized Routing in MANETs

Meta-heuristics firefly correlated strategy is used to find the best path in the network. In this procedure, the value of the correlation coefficient is determined for all the nodes in the network.

Figure 2 demonstrates the increased efficiency of resource effective routing in the MH-FAORM methodology using a meta-heuristic firefly correlated technique. This technique uses the energy, bandwidth and delay time of the node to determine the value of the correlation coefficient. The best route in the network has been established using this method.

The MH-FAORM technique considers the following resource R variables, such as residual power, residual bandwidth and time delay, in order to achieve resource-conscious routing in MANETs. In MANETs, the availability of mobile node resources for N_i is determined as;

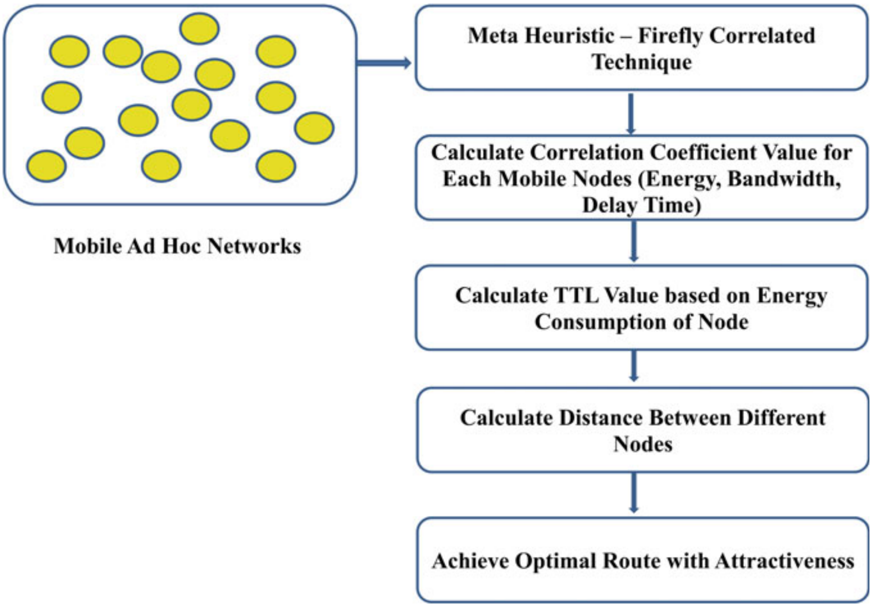

Mobile Ad Hoc Networks
Meta Heuristic – Firefly Correlated Technique
Calculate Correlation Coefficient Value for Each Mobile Nodes (Energy, Bandwidth, Delay Time)
Calculate TTL Value based on Energy Consumption of Node
Calculate Distance Between Different Nodes
Achieve Optimal Route with Attractiveness

Fig. 2 Architecture diagram of meta-heuristic—firefly correlated technique

$$R_{N_i} = R\text{Energy}_{N_i} + \text{RB}_{N_i} + \text{DT} \tag{1}$$

From Eq. (1), $R\text{Energy}_{Ni}$ and RB_{Ni} represent the node's residual energy and bandwidth after data transfer, respectively, while DT represents the delay time between two nodes. Power, bandwidth and delay time [11] are the mathematical approach for measuring these resource variables.

In MANETs, data packets are routed from source to destination in this manner, consuming less resources and bandwidth. As a result, the availability of services and the number of available neighbor nodes are easily determined. Finally, in MANETs, a neighboring mobile node with a small distance and high resource availability is selected as an optimal node for routing data packets between source and destination nodes. For effective data packet transmission in MANETs, optimal mobile nodes use the least amount of storage, bandwidth, and delay time.

To calculate the available number of neighboring nodes with a minimal distance using the MH-FAORM technique and a correlation coefficient.

From Eq. (1), the correlation measures the degree and direction of which the two nodes are associated. It does not match a line across data packets. A correlation coefficient that says how much one node appears to shift as the other node does.

Correlated data packets can be mixed to reduce transmission demand, which reduces the network's overall energy consumption. Assume P_i is the data packet generated at Node_i.

$$\Pi(\text{node}_{\text{dis}}) = \min(|S_i|\forall|\text{Node}(N_i)|) \tag{2}$$

where S_i denotes the root node and Ni denotes that neighbor node of all nodes in network.

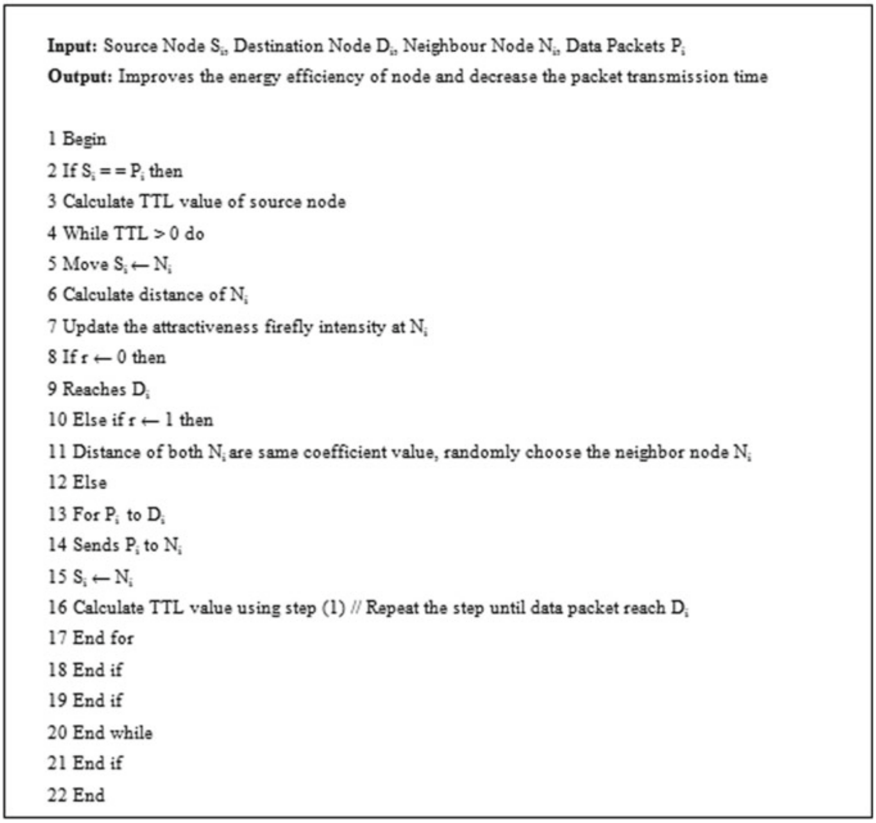
Firefly algorithm has two main functions, power of fluorescence and attractiveness. In this proposed strategy, the attraction of seeking distances between different nodes was changed in MANETs.

From Eq. (2), attractiveness which is the brightness of firefly i on the firefly j in dependent on the degree of brightness of firefly i and the distance $\text{route}_{\text{SiDi}}$ between firefly of S_i and firefly node (N_i).

$$\beta(\text{route}) = \beta_0 e^{-\gamma r} \tag{3}$$

The distance between the source and destination nodes is called the path. The beauty (brightness) of the ideal node was demonstrated using fluorescence intensity.

Algorithm - 1



Input: Source Node S_i, Destination Node D_i, Neighbour Node N_i, Data Packets P_i

Output: Improves the energy efficiency of node and decrease the packet transmission time

```
1 Begin
2 If $S_i$ = = $P_i$ then
3 Calculate TTL value of source node
4 While TTL > 0 do
5 Move $S_i \leftarrow N_i$
6 Calculate distance of $N_i$
7 Update the attractiveness firefly intensity at $N_i$
8 If r ← 0 then
9 Reaches $D_i$
10 Else if r ← 1 then
11 Distance of both $N_i$ are same coefficient value, randomly choose the neighbor node $N_i$
12 Else
13 For $P_i$ to $D_i$
14 Sends $P_i$ to $N_i$
15 $S_i \leftarrow N_i$
16 Calculate TTL value using step (1) // Repeat the step until data packet reach $D_i$
17 End for
18 End if
19 End if
20 End while
21 End if
22 End
```

All packets at the transmitting node should have a limited lifetime, which is governed by TTL information. The information value is stored in the routing table since the source node produces the P_i packet. Centered on the energy of the node used to measure the value of TTL [11]. The TTL value of S_i equals 1, S_i (P_i) shifts between one node to another.

4 Experimental Results

The NS-2 evaluation device with a scale of 1200*1200 m is used to assess the MH-FAORM method. For the re-enactment, 500 portable hubs were selected. To determine the viability of our work, the consequences of the MH-FAORM method is compared to the methodology suggested BLR-OR (binomial logistic regression resource optimized routing) [15].

4.1 Energy Consumption

The E-C is measured using the energy used by a single portable hub for all versatile outside hubs in MANETs in the MH-FAORM approach. The rate of energy consumption is measured in Joules (J) and expressed as,

$$\text{Energy Consumption} = \frac{\text{Energy}_{\text{DP}}}{\text{Total}_{\text{DP}}} \tag{4}$$

In the above Eq. (4), “$\text{Energy}_{\text{DP}}$” is a routing system. The proportion of power inspired by a single packet compared to the overall resources used by all “Total_{DP}” packets in the network is referred to as E-C.

The measurement and effect of the E-C for packet communication dependent on various portable hubs in the 50–500 range are seen in Table 1 and Fig. 3. As opposed to current methods, the E-C of the proposed MH-FAORM solution is lower, as shown in Table 1 and Fig. 3. Similarly, as these methods were used to maximize the number of hubs, the energy levels also increased.

4.2 Packet Delivery Ratio

$\text{Data}_{\text{P-D-R}}$ is the proportion of the number of packets needed by the aim to the total number of packets sent in the MH-FAORM method. In terms of rate (percent), the $\text{Data}_{\text{P-D-R}}$ is quantified and designed as follows.

Table 1 Tabulation of energy consumption

Number of nodes	Energy consumption (J)	
	BLR OR	MH-FAORM
50	0.06	0.05
100	0.08	0.07
150	0.08	0.06
200	0.10	0.09
250	0.13	0.11
300	0.14	0.12
350	0.17	0.15
400	0.19	0.17
450	0.22	0.20
500	0.23	0.21

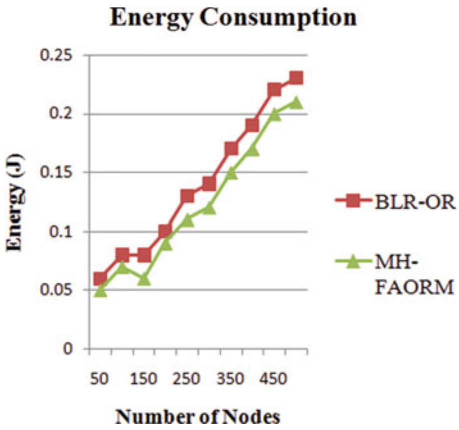


Fig. 3 Energy consumption

$$\text{Packet Delivery Ratio} = \frac{\text{Number of data packets received}}{\text{Total number of data packets sent}} \times 100 \quad (5)$$

From the above state (10), where the Data$_{\text{P-D-R}}$ is higher, the technique is said to be becoming increasingly competent.

The effect of Data$_{\text{P-D-R}}$ based on separate packets in the 90–900 range as shown in Table 2 and Fig. 4. The Data$_{\text{P-D-R}}$ using the proposed structures is higher as compared to the existing approach, as shown in Table 2 and Fig. 4.

Table 2 Packet delivery ratio

Number of Packets	Packet delivery ratio	
	BLR-OR	MH-FAORM
9	81.65	81.85
18	81.91	82.06
27	82.12	82.58
36	82.73	82.84
45	83.04	83.14
54	83.81	83.92
63	84.07	84.27
72	84.52	84.69
81	84.86	84.90
90	85.23	85.31

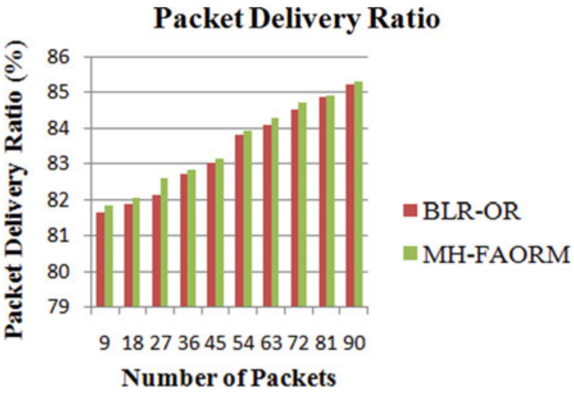


Fig. 4 Packet delivery ratio

5 Conclusion

The MH-FAORM technique was developed to help in the discovery and optimization of routing paths by considering the data correlation coefficient values in different nodes. The fluorescence attractiveness eliminates the maximum value of a neighbor node to enhance the routing performance. For effective transmission, the packet distribution ratio with coefficient in the AODV protocol is proportional to the TTL value. The experimental evaluation of MH-FAORM provides better resource efficient routing with better results in the following parameters such as energy and packet delivery ratio. In the future, the route breaks prediction in the proposed quality aware routing method. A route break prediction scheme helps to give a quick response of the method to route break and reduces the resources of routing.

References

1. B.H. Algarni, A.S. Almogren, Reliable and energy efficient protocol for MANET multicasting. J. Comput. Netw. Commun. (2016)
2. C. Zhang, Cluster-based routing algorithm using spatial data correlation for wireless sensor networks. J. Commun. **5**(3), 232–238 (2010)
3. D. Ciullo, V. Martina, M. Garetto, Impact of correlated mobility on delay-throughput performance in mobile Ad Hoc networks. ACM Trans. Netw. **19**(6), 1745–1758 (2011)
4. B. Devika, P.N. Sudha, Fisheye state protocol in correlation with power consumption in Ad Hoc networks. Int. J. Appl. Eng. Res. **13**(17), 13335–13339 (2018)
5. E. Zeydan, D. Kivanc, C. Comanicill, U. Tureli, Energy efficient routing for correlated data in wireless sensor networks. Ad Hoc Networks (2012)
6. G.D. Praveenkumar, M. Dharmalingam, Pruned cascade neural network image classification. Int. J. Recent. Technol. Eng. **8**(3), 6454–6457 (2019)
7. G.D. Praveenkumar, M. Dharmalingam, Recurrent cascade neural network for image classification. Int. J. Sci. Technol. Res. **8**(10), 1009–1012 (2019)
8. J.W. Jung, D. Choi, K. Kwo, I. Chong, K. Lim, H. K. Kahng, A correlated load aware routing protocol in mobile Ad Hoc networks (Springer-Verlag Berlin Heidelberg, 2004), pp. 227–236
9. M. Ilango, A. V. Senthil Kumar, Deterministic multicast link based energy optimized routing in MANET, in *2017 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Vol. 3 (2017), pp. 1102–1110
10. M. Ilango, A.V. Senthil Kumar, Probabilistic and link based energy efficient routing in MANET. Int J Comput Trends Technol (IJCTT) **38**(1), 38–45 (2016)
11. M. Ilango, A. V. Senthilkumar stepwise regression based resource optimized routing in mobile Ad Hoc network. Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) **6**(2), 504–514(2018)
12. M. Ilango, A.V. Senthil Kumar, Multipath strategies and link based resource optimized routing in mobile Ad Hoc networks. Int. J. Recent Technol. Eng (IJRTE) **8**(5)
13. M. Ilango, A.V. Senthil Kumar, Non linear differential optimization for quality aware resource efficient routing in mobile Ad Hoc networks. Int. J. Eng. Adv. Technol. (IJEAT) **9**(1) (2019)
14. M. Ilango, A.V. Senthil Kumar, Resource optimized routing using shortestpath technique in MANET. J. Xidian Univer. **14**(5) (2020)
15. M. Ilango, A.V. Senthil Kumar, Binomial logistic regression resource optimized routing in MANET. IECEMSN, Springer Lecture Notes Data Eng. Commun. Technol. 384–392 (2020)
16. M. Xu, G. Liu, A multipopulation firefly algorithm for correlated data routing in underwater wireless sensor networks. Int. J. Distrib. Sensor Netw. (2013)
17. N. Chaubey, A, Aggarwal, S. Gandhi, K.A. Jani, Performance analysis of TSDRP and AODV routing protocol under black hole attacks in MANETS by varying network size, in *International Conferrence on Advanced Computing and Communication Technologies* (2015)
18. N. Islam, Z.A. Shaikh, Exploiting correlation among data items for cache replacement in Ad Hoc networks, in *International Conference on Information Management and Engineering* (2010)
19. P. Papadimitrators, Z.J. Haas, E.G. Sirer, Path set selection in mobile Ad Hoc networks, in *ACM International Symposium on Mobile Ad Hoc Networking* (2002), pp. 1–11
20. S.K. Mohapatra, B.R. Swain, S.K. Mohapatra, S.K. Behera, Stability and energy aware reverse AODV routing protocol in MANET, in *International Conference on Recent Trends in Information System* (2015) pp 526–531
21. S. Kim, Adaptive MANET multipath routing algorithm based on the simulated annealing approach. Sci. World J. 1–8 (2014)
22. K.M. Sushma, T.H. Manjula Devi, A spatio-temporal correlation based routing technique for wireless sensor networks. Int. J. Eng. Adv. Technol. (IJEAT) **4**(5), 86–89 (2015)
23. Z.S. Mahmood, A.H.A. Hashem, S.A. Hameed, F. Anwar, The diretional hierarchical AODV routing protocol for wireless mesh networks, in *International Conference on Computing, Control, networking, Electronics and Embedded Systems Engineering* (2015)

Machine Transfer Learning Deep Softmax Regression Neural Network for Image Classification

G. D. Praveenkumar and Dharmalingam Muthusamy

Abstract Image classification is a major topic in image processing. Conventional algorithm had the major drawback of vanishing gradient problem with stochastic descent algorithm; to overcome this problem, a new approach called deep Softmax regression classifier neural network (DSRCNN) was developed. The proposed DSRCNN technique developed layer-wise learning that is applied to restricted Boltzmann machine for accurate image classification. Restricted Boltzmann machine contains visible unit and hidden unit to carry the image classification on CIFAR-10 and CIFAR-100 dataset. The datasets of each input image is preprocessed to reduce the noise then that are fed to the visible unit. The visible and hidden units are trained by the extracted images and forward into Softmax regression to accurately classifying an input image into multiple classes. The simulation of the proposed, DSRCNN technique metrics' accuracy and false-positive ratio performance are compared with the other state-of-the-art methods.

Keywords Restricted Boltzmann machine · Image features extraction · Contrastive divergence · Logistic regression · Image classification

1 Introduction

The artificial neural network consists of many processing units which, according to a certain topology, are interconnected to carry out a classification task [1]. Research datasets on image classification are generally very broad. Image classification is intended to classify a group of images into a given class. There are two main methods of classification: the feature-based approach and the appearance-based method. A feature extraction-based method was used in this proposed DSRCNN method. The model of classification tasks is trained to recognize different image classes [2]. It is difficult to train the neural network in deep Boltzmann machine (DBM), and it takes more computing time and hard to train an algorithm and also suffer from a gradient problem that disappears. DBM has a collection of completely connectable

G. D. Praveenkumar (✉) · D. Muthusamy
Government Arts and Science College—Modakkurichi, Erode, Tamil Nadu 638104, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_53

visible units and hidden units, so training is too difficult. The key downside this paper focuses on is the issue of the vanishing gradient and the difficulty to computational time.

In the DBM architecture, there is no restriction to linking fully visible units and hidden units. We introduced the deep restricted Boltzmann machine, which is being used as an RBM [3, 4] for training each layer. The key factor used in the construction of deep networks is the RBM machine. RBM is commonly used in supervised and unsupervised machine learning applications such as feature learning, topic learning, and collaborative filter classification. The RBM architecture supervised learning method with a trained weight as the basis weight of the neural network is accessed to the gradient-based supervised method [5]. The restriction of RBM enables inference and learning to be more effective as a hidden unit, conditionally independent of the visible unit [6].

Softmax regression is an effective image recognition algorithm for deep learning. Softmax regression is often referred to as multiclass label regressions, which are generalized from the logistic approach to solve the problems of image classification. Softmax regression using a logistic function that generated an S-shaped curve in the range of 0 or 1, making it possible to obtain a nonlinear boundary. Softmax regression is a type of regression used based on image classification to evaluate the relationship between the variable input and output. Usually, a trained neural network uses a backpropagation algorithm to find the output gradient [2, 7]. Each of the weight matrixes is part of the neural network to increase the weight matrix and performance.

The network is starting to learn very slowly and gradually stopping learning is called a gradient problem. The restricted Boltzmann machine neural network is used to overcome the vanishing gradient problem in the proposed DSRCNN technique. It takes a longer time for the neural network to train and learn from the data. We concentrate on multiclass labels using the Softmax regression model in this paper to solve the issue of existing conventional methods. The main contributions of the proposed DSRCNN method are as follows: Enhance the powerful image feature extraction such as shape, color, texture, size to use a profoundly restricted Boltzmann machine neural network. Using the Softmax regression classifier to overcome the multiclass image classification to increase the accuracy performance and reduce the false-positive rate.

The rest of the article is structured as follows: An outline of the related work is given in Sect. 2. Next, Sect. 3 presents the deep Softmax regression classifier of the neural network methodology. Section 4 explains the simulation platform discussed in our research. Subsequently, the analysis of the results is summarized in Sect 5. Finally, with a conclusion in Sect. 6.

2 Related Work

Marque et al. [8] suggested a cascade learning model to build a bottom-to-top-level method, minimize the gradient problem of VGG implementation, and monitor the backpropagation process for network training. VGG network computational, however, is costly. Praveenkumar et al. [2, 7, 9, 10] suggested pruned cascade, recurrent cascade, and hierarchical cascade Bayesian inference both of these models are developed by the GoogleNet architecture to reduce the gradient problem in the deep cascade methodology and reduce time complexity and space complexity. However, the training time complexity of the approach has not been reduced. Yuan et al. [11] has developed multicriteria models to learn an adaptive learning technique for deep learning models in the classification of images. The multicriteria that raise the computational time and the continuous training imbalance manually set as hyperparameter and accuracy was low.

Ilango et al. [12] suggested a binomial logistic regression approach in MANET to suppress linear and stepwise regression issues and ensure network efficiency. Logistic regression is supported by image mining and computer networks. However, access to image classification in logistic regression and binomial regression to train an image has the toughest challenge, so we use multinomial. Janmaijaya et al. [13] created a fuzzy RBM with two fuzzy sets to train the MNIST dataset. The extraction process is not capable of satisfying the RBM needs. In any case, the yield of the boundary is expected to improve the layer to layer learning measure. The class incremental learning method [14] is trained with CNN by the Softmax layer used to determine the feature representations on the network. However, CNN is trained as an overall image batch on various datasets, but each extraction function performs on the Softmax classifier to change the learning parameter sequentially. However, the image function on the training and testing dataset was extracted to solve this problem [15].

The Softmax classifier [16] has been taught to detect an effective image classification. The significance of the characteristic extract of each heart-beat section is high. Dimension vector values used more time taken to train the cardiac beat segmentation and classifying the result. Rakkimuthu et al. [17] helped the SUDMLA data collection for finger vein authentication that has been built in a completely recurrent deep learning image classification for deep learning. However, precision, time complexity, and false-positive ratios are not reduced. DGA algorithm is working with text classification to detect malicious domain network [18], but proposed algorithms are difficult to predict and classify the image datasets. The cognitive image classification-based visual saliency guided model (VSGM) gives better image retrieval from the database. The image features extraction is not considered in the image detection and classification.

3 Deep Softmax Regression Classifier Neural Network Technique

The deep Softmax regression classifier neural network (DSRCNN) methodology uses deep restricted Boltzmann machines (RBM) learning to perform image classification. The dataset image performs preprocessing on each image to remove the unnecessary

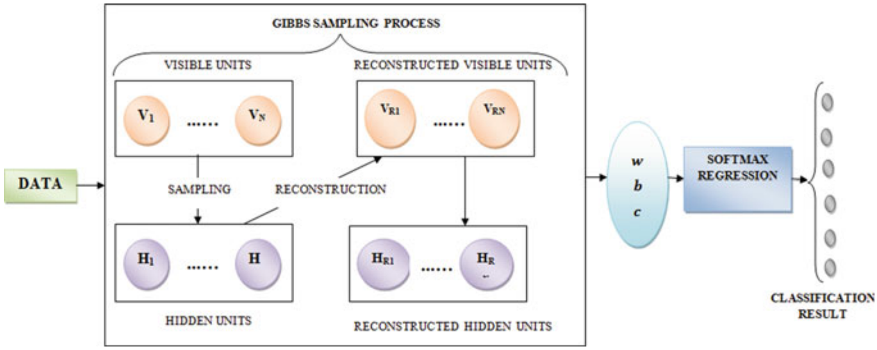


Fig. 1 Proposed system architecture

noise. The preprocessing images are then added to the visible unit and hidden units to extract features like shape, color, texture, and size. Finally, to categorize the input images into multiple classes, the Softmax regression function is performed. Figure 1 describes the proposed DSRNN techniques.

3.1 *Deep Restricted Boltzmann Machine (Contrastive Divergence) with Gibbs Sampling*

The deep restricted Boltzmann machines are designed by running the Monte Carlo Markov chain to converge and using the Gibbs sampling process as chain transformation operators. A deep restricted Boltzmann machine includes visible and hidden layers of *p* and *q*. The ɵ parameter contains the weight matrix *W* (*p*, *q*). The energy function *E* (*p*, *q*) is defined as following given below equation.

$$E(p, q|\theta) = 0.5\big(P^T W q + b^T + c^T q\big) \tag{1}$$

The partition function also called the normalizing factor *K* (ɵ) is defined as

$$K(w) = \sum w_e[-E(x|\ominus)] \tag{2}$$

The probability function is defined as

$$Pb(p, q|\ominus) = \frac{1}{\mathrm{K}(\ominus)} w_e[-E(p, q|\ominus)] \tag{3}$$

The randomly choose a visible layer data values conditional probability is defined as

$$Pb\left(p_i = \frac{1}{q}\right) = S\left(b_i + w_{(pq)i}.q\right) \tag{4}$$

where the visible layer bias is $b = b1, b2 \ldots . b_m$ and weight is $w_{(p,q)i/j}$.

The hidden layer training dataset of each image in hidden unit probability is defined as

$$\mathrm{Pb}\left(q_i = \frac{1}{p}\right) = S\left(c_j + w_{(\mathrm{pq})j}.p\right) \tag{5}$$

where S (logistic function) is defined as

$$S(x) = \frac{1}{1 + w_e(-x)} \tag{6}$$

The hidden layers remain unchanged binary unit, but visible layer unit are replaced by Gaussian linear units. The conditional probability of reconstructed in visible units is defined as

$$\mathrm{Pb}_{\text{ reconstruct}}\left(\mathrm{pi} = \frac{1}{q_{\text{ reconstruct}}}\right) = S\left(b_{i \text{ reconstruct}} + w_{(\mathrm{pq})i \text{ reconstruct}} \cdot q_{\text{ reconstruct}}\right) \tag{7}$$

The objective of the deep restricted Boltzmann machines is to recreate the visible image of the device as far as possible. The reconstruction image is modified based on weight and bias and then used to enable hidden layers. The fast learning algorithm contrastive divergence (CD) method updates the parameter value defined as

$$\Delta w_{ij} = \eta\left(\left(p_i q_j\right)_{D1,D2}\right) - \left(p_i q_j\right)\text{reconstruct} \tag{8}$$

$$\Delta b_j = \eta\left(\left(q_j\right)_{D1,D2}\right) - \left(q_j\right)\text{reconstruct} \tag{9}$$

where η is the learning rate from the above equation is used to update the parameter value during the training time. The CIFAR-10 (D_1) and CIFAR-100 (D_2) dataset input images are sent to the visible unit. Using the deep restricted Boltzmann machines, the visible unit Eq. (4) performs corresponding feature vector values are fed to hidden units depending on the weight and bias of the network. Each input image is interpreted as a given mathematical equation.

$$p(t) = w_{pq}.I_i(t) + b \tag{10}$$

where p (t) indicates the behavior of the neurons in the visible unit at the moment t, input neurons are denoted as p, hidden neurons are denoted as q, $I_i(t)$ blends input images, b is denoted as bias, w_{pq} is denoted as weight. Hidden units are extracting of features such as structure, color, texture, and size. At this point, the activation from the hidden layer is modified based on w and b, and then sent back to the input layer for activation. From Eqs. (5, 6), the hidden unit builds a multilayer structure to efficiently train the dataset. Training dataset vectors measure the likelihood of hidden unit activation values performing a learning rate of 0.01 over 200 epochs. From Eqs. (7, 8, 9), the subsequent reconstruction of the hidden activation uses the Gibbs sampling function and updates the gradient descent weight.

The unit effectively trains each input image on a feature detector to find out the key features shape, color, texture, and size from an input image. The hidden unit to train on the image feature representation of a simple forward network and learn what each model looks like. These phases are called models of training (learning). The training dataset is analyzed by the deep restricted Boltzmann machine applied to the Softmax regression function to solve the problem of image classification. The Softmax activation function leads to the categorization of various groups with logistic regression. Softmax regression is used to define multiple type images relative to evaluating dataset images as follows.

$$\text{soft}(t) = \frac{1}{1 + e^{t}} \tag{11}$$

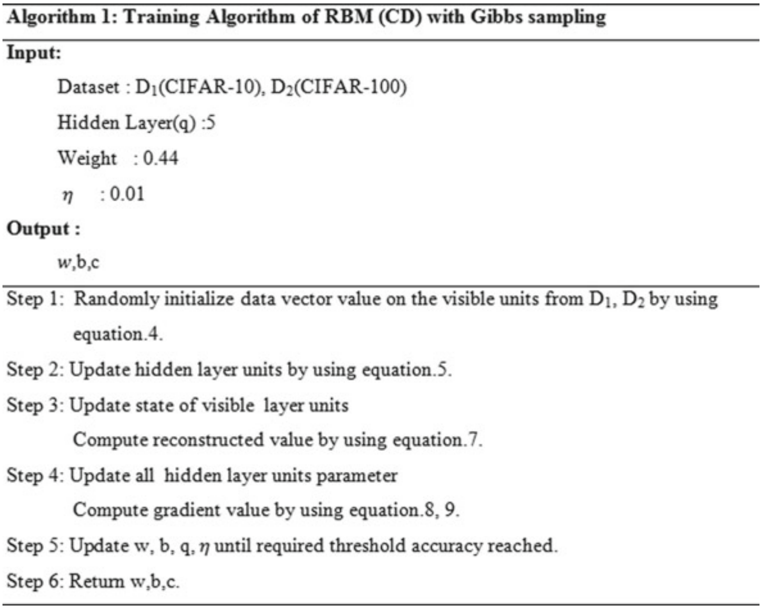
From above the mathematical Eq. (11), the values of the coefficients to be understood by the pixels are significant as calculated in the multiclass. The vector coefficient values are graded as positive and negative. Positive coefficient pixels are positively correlated with the class and the remaining classes are represented as a misclassified image or negative coefficient. Positive coefficient classes that classify the potential of each class to categorize all the output class image numbers are grouped within the particular class. The process of the DSRCNN technique is stopped. Misclassified image to measure log loss as follows

$$\log\text{loss} = \sum_{i=1}^{n} I i p_i(t)/a_i(t) \tag{12}$$

From above the mathematical Eq. (12), it shows the sum of error between the specified $p_i(t)$ and the specified $a_i(t)$ as the real value. The misclassified image in the test dataset is retrained to minimize the residual error and introduce a new hidden layer to reduce the log loss feature, eventually, the connection W (p, q) is involved with the aid of the gradient descent to determine the derivative of all the training results. Therefore, the DSRCNN technique offers improved image classification efficiency in terms of accuracy and false-positive ratio to add new hidden units; its input weight is kept constant, and all the output weights are once again trained with the spike backpropagation and reduce the gradient problem. The values of spike backpropagation using a residual connection to allow extract image features and the classification model quality to be extracted from the network layer.

3.2 Deep Restricted Boltzmann Machine Training Algorithm

In this section, the proposed restricted Boltzmann machine training algorithm is represented using the contrast divergence process. The algorithm will initially take a ‘*n*’ number of input images from D_1, D_2 dataset. Before that, all input images are preprocessed, (some sample simulation of preprocessing D_1, D_2 dataset) shown in Fig. 3 then fed into RBM layer. Hinton et al. developed an efficient neural network learning algorithm by greedily training each layer as an RBM using the previous layer activation as input follows, and the visible units randomly assign the input image from the dataset and starts learning [19–22]. The algorithm 1 takes input from the learning rate and computed by the layer by layer to update the weight by descent. The training follows layers by layers approach. The input vector values are passed through the hidden units extracting the image features such as shape, size, color, texture, and multilayer extraction to be extracted as a function of visible units shown in Fig. 4. Based on the extracted features, all units developed by a binary unit, but visible layers are reconstructed by the Gaussian linear units shown in Fig. 1. Recreated visible unit probability values are evaluated by using Eq. (7), and the gradient value of the different parameter is determined until the training algorithm is more accurate. The flow process is in described in Fig. 2.



```
Algorithm 1: Training Algorithm of RBM (CD) with Gibbs sampling
Input:
        Dataset : $D_1$(CIFAR-10), $D_2$(CIFAR-100)
        Hidden Layer(q) :5
        Weight  : 0.44
        $\eta$    : 0.01
Output :
        w,b,c
Step 1:  Randomly initialize data vector value on the visible units from $D_1$, $D_2$ by using
           equation.4.
Step 2: Update hidden layer units by using equation.5.
Step 3: Update state of visible  layer units
           Compute reconstructed value by using equation.7.
Step 4: Update all  hidden layer units parameter
           Compute gradient value by using equation.8, 9.
Step 5: Update w, b, q, $\eta$ until required threshold accuracy reached.
Step 6: Return w,b,c.
```

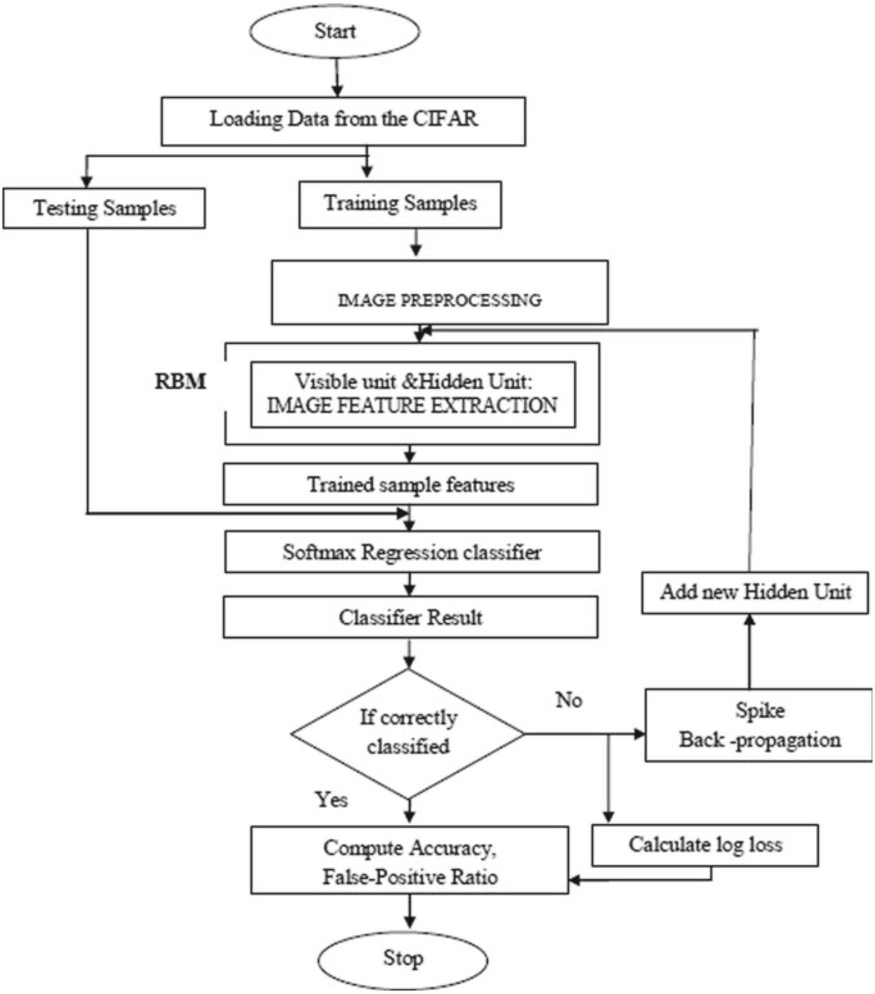


Fig. 2 Flow chart of Deep Softmax Regression Cascade Neural Network (DSRCNN)

4 Simulation Platform and Image Processing

The experiments have been executed on MATLAB (2019 a) with Windows 7/8(64 bit) and the hardware requirements are Intel®-Core™ i7-2670 QM GHz Processor, 8 GB Ram, Open Gl3.3 support graphics card with 1 GB GPU memory running on Microsoft platform. Experiments carried out in benchmark datasets CIFAR-10 and CIFAR-100 [23]. The CIFAR-10 dataset contains 60,000 images with 32*32 pixels. In CIFAR-10 dataset, there are 50,000 training images and 10,000 test images; CIFAR-100 dataset contains 100 classes with 600 images each. The DSRCNN technique considers 1000–5000 images from both the CIFAR-10 and CIFAR-100

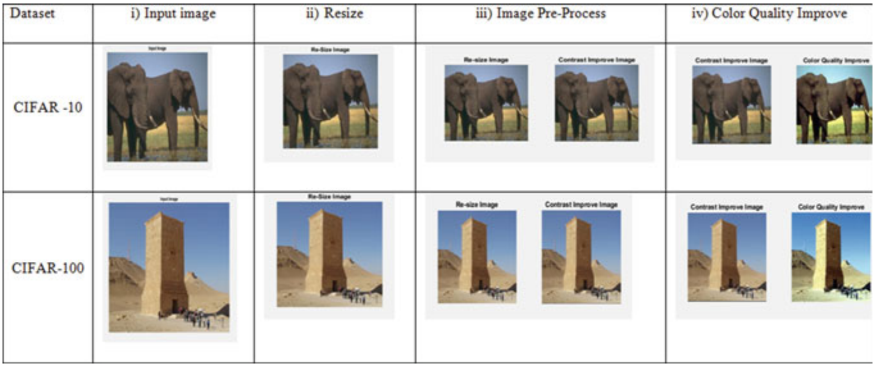


Fig. 3 Sample simulation result of image preprocessing on CIFAR 10 and CIFAR 100 dataset

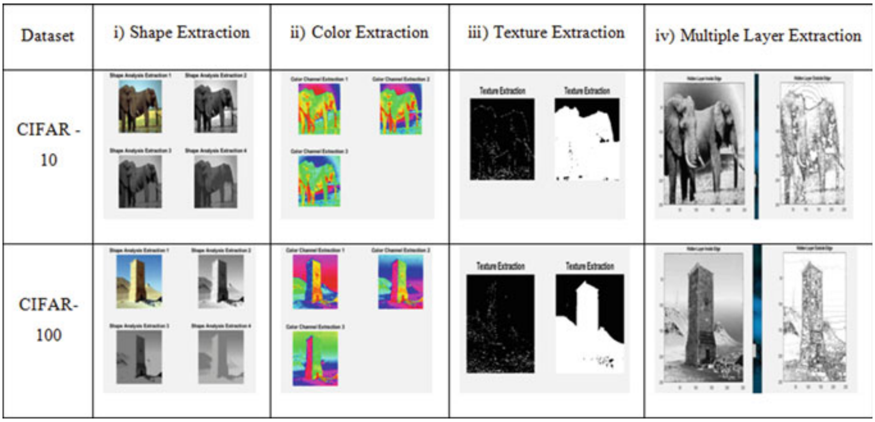

Dataset
i) Shape Extraction
ii) Color Extraction
iii) Texture Extraction
iv) Multiple Layer Extraction
CIFAR - 10
CIFAR- 100
Shape Analysis Extraction 1
Shape Analysis Extraction 2
Shape Analysis Extraction 3
Shape Analysis Extraction 4
Color Channel Extraction 1
Color Channel Extraction 2
Color Channel Extraction 3
Texture Extraction

Fig. 4 Sample simulation result of RBM extracts image features on CIFAR 10 and 100 dataset

database. The performance of the DSRCNN method is estimated in terms of accuracy, false-positive ratio and compared to MCADL, SDBCNL techniques. The flow chart described in Fig. 2 of the proposed DSRCNN methods. The dataset is split into training and testing samples. Gradient descent is used to set the learning rate of 0.01, number of hidden units 5, the weight is initialized to 0.44 and adjusted using the learning rate, all models are trained for 200 epochs within the range of network training.

5 Result Analysis

In this section, the comparative result of the DSRCNN method is introduced. The simulation result of the DSRCNN is compared with the existing of MCADL [11], SDBCNL [2] technique.

5.1 Accuracy of DSRCNN

In the DSRCNN method, accuracy (A_{CC}) is measured as ratio of the range of images that are exactly labeled to the complete variety of images. The accuracy is mathematically dedicated as follows

$$A_{\text{CC}} = \frac{Z_{\text{NC}}}{Z_T} * 100 \tag{13}$$

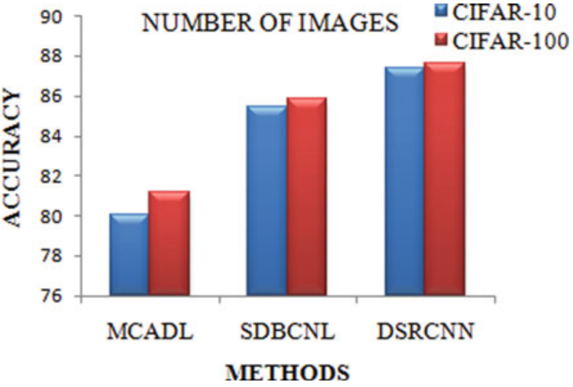
From the above mathematical Eq. (13), Z_{NC} shows the number of images effectively categorized and Z_T refers compute the number of images taken carried timeout the simulation technique. The accuracy of image classification is considered in terms of percentage (%). The accuracy of image classification in DSRCNN proposed method and existing MCADL, SDBCNL technique are implemented in MATLAB (2019a) simulator by taking a different number of images in the range of 1000–5000 from CIFAR-10 and CIFAR-100 dataset. The simulation result evaluation of accuracy for image classification is presented in Table 1.

Figure 5 given above shows the comparison result of accuracy on the classified input image to multiple classes, compared to MCADL and SDBCNL techniques. Deep restricted Boltzmann machine learns image representation, which increases the representational capacity of image and logistic regression in the proposed antithetical model to the existing model, where it uses Softmax regression to precisely

Table 1 Comparison result of MCADL, SDBCNL, DSRCNN methods on CIFAR 10 and CIFAR 100 dataset

Accuracy (%)						
Dataset	Methods	Number of Images				
		1000	2000	3000	4000	5000
CIFAR-10	MCADL	85	83	79	82	76
	SDBCNL	88	85	87	84	83
	DSRCNN	89	90	88	86	84
CIFAR-100	MCADL	85	85	79	81	76
	SDBCNL	89	87	84	85	84
	DSRCNN	91	89	88	86	84

Fig. 5 Performance analysis of MCADL, SDBCNL, DSRCNN



classify the image into different labels. The existing MCADL splits the network into a diverse layer where training is carried out by one layer. During this process, the feature learning utilizes regression to choose the relevant features. Besides, there is no architecture is used for the training process. This leads to degrade the accuracy of classification in conventional works. In order to overcome the issues, the proposed DSRCNN technique is introduced. In contrast to existing works, deep restricted Boltzmann machine is utilized in proposed DSRCNN technique where it utilizes Softmax activation function in the output unit to accurately categorize input images into multiple classes. A Softmax activation function is used to realize the relationship between extracted features (i.e., shape, color, texture, and size of images as input) and pre-stored templates by measuring the exponential value of the correctly classified image. The proposed method to increase the ratio of a variety of images that are correctly labeled as compared to conventional work of MCADL, SDBCNL.

5.2 *False-Positive Ratio of DSRCNN*

In the DSRCNN method, the false-positive ratio calculates the ratio of several images mismatch classified to the total number of images. The false-positive ratio is mathematically dedicated as

$$\text{False-Positive Ratio} = \frac{Z_{\text{IP}}}{Z_T} * 100 \tag{14}$$

From the above mathematical Eq. (14), Z_{IP} denotes the number of images incorrectly predicts the positive class and Z_T refers to the total number of the image determined in terms of percentage. The proposed methods of DSRCNN are taking the different number of images in the range of 1000–5000 from CIRAR-10 and CIFAR-100 dataset. The DSRCNN methods obtain a lower false-positive ratio for effective image classification. When compared to convention work of MCADL, SDBCNL, and the comparative result evaluation of false-positive ratio for image classification exhibits in Table 2.

Figure 6 given above shows the performance of false-positive ratio with various numbers of images in the range from 1000–5000. The spike backpropagation methods help to determine the weight of the network are randomly assigned to train the images in deep restricted Boltzmann machines. This process is repeated until the error is very small for exactly performing the image classification. The proposed DSRCNN model reduces the false-positive ratio of image classification using the CIFAR 10 and CIFAR 100 dataset by reduces the 5% and 3% improvement in the original result when compared to MCADL, SDBCNL technique.

Table 2 Comparison Result of MCADL, SDBCNL, DSRCNN methods on CIFAR 10 and CIFAR 100 Dataset

False-Positive Ratio						
Dataset	Methods	Number of Images				
		1000	2000	3000	4000	5000
CIFAR-10	MCADL	20	19	20	21	24
	SDBCNL	12	15	13	16	17
	DSRCNN	11	14	13	15	16
CIFAR-100	MCADL	20	18	19	21	22
	SDBCNL	18	16	15	15	17
	DSRCNN	15	12	13	12	14

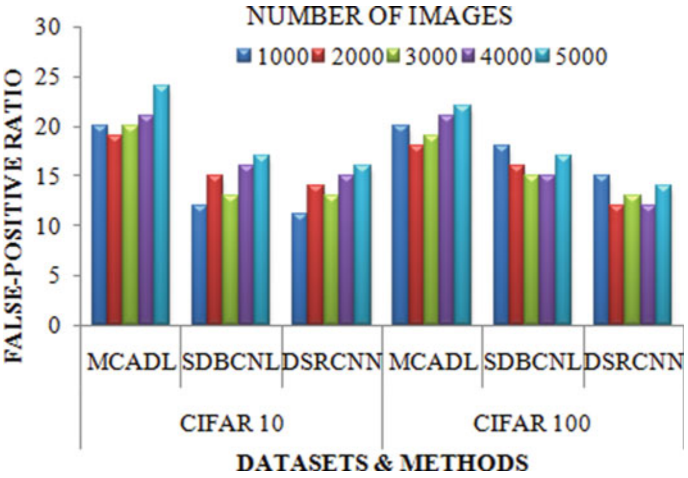


Fig. 6 False-Positive Ratio of MCADL, SDBCNL, DSRCNN

6 Conclusion

In this work, we conclude the deep Softmax regression technique is designed to supervise cascade learning to train the neural network and reduces the vanishing gradient and extend the overall performance of the multiclass Softmax classifier to enhance the ratio of image classification. The DSRCNN performance metrics accuracy, false-positive ratio is processed in various numbers of images from both CIFAR-10 and CIFAR-100 input dataset. The DSRCNN is responsible for addressing the task of extracting features, which is the RBM system trained by extracting features from images. The Softmax regression classifier increases image classification quality. The proposed learning model produces comparatively good results from traditional methods. We are interested in considering different natural database and human emotional images using various learning models in the future.

References

1. R. Amalraj, M. Dharmalingam, A work point system coupled with back-propagation for solving double dummy bridge problem. Neurocomputing **168**, 160–178 (2015)
2. G.D. Praveenkumar, M. Dharmalingam, Softmax deep Boltzmann cascade neural network learning technique for image classificaation. (under review)
3. J.Su,D.B.Thomas,P.Y.K.Cheung, Increasing network size and training throughput of FPGA restricted Boltzmann using dropout, in *Proceeding of IEEE 24th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM)* (2016), pp 48–51
4. L. Zhang, H. Huang, X. Jing, A modified cyclostationary spectrum sensing based on softmax regression model, in *Proceeding of IEEE* (2016), pp. 620–623
5. F.He, N.Li, Restricted Boltzmann machine based on item category for collaborative filtering, in *Proceeding. of International Conference on Computer Technology, Electronics and Communication (ICCTEC)*, pp. 756–760
6. Q. Wang, X. Gao, K. Wan, F. Li, Z. Hu, A Novel restricted Boltzmann machine training algorithm with fast Gibbs sampling policy. Math. Probl. Eng. **7**, 1–19 (2020)
7. G.D. Praveenkumar, M. Dharmalingam, Recurrent cascade neural network for image classification. Int. J. Sci. Technol. Res. **8**, 1009–1012 (2019)
8. E.S. Marquez, J.S. Hare, M. Niranjan, Deep cascade learning. IEEE Trans. Neural Netw. Learn. Syst. **29**, 5475–5485 (2018)
9. G.D. Praveenkumar, M. Dharmalingam, Hierarchical Image Classification on Bayesian Inference with GoogleNet, Innovative Computing and Communication: An International Journal. 11–6
10. G.D. Praveenkumar, M. Dharmalingam, Pruned cascade neural network image classification. Int. J. Recent Technol. Eng. **8**, 6454–6457 (2019)
11. J. Yuan, X. Hou, Y. Xiao, D. Cao, WeiliGuanc LiqiangNie, Multi-criteria active deep learning for image classification. Knowl. Based Syst. **172**, 86–94 (2019)
12. M. Ilango, A.V. Senthil kumar, A. Dutta, Binomial logistic regression resource optimized routing in mobile adhoc network, in *Proceeding of Springer lecture Notes on Data Engineering and Communication Technologies (ICECMSM)* (2020)
13. M. Janmaijaya, A.K. Shukala, T. Seth, P.K. Muhuri, Interval type-2 fuzzy restricted boltzmann machine for the enhancement of deep learning, in *Proceeding Of IEEE International Conference on Fuzzy Systems* (2019)

14. J. Yu, J. Gwak, S. Lee, M. Jeon, An incremental learning approach for Restricted Boltzmann machine, in *Proceeding of International Conference on Control, Automation and Information Sciences (ICCAIS)* (2015), pp. 113–117
15. X. Ye, Q. Zhu, Class-Incremental learning based on feature extraction of CNN with optimized softmax and one-class classifier. IEEE **7**, 42024–42031 (2019)
16. H.E. Saadawy, M. Tantawi, H.A. Shedeed, M.F. Tolba, Electrocardiogram(ECG) heart disease diagonisis using PNN,SVM and softmax regression classifier, in *Proceeding of Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)* (2016), pp 106–109
17. P. Rakkimuthu, M. Dharmalingam, Delta ruled fully recurrent deep learning for finger-vein verification. Int. J. Innov. Technol. Explor. Eng. **9**, 1580–1588 (2019)
18. P. Karunakaran, Deep learning approach to DGA classification for effective cyber security. J. Ubiquit. Comput. Commun. Technol. (UCCT) **2**(04), 203–213 (2020)
19. G. Hinton, T. Sejnowski, Learning and releaming in boltzmann machines. Parallel Distrib. Process.: Explor. Microstruct. Cogn. **1**, 282–317 (1986)
20. G.E. Hinton, Training products of experts by minimizing contrastive divergence. Neural Comput. **14**, 1771–1800 (2002)
21. J. Chu, H. Wang, H. Meng, P. Jin, T. Li, Restricted Boltzmann machines with Gaussian visible units guided by Pairwise Constraints. IEEE Trans. Cybern. **49**(12), 4321–4334 (2019)
22. J. Chu, H. Wang, J. Liu, Z. Gong, T. Li, Unsupervised feature learning architecture with multi-clustering integration RBM. IEEE Trans. Know. Data Eng. 1–14 (2020)
23. CIFAR-10 and CIFAR-100 datasets. https://www.cs.toronto.edu/~kriz/cifar.html

A Profound Deep Learning Approach for Detection System in Network Data

N. Raghavendra Sai, Tirandasu Ravi Kumar, S. Sandeep Kumar, A. Pavan Kumar, and M. Jogendra Kumar

Abstract This research work has attempted to design an interconnection, which approves the sharing of the data correspondence relationship without the requirement of an individual. The construction of the Web of Things has permitted to team up more gadgets without human intervention, wherein the information stack is low, and therefore, the amount of gathered information will be decreased, for example: progressed drawing and different edges. A few hypotheses, for instance, have misrepresented the eye, counterfeit scholastic expertise, and significant characterization on how they need an incredible mien to gain attention to their capacity, and along these lines, the genuine advantages of planning, heterogeneous information from various assessments and different analysts have also been anticipated. Likewise, the TCP/IP show has been introduced in this paper to control the data transmission and the assessment rehearses for gathering. Firstly, this research work has attempted to visualize the inconsistencies in the Web of Things and the advancement in Web progressions are considered from actual articles that are truly recommended as a system with a lightweight and a particularly intriguing approach has been developed to deal with an IoT association. Second, the present occupations of fake instinct are considered by obtaining the best way to use IoT and associate security.

Keywords Information security · Web of Things · Deep learning · IDS · Organizations

1 Introduction

By definition, it is nearly always claimed that the organized or interconnected edges will always have an outsized example of interconnected information; this type of design works with the rapid exchange of information as it is disseminated, and mistakes are usefully reduced with each associated gadget. Furthermore, such

N. Raghavendra Sai (✉) · T. Ravi Kumar · S. Sandeep Kumar · A. Pavan Kumar · M. Jogendra Kumar
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
e-mail: nallagatlaraghavendra@kluniversity.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_54

affiliations are remaining as immediate numerical aftereffects of the explanation that plane between the information factors and, as a result, it reforms the exhibition acquired, which is reliably obtained from non-direct affiliations and the most recent approach has been found from the unavoidable consequences of the factors that appear at the strategy of pieces. Distinctive confounding layers are found at each predefined work [1]. The proposed assessments unequivocally suggest the assessment of another technology called Web of Things with the introduction of the arrangement and neural affiliation. This research work has regulated the assessment of safety and emerging concerns between important learning and, eventually, the Web of Things at this stage. This may be the case by performing a detailed study on automated thinking or a focus on how the Web of Things application generates a considerable amount of diverse data. Furthermore, we can observe that this evaluation falls within the unusual connection, and along these lines, the new advancement of the Web of Things for an unusual turn of events. Furthermore, we intended to employ the test by guaranteeing the association information, which interconnects and displays the exercises as application attacks for distinct degrees of each attempt. In addition, we admit that by utilizing the KDD 99'Cup, the collect site improves a summary that many web information security experts dismiss as a combination of reference data for our research. Because everything [2] is associated with our job, where the plan implemented for the work is also limited [3], and similarly, the completed up portion is reliant on the execution and results [4].

2 Motivation

The IoT space includes instructional records, which are essentially dimensional, ordinary, and multi-unequivocal. Improving IoT records and its sensitivity is against the operations of standard computer-based intelligence systems. The researchers finally discover the facts from IoT data if there is a solid understanding on the operations of IoT industry, particularly IoT security, which is now requiring an increased research attention. We can notice that IoT introductions are updated on a regular basis by carefully using fundamental learning controls. Here, the enormous and unexpected neural connections of learning are explored, for example, it can be exploited outside schooling. [5]. Cutting techniques can be modified dependent on the convenience of light for IoT plans, regardless of how startling the brain association structures are. We've decided to apply critical learning concerns to IoT network security as a result of this speculation.

2.1 Clarification of the Contribution Topic

The purpose of this application is to log out and respond the following test questions:

- What are the main safety and success issues associated with the IoT environment?
- Is GRU better than other AI approaches in managing the introduction of IoT impedances?
- Create a GRU-based surrogate IDS for each alliance
- How to show updates at all GRU levels?

This assessment can be refined by applying GPU environment assessments to the ongoing IoT information. Despite the fact that there are two or three important learning evaluations, such as large neural affiliations, displaced coders, convolutional neural affiliations, and weaker neural affiliations, wherein the test problem necessitates the employment of a clock, which can misuse recorded data. In this sense, we select a variety of irregular neural relationships for evaluation. Reflecting on the need for light and light responses for IoT affiliation, several tests have been performed with only the estimate of the Gated Recurrent Unit (GRU) [6], while the critical characteristics of RNN and LSTM were not considered. We fine-tune data by disconnecting it at various stages, with an ultimate objective of using a close approach in an Internet of Things connection.

3 Literature Survey

This article analyzes the latest research works by using different methods and plans. This article on “PC Physical Engineering—Social Security for the Future Internet of Things” pushed and attracted us with the inspiration that [7], after investing tremendous energies in thinking about and envisioning this task, we were able to experience the enormous advantages. Preset data are transferred to an assortment level, which mixes a critical confidence loss classifier (created with 10-year social gatherings), with a generally outrageous classifier. The proposed strategy was considered by using the KDD Cup ‘99 organizational roster. The creators have guaranteed a regional speed of 97.90% along with a false unfavorable speed of 2.47%. This is an improvement over the results confirmed by the proposed researchers. Also, Chen, Joy Iong Zong [8] proposes a technique for isolating the information from the flow of the organization. The paper has requested information based on careful assessments by presenting an assessment with the NSL-KDD dataset in which the manufacturers have attained 75.75% accuracy by using only 6 key credits. Kang [9] proposed a DBN to characterize some boundaries for the DNN in order to deliver better assembly results.

LU et al. [10] propose a modified security package for short messages (SMS). The technique basically depends on the RNN model. The researchers have proposed that their evaluations have accomplished an exactness of 92.7%, and furthermore, the existing grouping frameworks are improved. Additionally, the other material work includes the DDoS certificate structure given by Niyaz et al. [twenty one]. They propose a critical learning-based DDoS spread construction to integrate with software defined networks (SDN). The evaluation was done by considering the network traffic. The creators ensure that they have achieved a comparable game plane precision of

99.82% and a 8-class conveying exactness of 95.65%. Despite this, we recognize that a link with this record would be strange by considering the dataset's capacity to get influenced by the environment. Specifically, KDD's reference information combination covers some undeniable classes of attack but the information arrangement used in this report is based on the sub-grouping.

4 Approach

This research work has devised an innovative concept for home IoT affiliation, which would reduce the dataset size of the IDS classifier. We tested a technique to fulfill the intermediate components of IoT by using the 1999 KDD Cup Power blackout Locale dataset. This research work has redesigned the component by placing it on a clashing wood sorter and the most important characteristics are also selected. Furthermore, a thorough examination has been made on the data, and also the data will be sorted according to the basic relationship before using it as model insurance.

Proposals for the IoT Network's Complex Engineering [11] have picked network security as a use case among the many security attempts to illustrate the features provided and remain appropriate for an IoT connection. The intrusion detection system (IDS) verifies the network information by using an anomaly-based technique. The IDS has been mounted at a point of the association in order to obtain all the data about the connection and portray the data as "standard" or as a "department of consideration about irregularities." The intrusion detection system has mounted at a union point to obtain all the information from the association and qualifies the information as "standard" or "assault." Despite normal projects, Machine Learning techniques have been highly employed.

Appraisals are applied to a record of significant worth and portrayal that happens through direct learning [12]. In any case, this approach may not forecast well for IoT cross-section frameworks by considering their heterogeneity. The successful responses to the neighborhood power interruption ought to be light, phenomenal and it will have a fair future. Currently, some researchers are working on few projects and have incorporated artificial intelligence estimations to perform better over longer durations. An IoT setup consists of a handful of gadgets placed at better locations with significant gaps between them. The number of IoT-restricting stunts is more evident than a hack that isn't available or conventional wiring. A single IDS design [13] should be able to store and route mixed data across all devices in a short period of time. Given the large number of devices and the large distance between them, the existing situation on the IoT network will be insufficient. Every ID simply inserts data obtained from gadgets that have a point on a layer of the TCP/IP screen that corresponds to that level. We chose this project as our major effort since it has two or three benefits and employs a creative concept that is currently considered as a typical construction (Fig. 1 and 2).

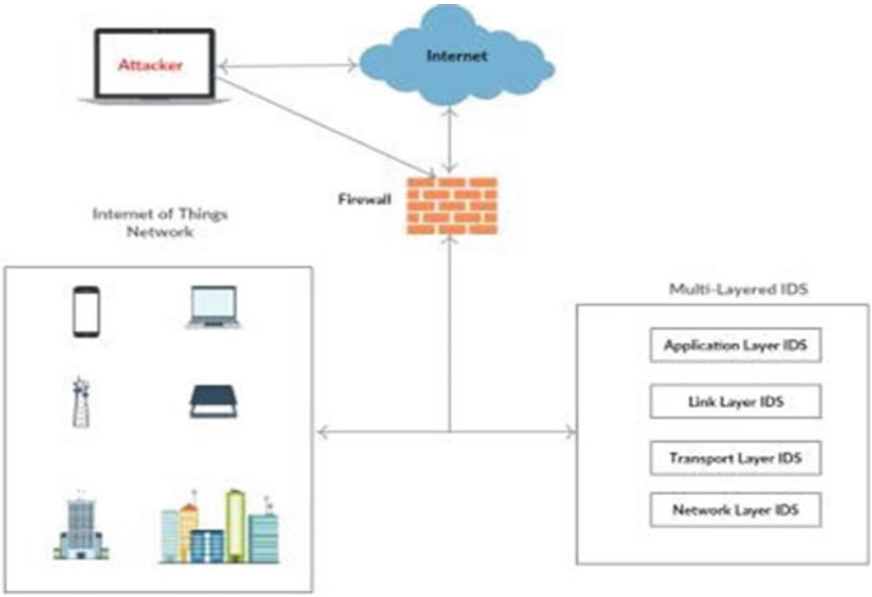


Fig. 1 Architecture of multi-layer IoT network

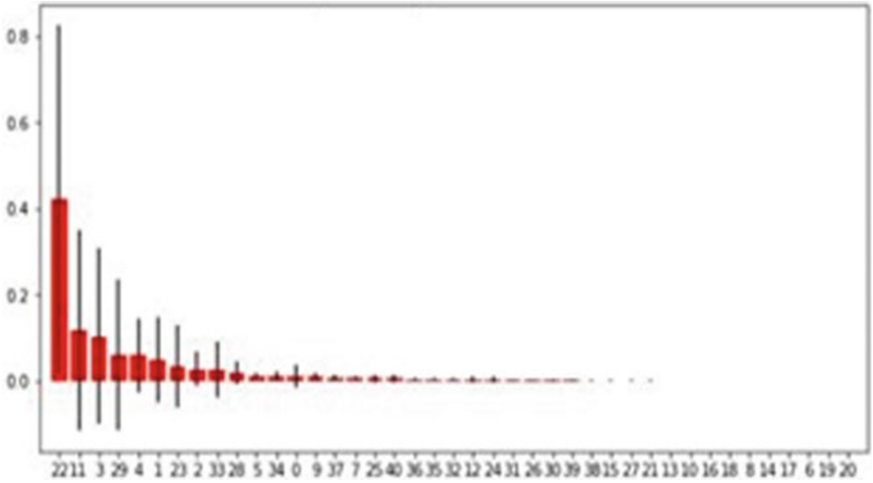


Fig. 2 Feature importance graph intrusion detection system for application layer

5 Experimentation and Results

We have thoroughly clarified various aspects including the arbitrary backwoods grouping calculation that has been used to select the primary and significant qualities of a large number of classifiers as well as meeting the graphical outcomes for each classifier's attributes. The "Conventional Type" work has been chosen at all the degrees of interruption identification.

5.1 Information Stacking and Pretreatment

The information stacking and pre-handling capacity have been demonstrated in this section. Further, the proposed model runs the accompanying approach on the stacked informational index after loading csv data into the information preprocessing structure.

- Removal of duplicates.
- Dividing the information collection into different sets of lights and names.
- Conversion of business data into numerical data.
- Encodes the ordinary as "1" and the uncommon as "0."
- Conversion of approaching data into waterfront data, which will be used later in the association's activity.
- Add another section to the sign set with a sort of hot encoding.

For example,
Common = '1' is tended to as '1 0'
Bizarre = '0' is tended to as '0 1

This is fundamental for Softmax's entropy capacity to beneficially figure the exactness. The information will be transferred to the edge and apply information preprocessing and job confirmation for the dataset. It is important to separate into planning and checking the enlightening records by making the above compromises prior to getting ready for the model which has to be tested (Table 1).

- Normalize the input characteristics and hyperparameters by restricting the information size and further the batch_size is given as "None" to store the data and the patterns are embedded without any help by utilizing the tf.random_normal work.

Table 1 Random forest classifier characteristics

Feature selection	Type of IDS	Accuracy training	Time training	No .of features
Random forest classifier	Transport	99.46	18.84	6
	Network layer IDS	98.705	30.04 seconds	6
	all layer IDS	98.88	50.03 seconds	12
	Application layer IDS	98.6	20.53 seconds	6

Table 2 Dsordered structure of the advanced frame

Time of steps	F-1 Score	Far	Training accuracy	Precision	Recall
10	0.9976	0.0021	98.692	0.9993	0.990
20	0.9921	0.0076	98.831	0.9942	0.3295
30	0.9417	0.05811	96.70	0.9995	0.9951
40	0.9982	0.0015	98.704	0.9966	0.9901

The estimations are depicted effectively by depending on the defense, where the yield is inclined as "1 0" or "0 1."

- Model development: Before building the model, it is important to remodel the commitments for the 3D measurement tensors from the 2D measurement tensors [14]. The cross entropy between the target and softmax execution effort has been used for recruiting the model in our arduous work.

5.2 *Metrics for Evaluation*

In terms of the part, how about doing an immediate assessment of the accompanying characteristics: intensive preparation, learning rhythms, and for a reasonable comprehension of the conduct of the model ward on the differentiation of these hyper-limits as for temporary experiences [15], after the spectacularity of each class of IDS classifier has been identified by adding the GRU gauge limits.

For all IDSs, we perform a similar type of test. Existing location classifiers are linked to classifiers that are currently functioning (all levels, application level, vehicle levels, and other level classifiers, for example, for association). Furthermore, we gladly integrate with the performance outcomes that come with your software. Running the classifier with the application layer has the following side effects: With this knowledge, we were able to achieve the highest planning accuracy with time steps of "40," which is our most recent achievement at this level. The linked graph for the application-level IDS and the disordered structure of the advanced frame (time steps = 40, learning rate—0.01) may be separated in Table 2.

5.3 *Correlation of the Consequences, All Considered and Their Levels*

True-negatives	77304
False-positives	43
False-negatives	794
True-positives	22767

Table 3 Different types of algorithm and its values

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	FAR (%)
PAYLOAD DIR,	78.54	74.32	78.54	75.00
SRC_PORT	96.31	95.42	96.31	95.73
GNNN	93.04	87.07	59.11	12.45
FNN	97.34	92.46	86.88	2.64
RBNN	93.04	69.55	69.8	6.94
K-mean-KNN	93.54	97	98.67	47.8
GRU RNN	97.05	95.71	98.64	10.02
IDS of all layer	99.96	98.810	98.41	0.01
IDS Application layer	99.82	99.66	99.01	0.0015
IDS transport layer	99.0	99.82	99.108	0.2
IDS network layer	99.82	99.66	99.37	0.15

The improved side effects of all IDS classifiers are examined, and the exposure of the IDS classifiers detected at all levels is lower than that of the IDS classifiers at the individual level with respect to accuracy and preparation time. When utilized in multi-faceted engineering, lightweight calculations provide improved execution, which is ideal for an IoT framework. The table shows the correlation of the results.

5.4 Assessment of the Vault Execution with Existing Positions

We are getting close to the end of our test. We conducted additional research and revealed that, as indicated in the table, we examined the outcomes and current study driven by AI estimations on the interference with acknowledgment exhibit. Our study has clearly exceeded any existing research works. Eventually, we want to go deeper into this exploration area (Table 3).

6 Conclusion

The proposed evaluations expressly indicate that we look at another area, the Web of Things, in terms of the link between level and neural associations. This experiment focused on the Internet of Things (IoT) component method, where the association power is modest and the information quantity is not massive. This multidisciplinary test is unique and in that it used the most important learning method for IoT security. By integrating IoT, this research work has proposed a lightweight technique for an interruption identification system (IDS). We recommended reviewing the IDS classifiers for each level in light of the TCP/IP level approach and such an assault at each level. This reduced the amount of each classifier's knowledge pool and enhanced the portrayal in terms of precision, control, coordination time, and sham caution rate. In each ID classifier, we apply massive learning computations to the pack information.

The proposed framework has achieved unprecedented results with astounding results when compared to the existing research works. Similarly, we used KDD 99'cup 22% complete information rating for testing, not in the least like the previous examination paper and it is also remaining fundamental for dynamic IoT affiliations. The accuracy and sham alarm speed of all layer IDS are 98.91 and 0.76%. As the IoT analyzes customer information and industry data, it is fundamental to leave complete reactions in order to avoid security possibilities. This can be imagined with false thinking and essential learning tests, as the IoT produces a colossal heap on heterogeneous information. This research work has applied the Gated-Repetitive Unit-based neural relationship for data grouping. Regardless, various band-aid understandings of a neural vulnerable relationship are available, e.g., dynamic RNN, bidirectional RNN, which can acquire upheld openings by concerning central GRU cells. Additionally, a mixed plot can be developed by using convolutional neural affiliations and eccentric neural associations for multispecific information. Besides, we acquainted ourselves with a level headed and this review is mainly focused on information security. Nevertheless, the proposed testing was helpful and outperformed the cutoff levels of all existing research works.

References

1. K. Xu, X. Wang, W. Wei, H. Song, B. Mao, Toward software defined smart home. IEEE Commun. Mag. **54**(5), 116–122 (2016)
2. G. Pan, G. Qi, W. Zhang, S. Li, Z. Wu, L.T. Yang, Trace analysis and mining for smart cities: issues, methods, and applications. IEEE Commun. Mag. **51**(6), 120–126 (2013)
3. X. Luo, J. Liu, D. Zhang, X. Chang, Alarge-scale web QoS prediction scheme for the industrial internet of things based on a kernel machine learning algorithm. Comput. Netw. **101**, 81–89 (2016)
4. M.A.M. Hasan, M. Nasser, S. Ahmad, K.I. Molla, Feature selection for intrusion detection using random forest. J. Inf. Secur. **7**(03), 129 (2016)
5. N. Raghavendra Sai, J. Bhargav, M. Aneesh, G. Vinay Sahit, A. Nikhil, Discovering network intrusion using machine learning and data analytics approach, in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV),* (Tirunelveli, India, 2021), pp. 118–123. https://doi.org/10.1109/ICICV50876.2021.9388552
6. Y. Li, L. Guo, An active learning based TCM-KNN algorithm for supervised network intrusion detection. Comput. Secur. **26**(7), 459–467 (2007)
7. S.N. Raghavendra, K.M. Jogendra, C.C. Smitha, A secured and effective load monitoring and scheduling migration VM in cloud computing, in *IOP Conference Series: Materials Science and Engineering,* vol. 981 (2020). ISSN-1757–899X
8. A. Bashar, Sensor cloud based architecture with efficient data computation and security implantation for internet of thing application. J. ISMAC **2**(02), 96–105 (2020)
9. M.J. Kumar, G.V.S.R. Kumar, P.S.R. Krishna, N.R. Sai, Secure and efficient data transmission for wireless sensor networks by using optimized leach protocol, in *2021 6th International*

Conference on Inventive Computation Technologies (ICICT) (Coimbatore, India, 2021), pp. 50–55. https://doi.org/10.1109/ICICT50816.2021.9358729

10. L. You, Y. Li, Y. Wang, J. Zhang, Y. Yang, A deep learning based RNNs model for automatic security audit of short messages, in *International Symposium on Communications and Information Technologies (ISCIT),* vol. 16488389 (Qingdao, China, IEEE, 2016), pp. 225–229
11. N.R. Sai, T. Cherukuri, S. B., K. R., A. Y., Encrypted negative password identification exploitation RSA rule, in *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (Coimbatore, India, 2021), pp. 1–4. https://doi.org/10.1109/ICICT50816.2021.9358713
12. N. Raghavendra Sai, K. Satya Rajesh, An efficient los scheme for network data analysis. J. Adv. Res. Dyn. Control Syst. (JARDCS) **10**(9) (2018) ISSN: 1943–023X
13. M. Jogendra Kumar, N. Raghavendra Sai, C. Smitha Chowdary, An efficient deep learning approach for brain tumor segmentation using CNN, in *IOP Conference Series: Materials Science and Engineering*, vol 981 (2020)
14. G. Edamadaka, C. Smitha Chowdary, M. Jogendra Kumar, N. Raghavendra Sai, Hybrid learning method to detect the malicious transactions in network data, in *IOP Conference Series: Materials Science and Engineering*, vol. 981 (2020)
15. A.A. Shah, M.S.H. Kiyhal, M.D. Awan, Analysis of machine learning techniques for intrusion detection system: a review. Int J. Comput. Appl. **119**(19), 23 (2015)

Hand Gesture Controlled Contactless Elevator

G. Sai Pravallika, M. Lakshmi Akhila, M. Divya, T. Madhu Babu, and G. Kranthi Kumar

Abstract In the existing pandemic situation, it is not safe for people to touch all the objects/things present around them. In urban areas, majority of human population use elevators in their everyday lives and they are ought to touch the elevator buttons to move from one place to another. This increases the risk of COVID-19 spread. To overcome this challenge, a contactless system has been developed by incorporating a sensor to capture the hand gestures and control elevator functions. In this way, the suggested technique prevents the manual pressing of elevator buttons. In the proposed system, People wave their hands in front of the sensor to select the desired floor number and trigger the elevator to move from the current to the desired floor.

Keywords Hand gestures · Elevator · Sensor · Contactless · Elevator buttons

1 Introduction

Technology advancements have transformed the globe into a smart world where everything is controlled by human hands and voices. For many years, touchless interfaces have been developed separately. They can be used in a variety of applications [1]. An example is Ambient Assisted Living, where voice recognition and gesture recognition are the key technologies used for assistive environments beyond the tendency to be used for activity recognition, touchless interfaces in order to enable the users to achieve a natural interaction with the available devices [2]. These touchless interfaces include many sensors that are capable of recognizing human gestures and voice. These systems perform the task based on the input given by the user.

In the present day, most places include smart devices that work with the assistance of Human gestures and voice. In the case of human-computer interaction, Gesture Recognition plays a key role [3]. Gestures come from body part's movement. Hand waving is one of the Gestures. This gesture recognition enables humans to interact

G. Sai Pravallika (✉) · M. Lakshmi Akhila · M. Divya · T. Madhu Babu · G. Kranthi Kumar
Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

G. Kranthi Kumar
e-mail: kranthi@vrsiddhartha.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_55

with the machine without the help of any mechanical device. This gesture can be recognized by the sensors. Those sensors can be either gesture sensors or any optical sensor like camera. The most optimized way is to combine the touchless interfaces and gesture recognition [4]. The elevator buttons can be operated without touching them by providing touchless interface, which contains the sensors to recognize the hand gestures made by the person present in the elevator [5].

The users' knowledge about the system plays the most crucial element to be taken into consideration. The majority of existing systems require users to have technical knowledge in order to use them. In order to operate the system, the user needs to be knowledgeable on how to create gestures [6]. As we all know that elevators are used by both literate people and illiterate people. The system working should be convenient for all types of users. And another important factor to be considered is the safety of the people in these pandemic times. So it is best to use the touchless systems in the places, where the people are more so that the virus spread can be controlled. The best applicable place to use these touchless interfaces are elevators because people mostly use the elevators in their daily lives.

2 Literature Survey

Montanaro et al. [7] proposed a touchless human interface for controlling the elevator. The user has to place the hand along the xyz plane in front of the camera placed in the elevator. The gesture is recognized based on location of hand in the xyz plane. This consists of three different interaction modes with the elevator controls. The first is about tracking movements along y-axis which is for floor selection and x-axis which is for floor confirmation. The second is about moving hand in circular way for floor selection and waiting time. The third mode is about placing hand in xy-plane which is for floor selection and moving fingers in z-axis for the confirmation of floor number.

Chen et al. [8] proposed a system that involves four steps. They are Hand detection, Fingers and palm Segmentation, Finger recognition, Hand Gesture recognition. This system uses background subtraction for detecting the hand. The result of this step is binary image. The second step is Fingers and palm segmentation. This is used for dividing palms and fingers. This segmented fingers in this step facilitates the third step that is finger recognition. Based on the output of the before step hand gesture is recognized using simple rule classifier.

Cernys et al. [9] proposed a lift model controlled by voice and sensor control panel. Average powered controller is used for the construction of this system. This system consists of voice recognition system and programmable terminal. These two are connected by logical lift program. This system uses well known algorithm called Dynamic Time Wrapping (DTW). A set of voice commands are defined that contains eight Luthanian words such as floor number, go and stop. This set of commands consists of phrases such as greeting and goodbye. This system can be used mostly in smart home projects and also used for physically disabled people.

Ren et al. [10] proposed a system that can be used for Hand gesture recognition. This system uses kinetic sensor which is a recently developed depth sensor. This work mainly deals with building a robust part-based hand gesture recognition system. This system receives noisy hand shapes from the kinetic sensor. Novel distances metric, Finger Earth Mover's Distance are used to handle the noisy data obtained from the sensor and also to measure the dissimilarity between hand shapes. The main advantages of using these methods are these can match only the finger but not the whole hand. This feature can be benefited while differentiating the hand gestures which are very similar but with a slight difference.

Dardas [11] proposed a real time system that can be used to interact with the applications like video games through Hand Gestures. The main function of this system is detecting and tracking a bare hand that is present in cluttered background. Initially face subtraction and hand gesture recognition which can be done through a bag of features and multiclass support vector machine are performed. After this phase skin detection and hand posture contour comparison algorithm are executed. The result of these phases a grammar is built that will generate the gesture commands that can be used to control the applications.

Shinde et al. [12] proposed a system that is used for hand gesture recognition which works in any type of environment. This system consists of three main steps. They are Hand detection, Hand Gesture Recognition, Finger Recognition. This system takes an image that is captured by the camera is taken as input. From this, we locate the hand position using the hand tracking function for segmenting the hand function. In the next step, we use two methods called Near-Convex Decomposition and Threshold Decomposition for obtaining fingers from the hand in the image. Then we use the Finger Earth movie distance that measures the distance between the fingers in the image captured. Based on the distance obtained we can determine the gesture.

Rokade et al. [13] proposed a new technique for recognizing hand gestures that are focused on thinning segmented images. This system takes RGB images as input and converts them into YIQ images. Now the whole image is divided into small 4 × 4 windows. For each window corresponding histogram, each histogram is compared with reference histogram and a similarity factor is also observed. RGB image is also converted as YCbCr. If the values of Cb and Cr of that pixel are in the given range, then all the white pixels are converted into black pixels. Further, the entire gray and white windows are integrated. In the next step, similarity factor is changed by considering the effect of neighbouring window. The final result of hand gesture is obtained by applying the threshold value.

Nair et al. [14] proposed a wireless system that is used for controlling the Elevator through an interface. With this system, the user can view the current position of the elevator car in the handling unit. The main aim of this system is to reduce the overall waiting time of the user. This system uses RF Module. This contains two units which are Main Unit and Handling Unit. The main unit is associated with the elevator controller and the handling unit is associated with the user.

3 System Description

The proposed system includes a Micro Controller [Arduino Nano], Gesture Sensor, OLED Display board and Servo motor.

Table 1 represents the list of hardware used in the proposed system. The hardware used is Arduino Nano which is used for interfacing the remaining components, OLED display which is used for the present the system working visually to the user, APDS9960 gesture sensor is used for receiving the gestures from the user. And the servo motor is used for elevator environment simulation.

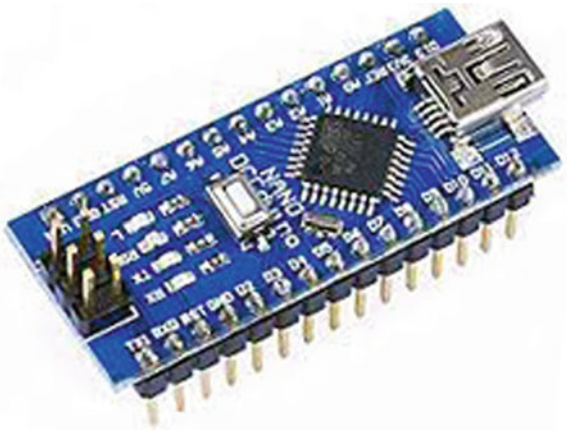
A. Arduino Nano

The Arduino Nano shown in Fig. 1 is a small board that uses the ATmega328P microcontroller (Arduino Nano 3.x) [15]. This microcontroller is used for interfacing all the components and establish communication between them by sending clock signals to the components for efficient sending and receiving of data through protocols. It has the same functions as the Arduino UNO but in a smaller package. It has a 5 V working voltage. Mini-B, unregulated 6–20 V external power supply (pin 30), or controlled 5 V external power supply (pin 27) can all be used to power the USB link. The power source will be chosen automatically based on the highest voltage source. Nano has 14 optical pins, 8 analogue pins, 2 reset pins and 6 power pins. TTL serial data is received (RX) and transmitted (TX) using this device.

Table 1 Hardware/sensors used

Hardware/sensor used	Specification
Adruino nano	Microcontroller ATmega328 Operating voltage: 5 V Input voltage: 7–12 V Input voltage: 6–20 V Digital I/O pins: 14 Analogue input pins: 8
OLED display	High-resolution at 128 × 64 pixels power consumption: 0.06 W Power supply AC: 3–5 V, Working temperature: −30 to 70 degree Celsius Dimensions: L27.8 × W27.3 × H4.3 mm
APDS9960 RGB and gesture sensor	Ambient light and RGB sensing, Proximity sensing and advance gesture detection Operating voltage: 2.4–3.6 V Operating current: 0.2 mA Communication protocol: 400 kHz
Servo motor	Operating voltage: 5 V Torque: 2.5 kg/cm Operating speed: 0.1 s/60° Rotation: 0–180°

Fig. 1 Arduino nano



B. APDS9960 Sensor

The Broadcom APDS-9960 shown in Fig. 2 is a digital RGB, ambient light, proximity and gesture sensor kit with an 8-pin connector [10]. This device can be used mostly for building smart devices and touchless interfaces. The device can detect red, green, blue, transparent (RGBC), proximity and movements using an IR LED and an I2C compliant interface [16]. The RGB and ambient light sensing function measures light intensity in a variety of lighting conditions and through a variety of attenuation materials, such as darkened glass. Gesture detection employs four directional photodiodes with a transparent blocking filter to detect basic UP-DOWN-RIGHT-LEFT gestures as well as more nuanced gestures. This can be also used to detect if there is any object nearby. This is an alternative sensor to the proximity sensor. Another advantage of this sensor is it can the RGB colours. This feature can be utilized in colour detection which is now done using cameras and image processing.

Fig. 2 APDS9960 sensor

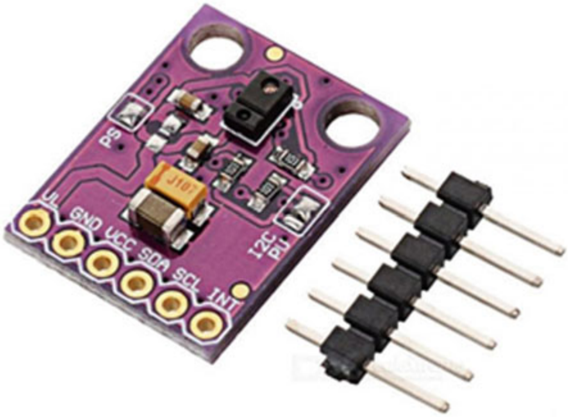


Fig. 3 OLED display

C. OLED Display

The emissive electroluminescent layer of an organic light-emitting diode (OLED or organic LED), also known as organic electroluminescent (organic EL) diode shown in Fig. 3, is a film of organic compound that will emit light when exposed to electric current. This organic layer is sandwiched between two electrodes, at least one of which is usually translucent. The image which is to be displayed on the OLED is converted as hex code. This hex code is bitmapped onto OLED. The OLEDs are used in devices such as television screens, computer monitors and portable systems such as smartphones, handheld game consoles and PDAs to construct digital displays. The creation of white OLED devices for use in solid-state lighting applications is a major field of study.

D. Servo Motor

A Servomotor is a rotary or linear actuator that can control angular or linear position, velocity and acceleration precisely [17]. An appropriate motor is connected to a position feedback sensor. It also necessitates a complex controller, which is frequently a separate module created exclusively for servomotor use. A servomotor is a closed-loop servomechanism that controls its motion and final position using position feedback. The input to its control is a signal (either analogue or digital) representing the output shaft position. To provide position and speed feedback, the motor is coupled with some type of position encoder. Only the position is measured in the most basic case. This sensor is used for simulating the environment (Fig. 4).

4 Methodology

An Arduino Nano, an APDS9960 Gesture Sensor, a Servo motor and an OLED display module are used in this Contactless Elevator. You can easily control your Lift with this gesture-based control panel by making a hand gesture. To read the

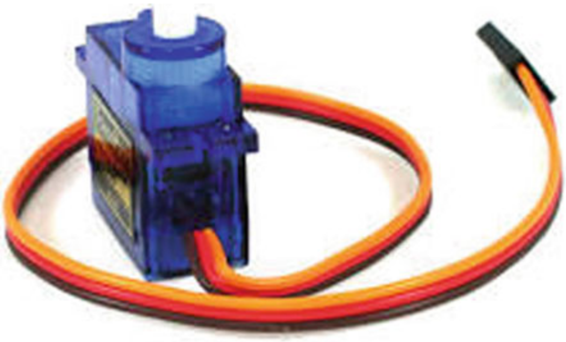


Fig. 4 Servo motor

movements, the APDS9960 Sensor is used. The left gesture is used to close the lift door and raise the lift according to the floor number, while the right gesture is used to open the door. The person in the elevator makes a hand wave in the upward direction. This gesture is recognized and considered to perform the task of increasing the floor number by 1. When the person waves his hand in the downward direction it is indicated to decrease the floor number by 1.

Servo motor is connected to the Arduino Nano and a belt like object is connected to the Servo motor. When a person waves his hand in the right side direction then the system compares the current floor number with the required floor number. If the Current floor number is less than the selected floor number it is indicated that lift has to move upward. In this case, the Servo motor is rotated in clockwise direction and belt attached to it lift the load upwards. If the Current floor number is greater than the selected floor number it is indicated that lift has to move downward. In this case, the Servo motor is rotated in anti-clockwise direction and belt attached to it move the load in the downward direction.

Figure 5 represents the flow of steps involved in the proposed. The initial step is capturing the gesture made by the user. Then based on the gesture operation is performed. In the right gesture, a comparison is made between current floor number and required floor number. Based on the result rotation of servo motor is decided. This overall process is shown in OLED Display.

Figure 6 represents the pin to pin connections between Arduino, Gesture Sensor and OLED Display. This proposed system uses I2C protocol which transfers the data between the integrated circuits bit by bit through SDA line. As this protocol is asynchronous the output needs to be synchronized. This is done through Serial Clock Line (SCL). It also represents the interconnections in the proposed framework The Arduino Nano is connected to an APDS9960 sensor and an OLED display. Both the VCC and GND pins of the APDS9960 Sensor and the OLED Display are wired to Arduino's 3.3 V and GND. The APDS9960 Sensor's SCL and SDA pins, as well as the OLED Display's SCL and SDA pins, are attached to the Arduino's A5 and A4 pins, respectively. Servo Motor 5 V pin is connected to Ardunio's 3.3 V pin and Pulse of servo motor is connected to D12 of Arduino.

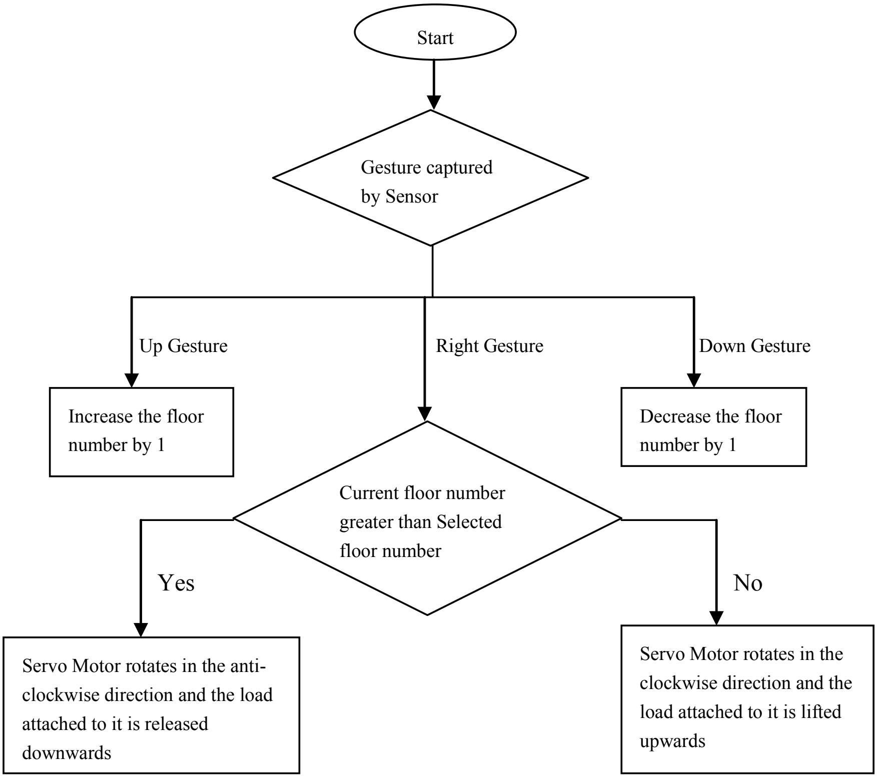

Start
Gesture captured by Sensor
Up Gesture
Right Gesture
Down Gesture
Increase the floor number by 1
Decrease the floor number by 1
Current floor number greater than Selected floor number
Yes
No
Servo Motor rotates in the anti-clockwise direction and the load attached to it is released downwards
Servo Motor rotates in the clockwise direction and the load attached to it is lifted upwards

Fig. 5 Flow sequence of proposed system

5 Results

Figure 7 Indicates the components of the working hardware. Gesture Sensor, OLED Display, motor are connected to the Arduino. The motor rotates based on the gesture made by a person

Figure 8 Indicates that OLED Display is currently set at floor number 0. The person will wave his hand up to increase the floor number and wave his hand to decrease the floor number

Figure 9 Indicates that lift is moving upward. If the person sets the floor number and waves his hand left then it checks if the current floor number is less than the desired floor then the lift starts moving upwards

Figure 10 Indicates that lift is moving downward. If the person sets the floor number and waves his hand left then it checks if the current floor number is greater than the desired floor then the lift starts moving downwards

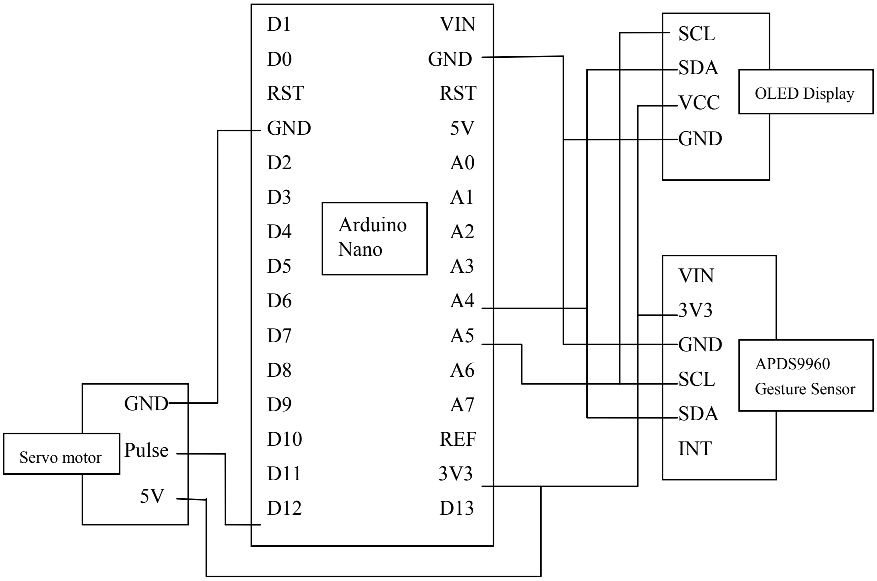


Fig. 6 Proposed system

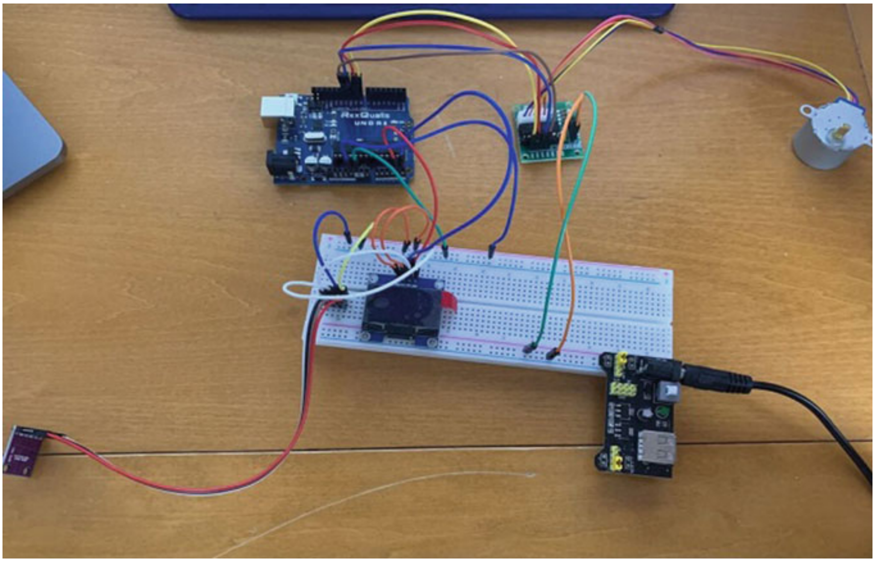


Fig. 7 Working hardware

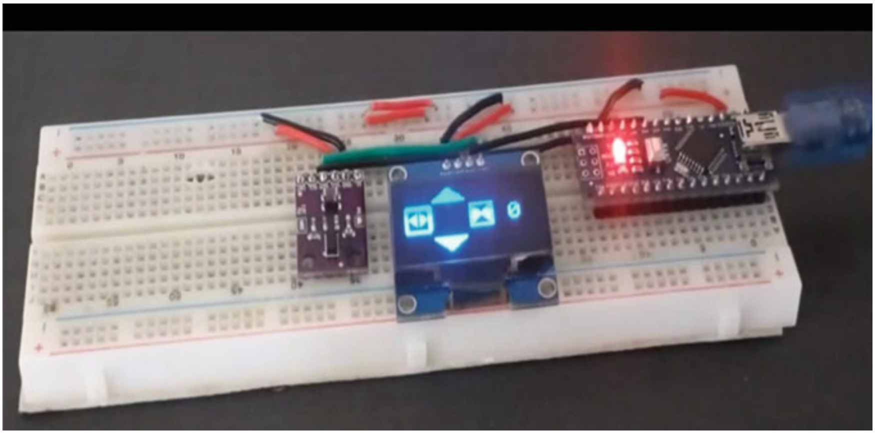


Fig. 8 OLED displaying the floor number

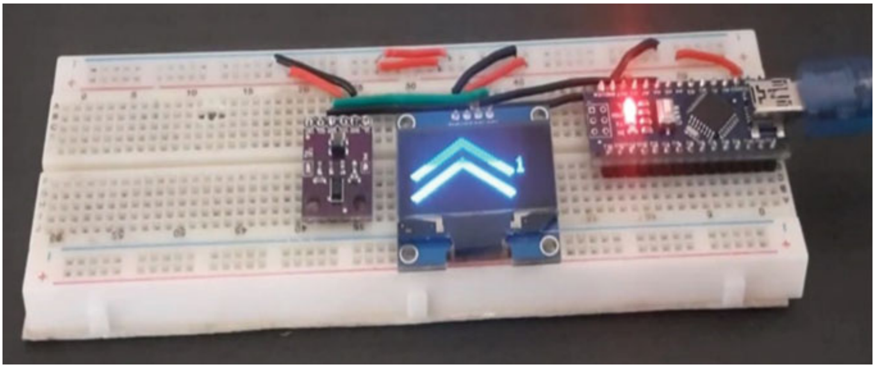


Fig. 9 OLED Displaying upward movement of lift

6 Conclusion and Future Work

Most people don‘t realize how elevators have become a high-risk place. This experiment was carried out in order to develop a touchless interface treating to prevent contact with elevator buttons. This system is easy to understand for the people. There is no requirement that people who use this system must have technical knowledge. They should only know direction of hand wave and its function. This requires only memory but not technical knowledge. The researched and involved sensors, as well as a method help in identifying the hand gestures. The system shows promising results as APDS9960 RGB and Gesture Sensor detect the presence of the objects as soon as they're placed. The application of this system is it can be used in apartments and offices. Because most people use the elevator for their daily schedule. This system can be efficient in places with less than 10 floors. If the number of floors is more

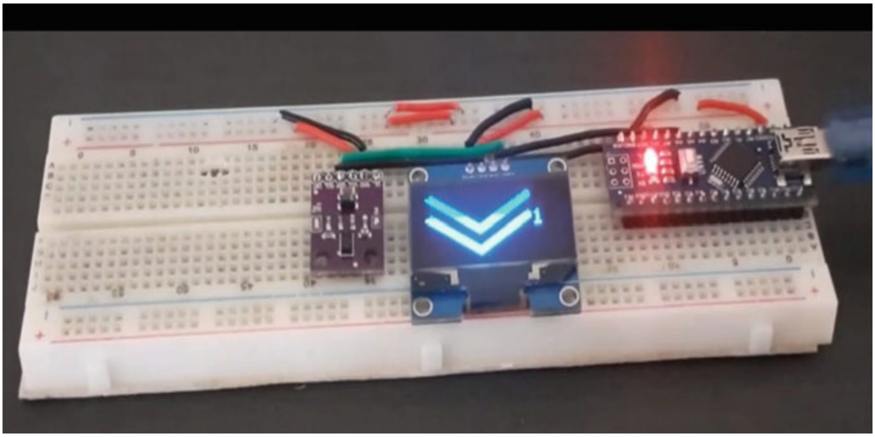


Fig. 10 OLED Displaying downward movement of lift

than 10 it is not convenient for the people to wave hands more than 10 times. Rapid advances in technologies provide ample opportunities for bringing in innovations in gesture recognition. Through testing, it was concluded that the best place to keep the sensor inside the elevator. The use of the whole system is cheaper and efficient.

References

1. H.I. Abbasi, A.J. Siddiquis, Implementation of smart elevator system based on wireless multi-hop AdHoc sensor networks, in *IEEE 2nd International Conference on Networked Embedded Systems for Enterprise Applications* (2011)
2. S.S. Sonavane, V. Kumar, B.P. Patil, MSP 430 and RF 24L01 based wireless sensor design with adaptive power control. ICGST-CNIR J. **8**(2), 11–15 (2009)
3. Y. Ren, C. Gu, Real-time hand gesture recognition based on vision, in *Proc. Edutainment* (2010), pp. 468–475
4. L. Yun, Z. Peng, An automatic hand gesture recognition system based on Viola-Jones method and SVMs, in *Proceeding 2nd International Workshop Computer Science Engineering* (2009), pp. 72–76
5. S. Shakya, L.N. Pulchowk, Sensor assisted incident alarm system for smart city applications. J. Trends Comput. Sci. Smart Technol. **1**(2020), 37–45 (2020)
6. S. Smys, A. Basar, H. Wang, Artificial neural network based power management for smart street lighting systems. J. Artif. Intell. **2**(01), 42–52 (2020)
7. L. Montanaro, P. Sernani, D. Calvaresi, A touchless human-machine interface for the control of an elevator, IEEE Trans. Instrum. Meas. (2016)
8. Z. Chen, J.T. Kim, J. Liang, J. Zhang, Y.B. Yuan, Real-time hand gesture recognition using finger segmentation, in *International Conference on digital Image Processing IEEE Computer Society* (2018), pp. 228–291
9. P. Cernys, V. Kubilius, V. Macerauskas, K. Ratkevicius, Intelligent control of the lift model, in *IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 8–10 Sept 2018

10. Z. Ren, J. Yuan, J. Meng, Z. Zhang, Robust part-based hand gesture recognition using kinect sensor. IEEE Trans. Multimedia **15**(5), 1110–1120 (2013)
11. N.H. Dardas, N.D. Georganas, Fellow, IEEE, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques. IEEE Trans. Instrum. Meas. **60**(11), 3592–3607 (2011)
12. V. Shinde, T. Bacchav, J. Pawar, M. Sanap, Hand gesture recognition system using camera. Int. J. Eng. Res. Technol. (IJERT) **3**(1) (2014*)*
13. R. Rokade, D. Doye, M. Kokare, Hand gesture recognition thinning method, in *IEEE International Conference on Digital Image Processing (ICDIP)* (Nanded India, 2009), pp. 284–287
14. D. Nair, S. Kunnel, A. Singh, S. Shanmugam, N. Tambe, Elevator controller using wireless system. PCE Electron. J. (2017)
15. P. Swetha, S. Amardeep, A. Siva Nagasen, G. Manoj Kumar, G. Kranthi Kumar, Arduino based virtual keyboard for locked-in-sysndrome, in *Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (2020)
16. M Yuvaraju, K. Sheela, S. Rani, Sheela, Smart wireless elevator design. Int. J. Res. Electric. Electron. Eng. **3**(3) (2015)
17. J.W. Chen, T. Trung, Y.C. Hsieh, Wireless elevator communication and monitor system design based on zigbee technology and Ethernet. *IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)* (2019)

Real-Time Hand Tracking and Gesture Recognizing Communication System for Physically Disabled People

Riya Gupta, Dishank Oza, and Sunil Chaudhari

Abstract Regardless of its importance, a systematic literature analysis and classification scheme for sign language are lacking. Communication is the most basic form of interaction. Deaf people lack this form of communication, and hence, sign language interaction is of utmost importance for them to be understood. In recent years, emerging technology developments, such as smartphones, have provided consumers with a slew of new features. If such mobile devices can recognize sign languages, sign language users can access a far more comprehensive range of mobile applications. This paper proposes a paradigm for sign language interaction between a consumer and a computer in a complex environment. The interface that is being proposed here can be used in various applications such as image browsers, games, and so on. Even in a complicated setting, the suggested model exhibited good accuracy and effectively recognized motions in low-resolution picture mode. The suggested model was tested using a labeled dataset of hand movements consisting of 99,000 gestural images. In our suggested Inceptionv3 based model, we achieved superior results by extracting features from the hand and identifying hand motions with 91% accuracy. We also compared our model to Resilient Back, VGG-16, and ResNet50, all of which had accuracies of 75–85%. The trained model may be used to recognize static hand pictures as well as dynamic motions captured on video in real-time.

Keywords Handshape recognition · Sign language recognition · Gestural interaction · ASL American Sign Language · Machine learning

1 Introduction

People make gestures when they talk. The gesture is a fundamental component of language that adds new and important detail to a spoken message while also reflecting the speaker's underlying understanding and experiences. Theoretical perspectives on

R. Gupta (✉) · D. Oza · S. Chaudhari
Fr.Conceicao Rodrigues College of Engineering, Mumbai, India

S. Chaudhari
e-mail: sunil_chaudhari@fragnel.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_56

speech and gesture suggest that they share a conceptual origin and a closely intertwined relationship, with time, context, and purpose overlapping. We have made use of image processing and computer vision to recognize gestures here. Computers can recognize human behavior via gesture recognition and serve as a translator between the two. This will allow humans to communicate naturally with computers without coming into physical contact with mechanical devices. In [1], a rehabilitation application to improve upper limb activity and mobility are presented. It uses a near-infrared visual device that also provides hand-skeleton information. Physically disabled people use signals to communicate. When broadcasting audio is unlikely, or typing and writing are difficult. However, vision is possible. This culture uses sign language to communicate. Previously, sign language was the only means of communication between people. When people don't want to talk, they usually use sign language, but for the deaf and dumb community, it is their only means of communication. Sign language and spoken language both communicate the same message. This is spoken by deaf and dumb people worldwide but in ethnic dialects such as ISL and ASL. Nearly two million hard-of-hearing people in the USA and Canada are using ASL as their primary basis of communication [2]. We used isolated ASL gesture recognition techniques in this paper.

The novelty of the proposed method is listed as follows:

(i) The first novelty of the proposed method is that Machine learning (ML) is utilized to infer 21 3D keypoints of a hand from a single picture, allowing for high-fidelity monitoring of hands and fingers.

(ii) To collect hand gesture data, several past research required users to wear a data glove. However, the data gloves' sophisticated sensors are pricey, limiting their use in real life. The authors employ a TOF camera (Kinect sensor) to collect the depth of the surroundings and a customized tape worn across the wrist to identify the hand area in their research. Our method merely requires the use of a regular camera to collect the visual information of the hand motion, and no special tape to recognize hand areas.

(iii) The palm detector is only executed as needed (which is rather rare), which saves substantial computing time and is highly efficient and fit for real-time applications.

2 Related Work

Disabled people's technologies and gadgets have always attracted innovation since they help them live a normal and pleasant life. Various breakthroughs in the technical world today have assisted various groups of persons with disabilities in terms of research and product system development in assistive technologies that help various handicapped persons to carry out their daily tasks. The engagement approach should be as natural as possible to increase the quality of interaction in a dynamic environment [3]. Human hand gestures could be defined as a set of permutations generated by movements of the hand and arm.

For decades, SLR has been the topic of study. Sensor-based devices, such as SignSpeak, were employed in much research. This gadget utilized a variety of sensors, including flex and touch sensors for finger and palm movements, as well as accelerometers and gyros for hand movement; the gloves were then taught to detect distinct gestures using Principal Component Analysis, and each gesture was then sorted into alphabets in real-time. An Android phone was also utilized to display the text and words received through Bluetooth from the gloves. The accuracy of SignSpeak was found to be 92% [4]. Electromyography, for example, is another method of recording indications using motion sensors (EMG) [5]. RGB cameras [6], Kinect sensors [7], compute an extra modality like optical flow to improve their performances [8], egocentric gesture recognition [9], and leap motion controllers [10], as well as their combinations, are all possible sensors. Deep learning-based real-time hand tracking and gesture detection systems. Ranganathan [11] Space and time-depth particulars (Spatial-Temporal depth details—STDD) and the random forest in the final step for movement classification are used to detect human movement in real life. The system shown makes use of Kinect sensors to capture data during the data collection step. Sign language is used by those who are unable to communicate verbally with others [12]. People who are unable to communicate with others via speech utilize sign language as an alternative. For human-computer interaction, hand gesture recognition is critical [13]. Chen [14] the background subtraction approach is used in the framework to remove the hand region from the backdrop. Then, the palm and fingers are segmented so as to detect and recognize the fingers. Finally, a rule classifier is applied to predict the labels of hand gestures.

2.1 Comparative Study of Existing Systems

In the data glove-based technique, the user is required to wear a glove or attach sensors to the fingers. Pressure sensors are used, which measure the pressure between the knuckle and the first joint of the finger. Another glove by Abhishek et al. [15] Abhishek, K.S. Qubeley, and others use capacitive touch sensors, which output binary on/off signals, which are triggered when they are brought within 1.6 mm of human skin. Wang, R. Y., & Popović, J. Have developed a Colorglove in which is a combination of both the data glove-based approach and the vision-based approach. It is, however, somewhat similar to the vision-based system. The drawbacks are similar to those of data glove-based approaches: they are unnatural and, due to hygiene concerns, unsuitable for applications with multiple users [16]. These gloves are not portable. The movement is limited only to fingers, but there are some sign languages that include the complete movement of hands. Therefore this technique is not feasible.

The vision-based gesture recognition technique overcomes the drawbacks of the glove-based system. The consumer does not need to use gloves, sensors, or wires with this tool. Video camera (s) are used to capture images of hands while they perform certain gestures, which are further processed and analyzed using computer vision techniques. This type of hand gesture recognition reduces the complexity,

Table 1 Comparison table

Criterion	Glove-based system	Vision-based system
User cooperation	Yes	No
User intrusive	Yes	No
Precise	Yes	Yes
Flexible to configure	Yes	No
Flexible to use	No	Yes
Health issues	Maybe	No

makes the process look natural, and is very convenient for users. Table 1 depicts the comparative study of the earlier system, Glove based with the proposed solution which is a Vision-based system.

As per the above literature review, every paper is prepared with the primary goal of creating a working prototype for building a low-latency, high-efficiency real-time online hand gesture recognition system with minimal processing costs.

3 Proposed Work

3.1 Problem Statement Analysis

Real-time hand tracking and gesture recognition system using Neural Networks (Deep Learning) in an unstrained Environment. A framework for creating pipelines that process perceptual data from several modalities, including video and audio. Machine learning (ML) is utilized to infer 21 3D key points of a hand from a single image, allowing for high-fidelity monitoring of hands and fingers. Using various external hardware and sensors can be expensive and not user-friendly. The model, with the help of a just system camera, will identify hand gestures in real-time and give results instantly. This application will help a diverse group of people with disabilities to communicate just like normal people very easily using video conferencing. The hand tracking solution makes use of a machine learning pipeline that consists of numerous models that work together. Simple socket programming has been used to aid communication through the video conferencing web application between a disabled and normal person with server-side connected to the model on the disabled person's end and client-side on the other end.

3.2 Dataset Description

The data set is a collection of images of alphabets from the American Sign Language, separated into 35 folders that represent the various classes. The data set contains 99,000 images which are 200 × 200 pixels. There are 35 classes, of which 26 are for the letters A-Z as shown in Fig. 1 and 9 classes for *SPACE*, *DELETE, OKAY, FIST, GOOD, BAD, PLEASE, STOP,* and *NOTHING.* These 9 classes are very helpful in real-time applications and classification. There are 26 grades in all. The dataset is divided into two sections: a training set of 2000 images and a testing set with 1000 images.17,500 images were run through the same model during the testing phase [17].

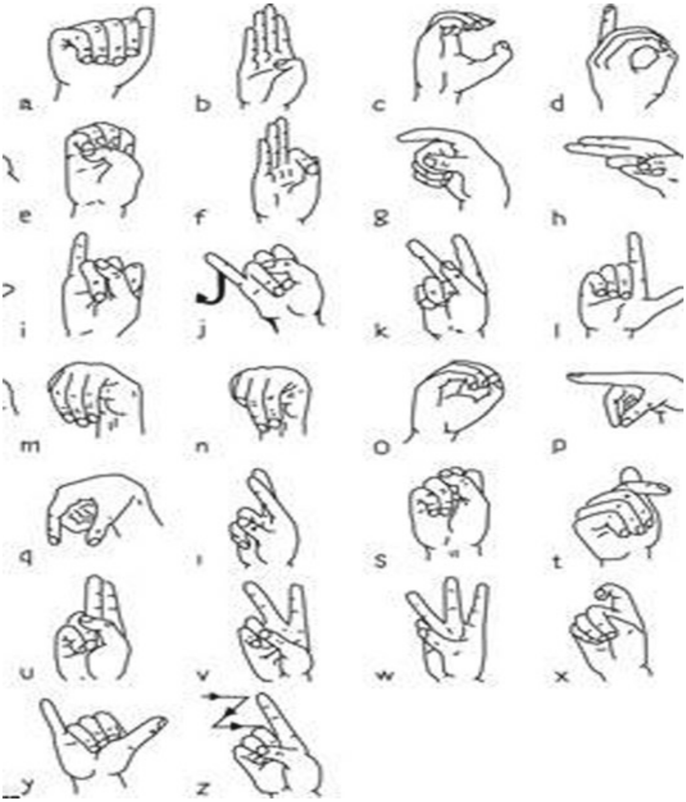


Fig. 1 American sign language alphabets

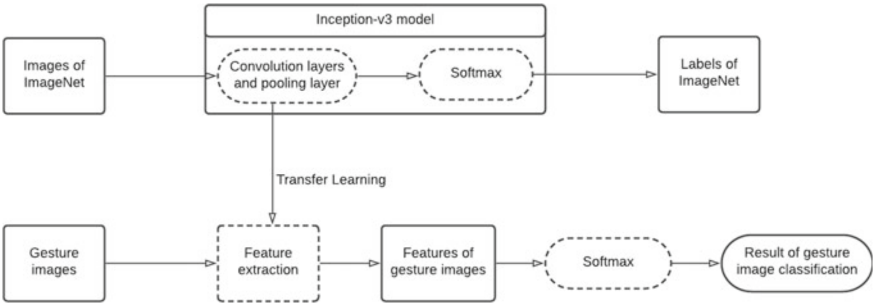


Fig. 2 Methodology broken down into steps

3.3 Design and Methodology of Proposed System

For training the model, we first extract the frames of the videos from the dataset, then we applied preprocessing algorithms like Skin Modeling, Removal Background, conversion from RGB to binary, and CbCr (Luminance Chrominance) to remove noises and also extract features from the frame. These features help the inception model to learn accurately about the gestures. The inception model takes this processed frame as input and tries to reduce the size of the frame by using convolution and max pooling layers. After this, the model then learns the unique features of every gesture. This process is repeated for every frame in the dataset for n steps.

Further, the methodology of the proposed system is explained in much more detail diagrammatically as shown in Fig. 2 for a better comprehension of the process.

3.4 Frame Extraction and Background Removal

Gesture videos are broken down into frames. Then preprocessing is performed to remove all the background noise from the image that is everything except hands. The final image consists of the grayscale image of the hand to avoid color-specific learning of the model, as shown in Fig. 3.

3.5 Feature Extraction Algorithm

The design of CNN includes a pooling layer and a convolution layer. These two modules are stacked one on top of the other to build a deep model. The weights are shared using the convolution layer, and the output is subsampled using the pooling layer. This approach decreases the data rate to a fraction of what it is in other neural network models. CNN's invariance is due to its unique pooling techniques and convolution layer weight sharing procedure. CNN is frequently utilized in image and

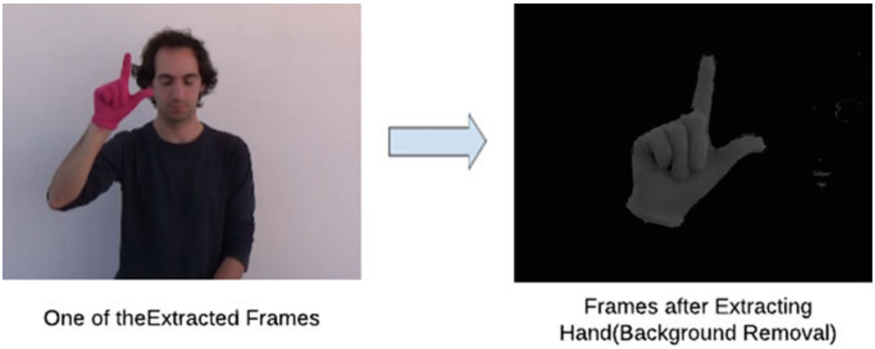

One of theExtracted Frames
Frames after Extracting Hand(Background Removal)

Fig. 3 Frame extraction (background removal)

Table 2 Analysis summary of discriminative architectures [18]

Methods	Reference article	Application	Merits	Demerits
Adaptive discriminative learning	Wang et al. [19]	Scene recognition	Better recognition efficiency	Slow learning rate
	Ding et al. [20]	İmage retrieval	Computation time	Less efficient compared to other neural network models

signal processing applications. Table 2 describes the applications of discriminative architectures in various domains with their merits and demerits.

After preprocessing of images in real-time, pairing convolutional networks with various deep learning methods, a model is created to identify and interpret various hand gestures. The framework in the proposed method is divided into three stages: preparation, testing, and recognition. In the training phase, images were passed through the Inceptionv3 model, where the images were divided into 35 categories, that is 26 English alphabets and 9 gestures.17,500 images were run through the same model during the testing phase. For greater accuracy, all of these images were manually annotated. On an Intel Core i5 processor with 12 GB RAM and a 64-bit operating system, the model was trained for 5000 steps with a loss of 0.4.Inceptionv3 is one of the most well-known neural networks for object detection. As shown in Fig. 4, Convolutions, average pooling, max pooling, concats, dropouts, and completely linked layers are among the symmetric and asymmetric building blocks in the model. Batchnorm is applied to activation inputs and is used extensively in the model. Softmax is used to calculate the loss.

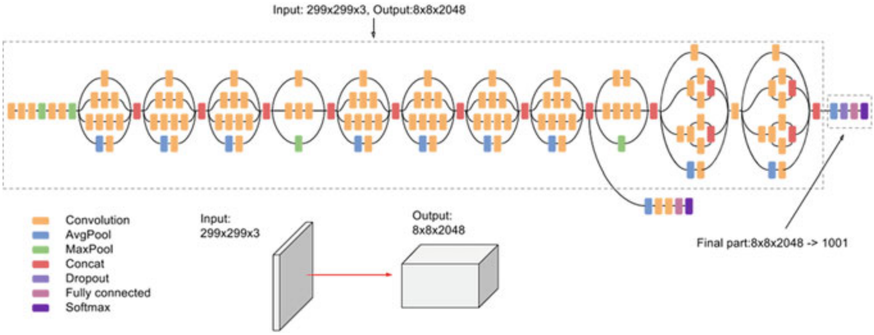


Fig. 4 High-level diagram of model

3.5.1 Efficient Grid Size Reduction

The function map downsizing is traditionally achieved by max pooling, as in AlexNet and VGGNet. However, either max pooling followed by conv layer is too selfish, or conv layer followed by max pooling is too costly. The below is a proposal for an optimal grid size reduction (Fig. 5):

Conv with stride 2 produces 320 attribute maps thanks to the powerful grid size reduction. Maximum pooling yields 320 function maps. And these two sets of feature maps are combined to create 640 feature maps, which are then used to advance to the next stage of the inception module.

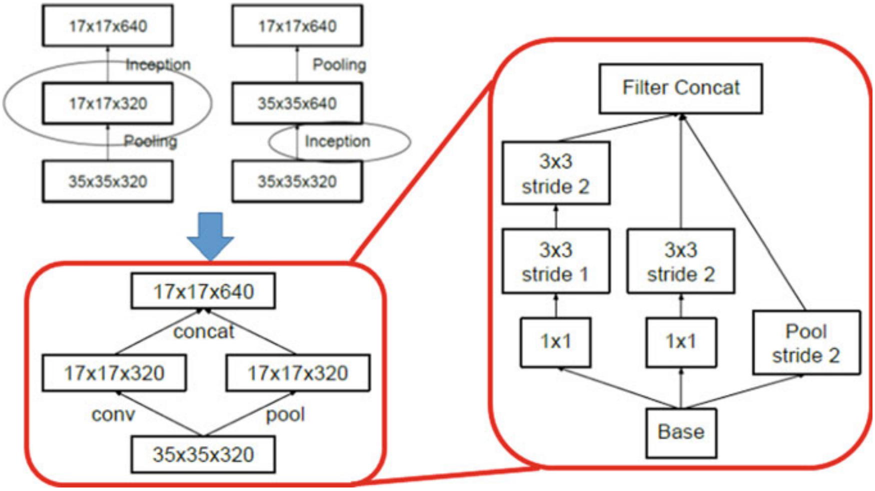


Fig. 5 Conventional downsizing (Top Left), Efficient grid size reduction (Bottom Left), Detailed architecture of efficient grid size reduction (Right)

3.5.2 Optimizer

The current model showcases three flavors of optimizers: SGD, momentum, and RMSProp. RMSprop is a popular optimizer first proposed by Geoff Hinton in one of his lectures. The update dynamics are given by:

$$g_{k+1}^{-2} = \alpha g_k^{-2} + (1-\alpha) g_k^2 \tag{1}$$

$$w_{k+1} = \beta w_k + \eta \frac{\eta}{\sqrt{g_{k+1^{+\in}}^{-2}}} + \nabla f(w_k) \tag{2}$$

For Inception v3, tests show that RMSProp has the best performance in terms of overall precision and time to achieve it, with momentum coming in second. As a consequence, RMSprop has been designated as the default optimizer. The parameters used are:

$$\text{decay } \alpha = 0.9, \text{ momentum } \beta = 0.9, \text{ and } \in = 1.0$$

3.6 Train CNN(Spatial Features) and Prediction

The first row in Fig. 6 is the video of a gesture Hello. The second row shows the set of frames extracted from it. The third row shows the sequence of predictions for each frame by our model.

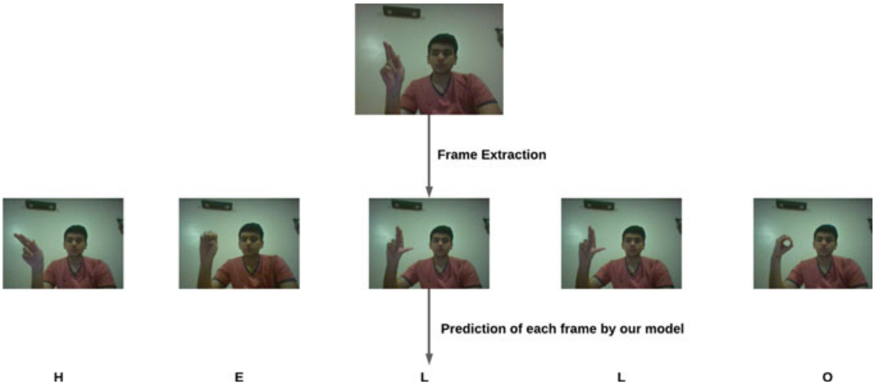


Fig. 6 Illustration of the working on an example

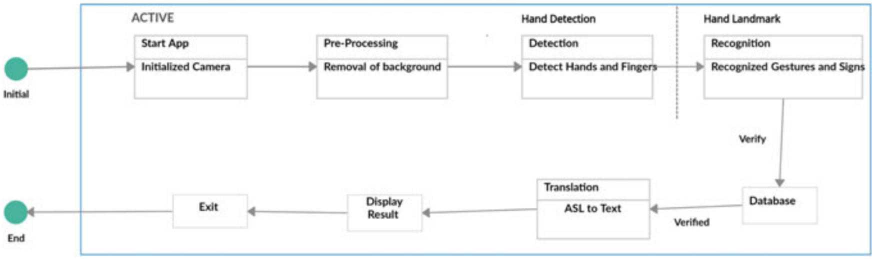


Fig. 7 Illustration of the working on an example

3.7 İmplementation

Figure 7 for hand tracking is shown below. The figure is divided into two sections: one for hand identification and the other for computation of hand keypoints (i.e., landmarks). The palm detector is only run as required (which is fairly infrequently), which saves considerable computation time. We do this by inferring the location of the hand in subsequent video frames from the calculated hand key points in the current frame, rather than having to run the palm detector through each frame. The hand tracking model also produces an additional scalar that captures the belief that a hand is present and fairly matched in the input crop for added robustness. The hand detection model is only reapplied to the whole frame when the trust slips below a certain level.

3.7.1 User İnterface

Figure 8 Home Page: This is the initial page of the app. This page has the option of creating or scheduling a new meeting with a unique invite code/link that can be sent to other users who wish to join the meeting.

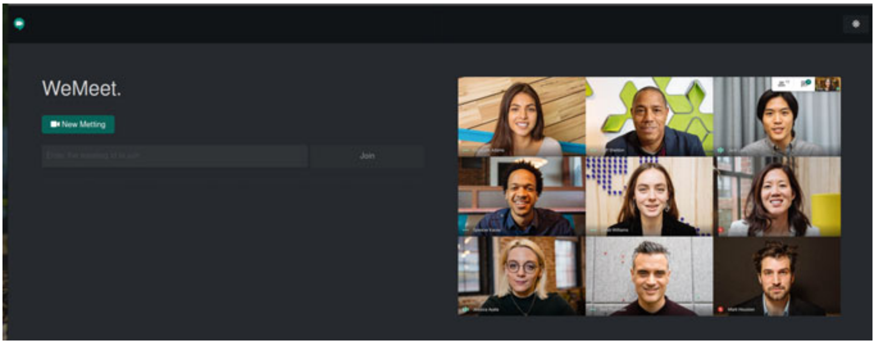
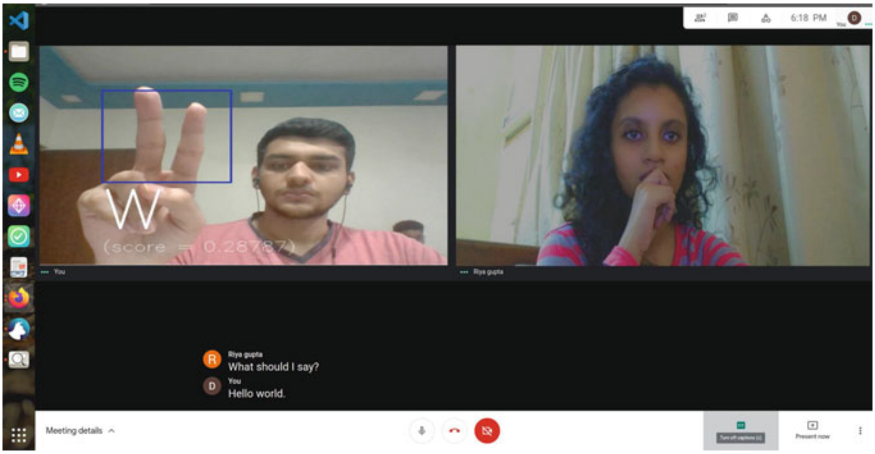


Fig. 8 WeMeet home page



Riya gupta
What should I say?
You
Hello world.

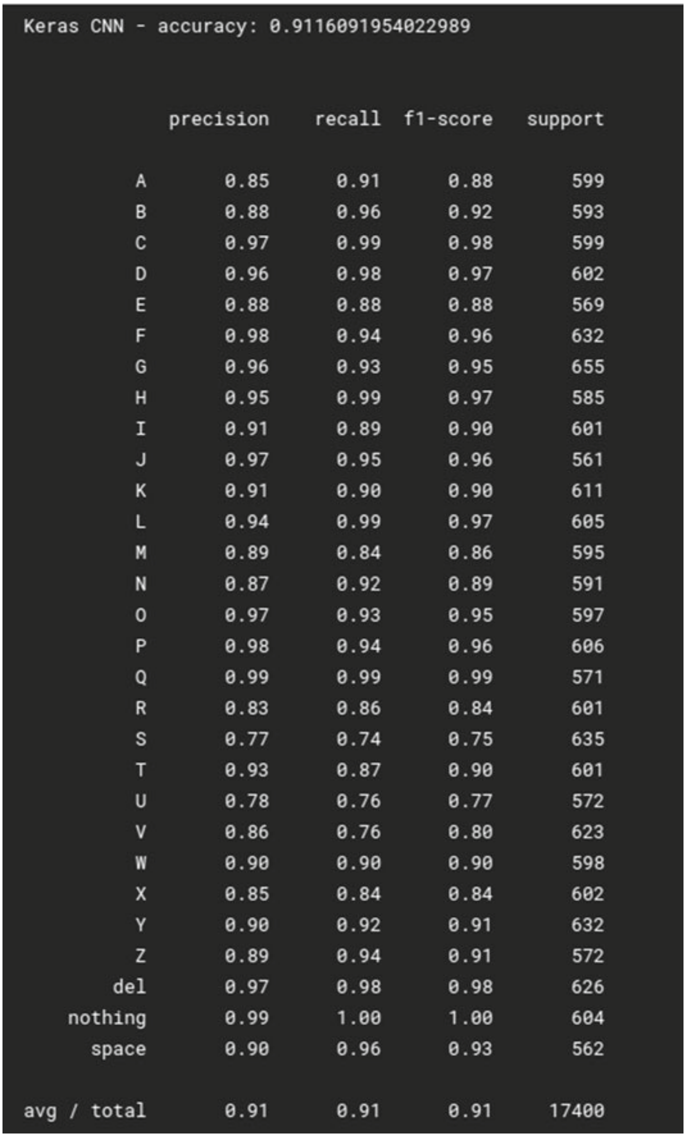
Fig. 9 Sign language conversion to subtitle meet page

Meet Page: Here is the main interaction happening between the people in the meeting through chats, voice, and video interaction with subtitles to gestures. WeMeet app sends live frames of mute person to the server. Hand detection and gesture recognition model are initialized when the app passes the very first frame to the server. The hand detection model and the gesture recognition model run serially; that is, the output of the hand detection model is given as input to the gesture recognition model, and this process continues. The server applies a hand detection model on the frame, which draws a bounding box around the hand. It also crops and pre-processes the bounding box image. After this, the processed frame is given as an input to the gesture recognition model, which returns a tensor array that contains all the output labels along with the confidence/probability. We then sort this tensor array according to the confidence/probability, the output label whose confidence/probability is the highest will get printed on the frame, then the server returns the original frame with the bounding box and the predictions are combined into a single entity as shown in Fig. 9.

4 Result Analysis

The dynamic user interface is designed using the image processing techniques which are implemented in python with the use of OpenCV Library. Experiments reveal that the device works best in low-noise settings (i.e., the presence of stimuli that are identical in color to human skin) and underbalanced lighting conditions.

To prove the feasibility and consistency of the planned system, we ran a series of experiments. A sign image dataset has been developed, as seen in Fig. 10, with 3000 images for each alphabet. The dataset is divided into two sections: a training



```
Keras CNN - accuracy: 0.9116091954022989


             precision    recall  f1-score   support

          A       0.85      0.91      0.88       599
          B       0.88      0.96      0.92       593
          C       0.97      0.99      0.98       599
          D       0.96      0.98      0.97       602
          E       0.88      0.88      0.88       569
          F       0.98      0.94      0.96       632
          G       0.96      0.93      0.95       655
          H       0.95      0.99      0.97       585
          I       0.91      0.89      0.90       601
          J       0.97      0.95      0.96       561
          K       0.91      0.90      0.90       611
          L       0.94      0.99      0.97       605
          M       0.89      0.84      0.86       595
          N       0.87      0.92      0.89       591
          O       0.97      0.93      0.95       597
          P       0.98      0.94      0.96       606
          Q       0.99      0.99      0.99       571
          R       0.83      0.86      0.84       601
          S       0.77      0.74      0.75       635
          T       0.93      0.87      0.90       601
          U       0.78      0.76      0.77       572
          V       0.86      0.76      0.80       623
          W       0.90      0.90      0.90       598
          X       0.85      0.84      0.84       602
          Y       0.90      0.92      0.91       632
          Z       0.89      0.94      0.91       572
        del       0.97      0.98      0.98       626
    nothing       0.99      1.00      1.00       604
      space       0.90      0.96      0.93       562

avg / total       0.91      0.91      0.91     17400
```

Fig. 10 Accuracy of predictions

set of 2000 images and a testing set with 1000 images. There are 26 grades in all. The accuracy of the proposed system is evaluated as follows:

$$\text{Accuracy} = \left(\frac{\text{correctly classified gestures}}{\text{total no. of gestures}}\right) \times 100\% \tag{3}$$

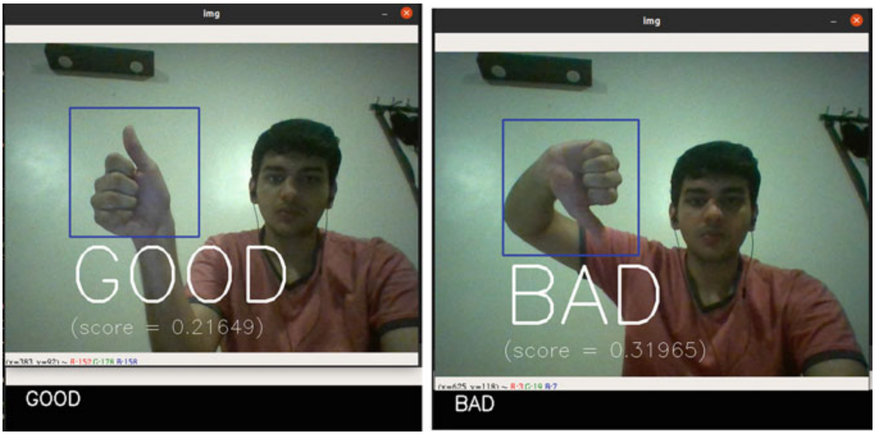

img
GOOD
(score = 0.21649)
(x=383, y=92) ~ R:152 G:178 B:158
GOOD
img
BAD
(score = 0.31965)
(x=625, y=118) ~ R:3 G:19 B:7
BAD

Fig. 11 Results of *GOOD* and *BAD* gestures

Figure 11 Represents the video gestures of *GOOD* and *BAD* being interpreted into subtitles based on the highest confidence score out of the remaining gestures in the image data set is visible on the screen with GOOD having the confidence of 26.6% and BAD having the confidence of 31.965% which is the highest of all

This model was pre-trained on the ImageNet dataset from 2012 and can distinguish 1,000 different forms, including Dalmatian and dishwasher. The algorithm retrains this current model using Transfer Learning to identify a new collection of photographs.

Figure 12 gives a depiction of the number of epochs versus accuracy with losses whereas Fig. 13 shows the number of epochs versus accuracy for different gestures with an accuracy rate of 91.62%.

5 Conclusion

The most significant benefit of using hand gesture-based input modes is that the user can communicate with the program from a distance without having to physically interact with the keyboard or mouse. This paper develops a hand motion recognition framework that uses MediaPipe consisting of several models working together: Palm detector, hand landmark, and gesture recognizer. Inceptionv3 is the neural network model used here for object detection as it gave us the highest accuracy of 91.62% as compared to various models like Resilient back, VGG-16, ResNet50.The proposed system offers a useful approach for creating a user-friendly interface between humans and computers. Moreover, considering the pandemic situation where everyone has shifted to the virtual platform as a means of communication, our system bridges the gap between a normal person and a physically disabled person by interpreting their sign language into captions. The system allows users, including visually disabled

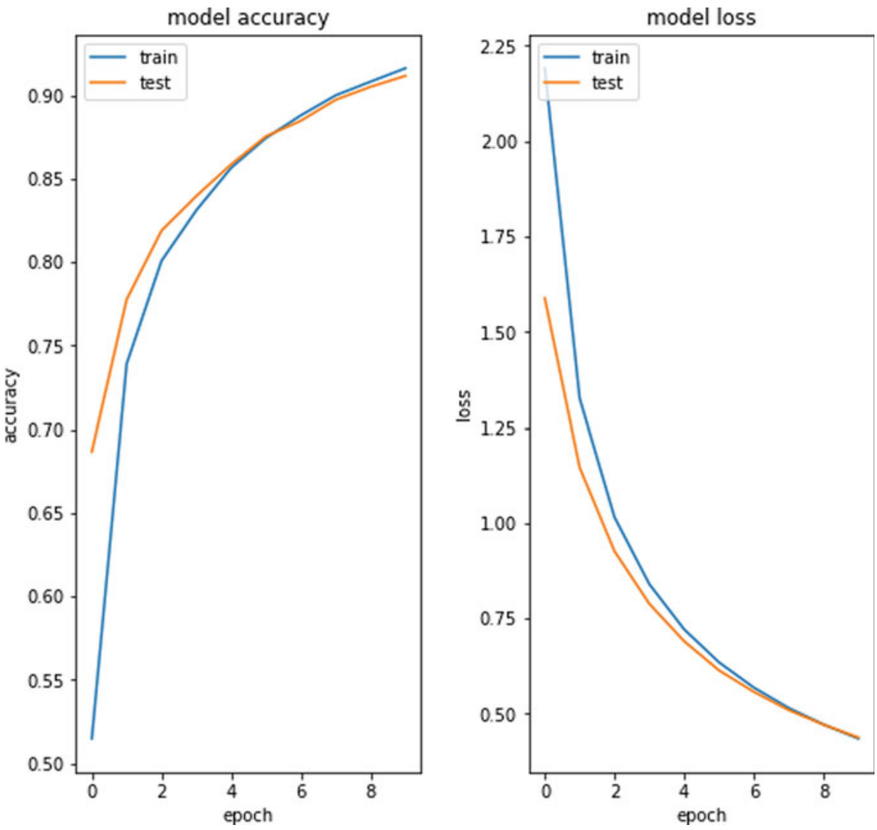


Fig. 12 Number of epochs versus accuracy with losses

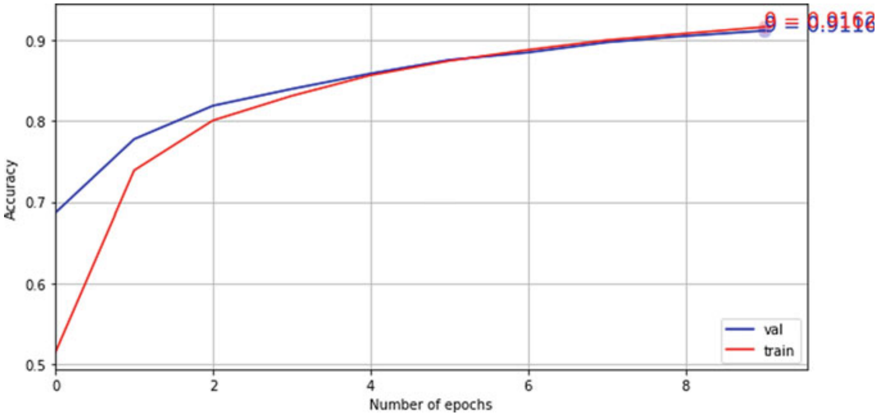


Fig. 13 Number of Epochs VS Accuracy for different gestures

users, to identify gestures based on their viability and ease of use, the system allows them to do so.

6 Future Work and Limitations

We wish to implement Autocorrect, Autocomplete, and Machine translation NLP models on the WeMeet app, The auto-correction model works like a spell checker. It corrects all the common spelling errors. So basically it removes all the spelling errors done by the mute people while performing sign language. Autocomplete tries to predict the next word or suggests the remaining sentence using deep learning. Machine translation is the process of mechanically transforming one language's source text into another language's text. As a result, we can transform the expected output text into any native language. The primary purpose of these models is to help mute people to communicate more effectively and fast. The only limitation we have with our proposed system is the model needs a high-end graphics card and processor to process the frames of the users and predict the gestures.

References

1. E. Niechwiej-Szwedo, D. Gonzalez, M. Nouredanesh, J. Tung, Talking hands—an Indian sign language to speech translating gloves, in *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*
2. S. Srinath, G.K. Sharma, Classification approach for Sign Language Recognition, in *International Conference on Signal, Image Processing, Communication & Automation* (2017)
3. W.T. Man, S.M. Qiu, K. Wong, ThumbStick: a novel virtual hand gesture interface, in *Robot and Human Interactive Communication, 2005* (IEEE International Workshop, Roman, 2005)
4. L.K.S. Tolentino, R. Juan, A. Thioac, M. Pamahoy, J. Forteza, X. Garcia (2019) Static sign language recognition using deep learning. Int. J. Mach. Learn. Comput. **9**, 821–827. https://doi.org/10.18178/ijmlc.2019.9.6.879
5. J. Bukhari, M. Rehman, S. Malik, A. Kamboh, A. Salman, American sign language translation through sensory glove; SignSpeak. Int. J. u- e-Service, Sci. Technol. (2015)
6. B. Fang, F. Sun, H. Liu, C. Liu, D. Guo, Introduction, in *Wearable Technology for Robotic Manipulation and Learning* (Springer, Singapore, 2020). https://doi.org/10.1007/978-981-15-5124-6_1
7. F. Mériaudeau, Sign Language Translator Using Microsoft Kinect XBOX 360 TM, (2012), pp. 1–76
8. M. Abavisani, H. Reza Vaezi Joze, V.M. Patel, Improving the performance of unimodal dynamic hand-gesture recognition with multimodal training, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019)
9. Y. Zhang, J. Cheng, L. Hanqing, EgoGesture: a new dataset and benchmark for egocentric hand gesture recognition. IEEE Trans. multimedia **20**(5), 1038–1050 (2018)
10. B. Toghiani-Rizi, C. Lind, M. Svensson, M. Windmark, in *Static Gesture Recognition using Leap Motion* (2017)
11. G. Ranganathan, Real life human movementrealization in multimodal group communication using depth map information and machine learning. J. Innov. Image Proc. (JIIP) **2** (2020)

12. Communication Foundations Fall 2016, Katie Ludwiczak: CTP Blog #1, http://comm200fall2016.blogspot.com/2016/09/katie-ludwiczak-ctp-blog-1.html
13. P. Keerthana, M. Nishanth, D. Karpaga Vinayagam, J. Alfred Daniel, K. Sangeetha, Sign Language Recognition, in *Special Issue of Second International Conference on Advancements in Research and Development*, vol. 03, no. 03S (2021)
14. Z. Chen, J.T. Kim, J. Liang, J. Zhang, Y.B. Yuan, in *Real-Time Hand Gesture Recognition Using Finger Segmentation,* vol. 2014, no. 267872. https://doi.org/10.1155/2014/267872
15. K.S. Abhishek, L.C.F. Qubeley, D. Ho, Glove-based hand gesture recognition sign language translator using capacitive touch sensor, in *IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)* (2016)
16. C. Thomas, S. Pradeepa, A comprehensive review on vision-based hand gesture recognition technology. Int. J. Res. Adv. Technol. **2**(1) (2014)
17. A.S.L. Kaggle, Alphabet Image Data set for Alphabets in American Sign Language, https://www.kaggle.com/grassknoted/asl-alphabet
18. S. Smys, J.I.Z. Chen, S. Shakya, Survey on neural network architectures with deep learning. J. Soft Comput. Paradigm (JSCP) **2**(03), 186–194 (2020)
19. C. Wang, G. Peng, B. De Baets, Deep feature fusion through adaptive discriminative metric learning for scene recognition. Inf. Fusion **63**, 1–12 (2020)
20. Y. Ding, W.K. Wong, Z. Lai, Z. Zhang, Discriminative dual-stream deep hashing for large-scale image retrieval. Inf. Process. Manage. **57**(6), 1–15 (2020)

Vision-Based Real Time Vehicle Detection: A Survey

Manjot Kaur and Rajneesh Randhawa

Abstract In this era of technology, automation, and speed, security measures have become indispensable part of the research. In the present context, we are concerned about road safety. To provide insights into road safety it is important to study and evaluate the systems which ensure road safety and vehicle detection systems/algorithms lie at the core of this subject. There is a plethora of techniques developed for detection of vehicles however vision-based techniques have become quite popular because they are non-intrusive, simple to install, and efficient by cost. Moreover, advances in computing power to process video data and frames have also favored vision-based object detection. For effective monitoring of the roads, vehicle detection needs to be done in real time. This study reviews some of the prominent vision-based vehicle detection algorithms which are real time by enlisting graphical comparison of Accuracy and detection rates of the prominent algorithms. It is important to discuss the performance evaluation metrics used in history to access the algorithms. Hence brief description of performance parameters is also provided. The present work also mentions important datasets used for vehicle detection and similar studies. The paper concludes by providing significant research directions, mentioning merits and demerits of various state-of-the-art algorithms.

Keywords Vehicle detection · Vision-based vehicle detection · Real time vehicle detection

1 Introduction

In this age of technology, every single task is being automatized resulting in huge research work done every day.

Technology has planted its roots in detecting objects, moving objects, and making objects communicate [1] to each other, accessing parking lots for vehicles [2] and so on. In the present context, we are focused on detection of objects (specifically vehicles).

M. Kaur (✉) · R. Randhawa
Department of Computer Science, Punjabi University, Patiala, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_57

Vehicle Detection in itself is a very crucial and difficult job. Over recent times real time vehicle detection is in the limelights as for real world scenarios to assist a driver, the vehicle detection algorithm needs to be accurate and real time. Thus, this paper reviews several existing real time algorithms for detecting vehicles. Although a lot of work has already been done on this subject however those studies do not focus on real time vehicle detection specifically. Hence this study brings insights into real time vehicle detection algorithms which use vision-based approach to detect vehicles [3].

Detection of vehicles is very significant because of its applications in various domains such as traffic surveillance, industrial research, automated driving vehicles, and accident-avoidance systems. However, what is more important is that here the object (vehicle) detection needs to be done in real time. To make a vehicle put brake on detecting a vehicle in front of it, to track any theft of vehicles or for general surveillance a timely detection and tracking are mandatory. Any failure or inaccuracy in vehicle detection may result in hazardous consequences.

In this study, we will focus on various challenges associated with detection of vehicles, various methods used for the same. Along with this work also sheds light on important performance evaluation metrics used and various datasets used for the detection of vehicles.

1.1 Challenges in Vehicle Detection

Vehicle detection faces a lot of challenges be it poor lighting conditions, bad weather conditions, or hindrance due to man made structures in urban areas or trees, etc. Table 1 summarizes various challenges in the way to effective and smooth vehicle detection [3].

Table 1 Challenges in vehicle detection

Challenge	Description
Illumination-	To deal with high light or very low light conditions. While working with vehicle detection, illumination of area under monitoring plays important role. This makes night time vehicle detection even more complex
Motion of vehicles-	To differentiate between moving and non-moving vehicles. Vehicles move at very high speeds on highways. It becomes quite difficult to detect such fast-moving objects that are too in real time
Occlusion-	Due to large number of vehicles on road, man-made structures, partial light conditions cause occlusion. This causes hindrance in vehicle detection
Varying Size and Color	Vehicles come in variable sizes and colors. Sometimes the color matches with the background and it becomes tough to extract their features. An algorithm which is capable of detecting motorbikes effectively may fail poorly in detection of large vehicles such as trucks and buses. Thus, size of the vehicle is quite significant

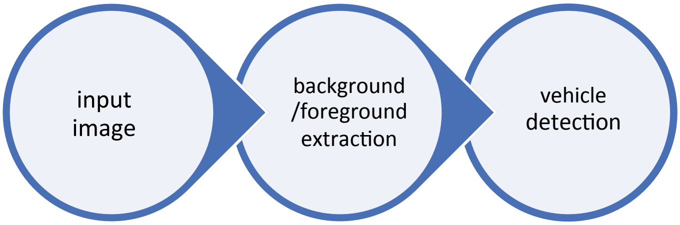


Fig. 1 Block diagram of motion-based vehicle detection

1.2 Classification of Vision-Based Vehicle Detection Algorithms

Vehicle Detection is broadly classified to be done using [4].

1.2.1 Motion-Based Vehicle Detection

Motion-Based vehicle detection uses Frame Differencing and Background subtraction. In Frame differencing, two consecutive frames are considered and their pixel-wise difference is calculated. On the other hand, in Background subtraction, by calculating the difference by pixel between background image and the current image, Foreground objects are extracted. Obtaining Background image is a tough step when the background is dynamically changing (Fig. 1).

1.2.2 Appearance-Based Vehicle Detection

In Appearance-based features, the appearance characteristics of the vehicles are taken into accounts such as color, shape, and texture. This method needs prior information to be employed. Haar filter and the Histogram of Oriented Gradient are used in this method. Motion-Based features are used to detect moving vehicles whereas appearance-based features can be used to detect stationary objects as well. Frame differencing or background subtraction approaches provide fast results but they require either a reference for background or a static image of the background or the background image updated frequently [5, 6]. Table 2 demonstrates vision-based vehicle detection methods.

1.2.3 Significance of Real Time Vehicle Detection

Real time performance of algorithms is demanded for effective implementation of vehicle detection algorithms. Theft control, traffic surveillance, Accident reporting,

Table 2 Vision-based vehicle detection algorithms

Broad categories	Motion based	Appearance based
Prominent **techniques**	• Frame differencing and background subtraction, e.g., optical flow	• Color, symmetry, lights of vehicles, multiple features based, e.g., Haar filter and Histogram of gradient,
Application area	• Used to detect moving vehicles	• Used to detect moving as well as stationary vehicles
Advantages	• Accuracy 60–90% for vehicles moving at 5–20 km/h	• Comparatively less computational load • Works for nighttime detection as well (vehicle lights) • Higher precision Rate
Limitations	• Camera sensitive to movement • Less efficient to detect slow moving objects • Processing of various frames is needed prior to detection • Lot of computational work required	• Performs inefficiently when background color matches with the color of the object When considering vehicle lights, it can confuse with street lights, road lights, etc. Needs Fast hardware

Accident prevention, Automated vehicles, etc. are various subjects where an algorithm with slow computational speed will be of no use hence the significance of vehicle detection algorithms lies in those being real time performing. There is huge number of vehicle detection algorithms developed but not all of them are real time. This paper focuses on reviewing several real time vision-based algorithms.

2 Literature Survey

2.1 Background Modeling/ Subtraction Based Methods

In Wang et al. [7] developed algorithm for real time detection of vehicles, combination of Mixture of Gaussians for background subtraction and H-squeeze net is used. The exact job of MOGs is to develop regions of interest from the video frames. Then identification of the vehicle is done with H-squeeze Net. Dataset used for the study and verification of the algorithms is CDnet2014 and UA-DETRAC datasets. Both MOGs and H-squeeze net are evaluated individually in terms of performance.

For MoGs the value of frames per second will be pre-defined. for H-squeeze Net the evaluation protocol is given by following Eq. (1)

$$\mathrm{F-Measure} = (2 * \mathrm{Precision} * \mathrm{Recall})/(\mathrm{Precision} + \mathrm{recall}) \qquad (1)$$

Though MoG is a slower preprocessing algorithm, it proved to be robust and Accuracy of H-Squeeze Net approached 98.3%. However, when taken both together the system performed with an accuracy of > 95%, and time is 39.1 FPS which is enough to cater real time requirements.

Another work [8] used raspberry pi with image processing paradigms to monitor the traffic flow. The work is implemented in python language. Linux Server is used for storing the recorded data and surveillance. Raspberry pi is programmed to record when motion occurs in the area under surveillance. The implementation results witness that the vehicle detection accuracy reached 97/1% and being a hardware system, it is fast in computations. The proposed method is cost effective [9]. It uses a monocular camera to acquire the images of area under monitoring. For detection of vehicles, it first extracts the ROIs, processes them for foreground extraction. It uses Sobel's edge detection and Otsu thresholding.

2.2 Symmetry and Color Based

The vehicle detection is by utilizing the symmetry feature of the vehicle. So developed method of detecting vehicles locates vehicle headlights with the help of image processing (image segmentation and pattern recognition). With the help of these image processing techniques, the regions of interest are extracted. These bright ROIs are then processed by spatial clustering that analyzes the spatial features of vehicle lights and this figures out the presence of vehicles (moving cars and bikes).

The significance of this method lies in the fact that it is useful for nighttime vehicle surveillance [10]. Block diagram of the algorithm is given in Fig. 2.

2.3 Harr and Hog Features Based

The proposed method [11] uses combination of Harr and HOG making a two-step vehicle detection algorithm. This combination utilizes the HOG's descriptive ability. Harr features are used to extract the areas of interest effectively. Further to reach a promising stage of detection, Adaboost classifier is also used on the features extracted by Histogram of Gradient.

Another research proposed [12] a real time technique for detecting vehicles that utilized the feature extraction capability of harr with the classifying capability of ANNs (Artificial Neural Networks). The concept of integral image is used to represent the whole image for quick computation. Number of neurons is the key parameter in this algorithm. This algorithm provides an accuracy of 91.2% with execution time of 0.59 s given the number of neurons 200. Block diagram of the process is given in Fig. 3.

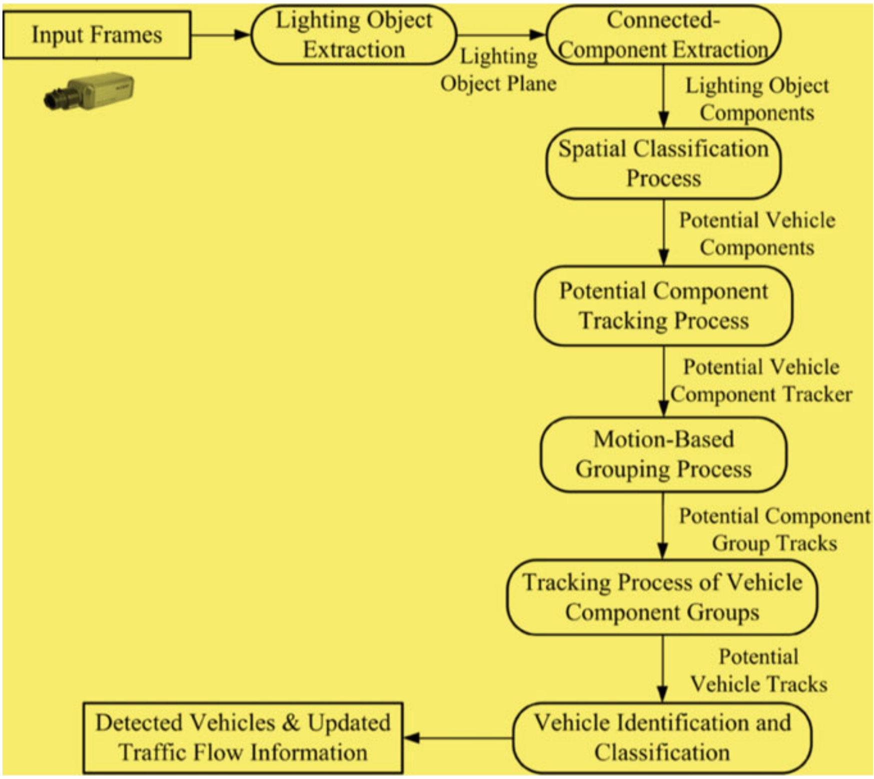

Input Frames
Lighting Object Extraction
Lighting Object Plane
Connected-Component Extraction
Lighting Object Components
Spatial Classification Process
Potential Vehicle Components
Potential Component Tracking Process
Potential Vehicle Component Tracker
Motion-Based Grouping Process
Potential Component Group Tracks
Tracking Process of Vehicle Component Groups
Potential Vehicle Tracks
Vehicle Identification and Classification
Detected Vehicles & Updated Traffic Flow Information

Fig. 2 Block diagram of the algorithm [10]

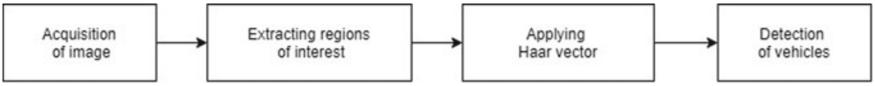


Fig. 3 Block diagram of Haar based vehicle detection [11]

2.4 Optical Flow Based

It is quite complex to detect moving objects in real time in an unconstrained environment. Huang et al. [13] the proposed system uses an optical flow-based technique for the same. Homography matrixes are utilized for building background models (optical flow) online making the algorithm adapting to various varying circumstances.

2.5 Histogram Equalization Based

The proposed algorithm vehicles are both detected as well as tracked. Image processing algorithms such as histogram equalization along with average edge images are used to locate the rear box of vehicle. Vehicle's shadow features are used to define region of interest [14].

2.6 Vehicle Detection Involving Occlusion

Another work [15] brings forth a real time algorithm for detection and tracking of vehicles while taking into account the presence of occlusion. Height and placement of camera can cause occlusion and this leads to inaccuracy in the results. Authors here handled occlusion with the help of some assumptions which are**:**

- Width of vehicle cannot be larger than the width of two lanes irrespective of the vehicle being a large vehicle.
- Vehicle cannot be wider than the width of a single lane. This makes the vehicle is completely in the region of interest.

Significance of handling occlusion is to reduce the effect of occlusion. For detection, using Gaussian MM the vehicles are segmented from the background. While doing so some of the relevant features such as size dimensions of vehicles (breadth and height, etc.) are extracted. Further Kalman Filter is used to perform the tracking.

In proposed algorithm [16], vehicle detection is carried out by segmenting moving objects with the help of color background. This algorithm detects and tracks the vehicles while taking into account the occlusion. If occlusion occurs it is handled by rule-based reasoning.

2.7 Contrast Analysis Based

Vehicle detection is a widely researched subject however when nighttime vehicle detection still has a large room for improvements as it is very significant for vision-based surveillance. In this work [17] contrast analysis method is used to detect vehicles during dark hours. Here the contrast for local change is considered to determine the presence of large moving objects. False positives are reduced with the help of spatial nearest neighbor data association and motion prediction. Practical implementation demonstrates the effective execution of the algorithm for vision-based real time detection of vehicles for nighttime.

2.8 Deep Learning Based

When considering deep learning based vehicle detection algorithms, the state-of-the-art algorithms are divided into two broad categories (1) Two-stage object detection (2) One-stage object detection.

In two-stage object detection, the image is processed in two stages, first the candidate frames are formed, and then the model is trained. Prominent examples of such methods include faster RCNN and R-FCN [18, 19].

In one-stage object detection, the intermediate region detection is absent. The results are extracted directly from the input image. Examples of such methods include SSD, YOLOv3, etc. [20, 21]. In these, the image is sampled at various positions uniformly. After sampling, a CNN is used to extract features from the image. This whole process is completed in one scan of the image Table 3.

3 Discussion

3.1 Datasets for Vehicle Detection

Table 4 contains the names and descriptions of the datasets which are used for the study of vehicle detection by various researchers.

3.2 Performance Evaluation Metrics

In the papers reviewed in this study, several performance parameters are used to evaluate the performance of the algorithms. Important performance evaluation metrics are:

- *Detection Rate*—It gives the number of correctly detected objects to actual present objects. Higher the detection rate better is the performance of the method
- *False Positive*—If the object is detected to be present when in actual it is not present It is counted as false positive. Lesser the value of False positive, greater is the accuracy of the algorithm
- *False Negative*—When an object is present, but method fails to detect it. Value of this parameter should also be less
- *True Positive*—It gives the data about correctly detected objects.
- *Operational Time*—Time taken by the algorithm to detect the object is called operational or computational time [34].
- *Jaccard coefficient*—This is popularly used in information retrieval algorithms to know the similarity between two different sets of the population (of objects). In the present context, jaccard coefficient gives the accuracy of the algorithm.

Table 3 Summary of vehicle detection algorithms

Reference	Method	Year	Accuracy/detection Rate	Operational Time	F-measure	Comments
[7]	Background Subtraction and MOG with H-Squeeze net model	2020	98.93%	10 Ms for vehicle identification	x	Gives Detection speed of 38.1FPS, it can be improved by upgrading slow computational speed of MOG
[10]	Locating Headlights and Tail-lights	2011	97.9%	12 Ms	X	It can be improved by fusing with machine learning techniques for enhanced performance
[22]	Fusion of Camera data and 3D-LIDAR	2017	64.77%	0.063 s	X	–
[11]	Harr and HOG features	2017	97.96%	137 ms/frame	X	Combination of HOG and Harr gave better results than conventional methods
[8]	Raspberry pi with computer vision	2020	97.39%	–	X	Cost efficient and suitable for real time works,
[12]	Harr features + Artifical Neural Networks	2015	91.3%	0.59 s	x	This performance is recorded when number of neurons were 150–200
[9]	Foreground Subtraction	2011	95.8%	0.16 s	x	–
[22]	Yolo v3 based	2017	94.97%	–	x	–

(continued)

Table 3 (continued)

Reference	Method	Year	Accuracy/detection Rate	Operational Time	F-measure	Comments
[14]	Shadow feature and histogram equalization	2012	94%	–	x	–
[15]	Vehicle detection with occlusion handling	2018	95.21%	–	94.014	–
[17]	Contrast in Local change is utilized	2008	96.68%	–	x	Can be enhanced by using multiscale region detection similar to face detection
[24]	Vision and Lidar point cloud fusion	2011	70.58%	–	x	–
[25]	Improving faster RCNN	2019	89.20%	–		–

Table 5 Enlists various datasets and performance parameters used for evaluating performance of vehicle detection methods.

3.3 Comparison of Algorithms

Various Vehicle detection algorithms have been studied. Study shows that fusion of technologies gives better results as compared to these technologies when implemented individually. Figure 1 gives the graphical comparison of several real time vehicle detection algorithms.

For example, Harr and HOG fusion gives an accuracy of 97.9% which is pretty good. Similarly, fusion of Harr and ANNs also performed well. Operational times of these algorithms are suitable for these to be implemented for real time responses. Such systems can be implemented to reduce number of vehicle-to-vehicle collisions.

Table 4 List of datasets

Reference	Dataset	Description
[26]	KITTI	This dataset contains highways images and simple road scenes. This can be useful for three-dimensional object detection and tracking
[27]	Stanford Car dataset	The dataset has over 19,000 categories of vehicles
[28]	Comprehensive Cars Dataset	This is similar to Stanford car database
[29]	BIT vehicle Database	In BIT vehicle dataset, we have 9850 images, however, this database is not suitable for CNN training as the vehicle size in image is very small
[30]	The traffic and Congestions (TRANCOS) dataset	The dataset has a total of 1244 images of vehicles
[31]	CD net 2014	CDnet2014 dataset has 53 videos containing various challenges
[31]	UA-DETRAC	The UA-DETRAC dataset consists of 100 videos sequences
[32]	PASCAL VOC	The dataset contains images of various categories of objects such as vehicles, animals, household objects, people, etc.
[33]	LSVH dataset	This dataset consists of 16 videos recorded under various secenes and weather. The video data is divided into two groups—sparse and crowded

4 Conclusion and Future Research Directions

Through this paper, we have provided a survey of literature focusing on vision-based real time algorithms for vehicle detection. In the context of vision-based vehicle detection, Background/ foreground extraction, feature based extraction algorithms such as harr and Histogram of gradient are reviewed. Further fusion of Vision and LIDAR based technologies was also part of the study.

To summarize Optical flow-based vehicle detection can be stated that it involves high computational load and comparatively low detection rate. It sometimes fails to detect slow moving vehicles. Similarly, symmetry-based vehicle detection requires a rough estimate of the location of the object thus it requires some preprocessing to extract the ROIs which adds to the computation (Fig. 4).

In the Fig. 4, 'BK sub' stands for background subtraction, 'HL and TL' denotes the Headlight and taillights detection based vehicle detection, ANN stands for Artificial Neural Network and 'FG sub' stands for Foreground subtraction based vehicle detection, rest are self explanatory. Apart from this, recent advances in one-stage vehicle detection such as YOLO perform pretty well in object detection. These

Table 5 Performance parameters used

Publication	Performance metric used
[7]	F-measures were utilized as evaluation proto-cols F-Measure = (2*Precision*Recall)/ (Precision + recall)
[10]	Jaccard coefficient J = true positive/(true positive + false positive + false negative)
[11]	Detection rate False positive rate Non detection rate Detection time
[12]	Detection rate Execution Time
[14]	Detection Rate
[16]	Accuracy/detection rate
[17]	Jaccard coefficient
[35]	Detection rate False alarm ratio

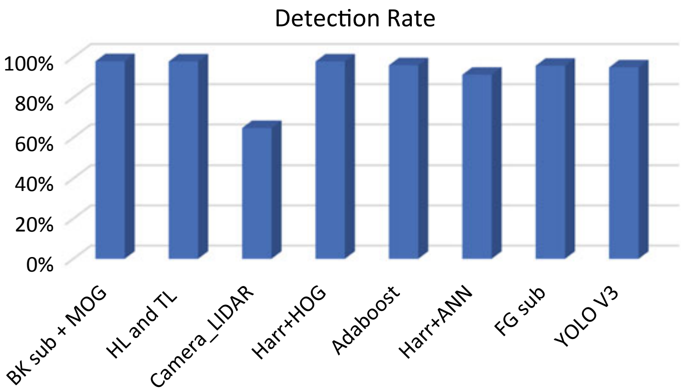


Fig. 4 Comparison of detection rates

algorithms are faster and efficient. However, the drawback of these algorithms is the rigorous training of the models. Images in the dataset need to be annotated for proper object classification. Finally, we conclude that vehicle detection using vision has achieved an accuracy of 98% approximately by using fusion of Harr and Histogram of gradient for feature extraction. Operational time of this technique is 137 ms per frame which is suitable for real time implementation of the algorithm. Further, another technique used bright area detection, i.e., headlight and taillight features to detect vehicles in night hours. This technique provided an accuracy of 97.9% approximately in computational time of 12 ns which is very promising for real time applications. These algorithms further can be enhanced in fusion with machine learning techniques to provide even better performance.

References

1. D.K. Kamel et al., Tenancy status identification of parking slots using mobile net binary classifier. J. Artif. Intell. Capsul. Netw. **2**(3), 146–154 (2020)
2. S. R., D.M., Concept of Li-Fi on smart communication between vehicles and traffic signals. J. Ubiquitous Comput. Commun. Technol. **2**(2), 59–69 (2020)
3. S. Sivaraman, M. Trivedi, Looking at vehicles on the road: a survey of vision-based vehicle detection, tracking, and behavior analysis. IEEE Trans. Intell. Transp. Syst. **14**(4), 1773–1795 (2013)
4. R. Chandran, et al, A review on video-based techniques for vehicle detection, tracking and behavior understanding, Int. J. Adv. Comput. Electron. Eng. **02**(05) 07–13 (2017); Electron. Eng. **02**(05), 07–13 (2017)
5. Z. Sun et al., On-road vehicle detection: a review. IEEE Trans. Pattern Anal. Mach. Intell. **28**(5), 694–711 (2006)
6. M. Fathy, M.Y. Siyal, An image detection technique based on morphological edge detection and background differencing for realtime traffic analysis. Pattern Recogn. Lett. **16**(12), 1321–1330 (1995). https://doi.org/10.1016/0167-8655(95)00081-X(1995)
7. Z. Wang, J. Huang, N.N. Xiong, X. Zhou, X. Lin, T.L. Ward, A robust vehicle detection scheme for intelligent traffic surveillance systems in smart cities. IEEE Access **8**, 139299–139312 (2020)
8. A.P. Kulkarni, V.P. Baligar, Real time vehicle detection, tracking and counting using Raspberry-Pi, In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (IEEE 2020), pp. 603–607
9. Y.C. Kuo, N.S. Pai, Y.F. Li, Vision-based vehicle detection for a driver assistance system. Comput. Math. Appl. **61**(8), 2096–2100 (2011)
10. Y.L. Chen et al., A real-time vision system for nighttime vehicle detection and traffic surveillance. IEEE Trans. Ind. Electron. **58**(5), 2030–2044 (2011)
11. Y. Wei et al., Multi-vehicle detection algorithm through combining Harr and HOG features. Math. Comput. Simul. **155**(2018), 130–145 (2019)
12. A. Mohamed, A. Issam, B. Mohamed, B. Abdellatif, Real-time detection of vehicles using the haar-like features and artificial neuron networks. Procedia Comput. Sci. **73**, 24–31 (2015)
13. J. Huang, W. Zou, J. Zhu, Z. Zhu, Optical flow based real-time moving object detection in unconstrained scenes. *arXiv preprint* arXiv:1807.04890. (2018)
14. Y. Chong, et al. Integrated real-time vision-based preceding vehicle detection in urban roads. Neurocomputing (2012)
15. R. Velazquez-Pupo, A. Sierra-Romero, D. Torres-Roman, Y.V. Shkvarko, J. Santiago-Paz, D. Gómez-Gutiérrez, M. Romero-Delgado, Vehicle detection with occlusion handling, tracking, and OC-SVM classification: a high performance vision-based system. Sensors **18**(2), 374. (2018)
16. S.P. Lin, Y.H. Chen, B.F. Wu, A real-time multiple-vehicle detection and tracking system with prior occlusion detection and resolution, and prior queue detection and resolution, in *18th International Conference on Pattern Recognition (ICPR'06),* vol. 1 (IEEE, 2006), pp. 828–831
17. K. Huang, L. Wang, T. Tan, S. Maybank, A real-time object detecting and tracking system for outdoor night surveillance. Pattern Recogn. **41**(1), 432–444 (2008)
18. R. Girshick, et al., Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proceeding IEEE Computer Social Conference Computer Visual Pattern Recognit.* (2014), pp. 580–587

19. R. Girshick, Fast R-CNN, in *Proceeding. IEEE International Conference Computer Visual 2015 International Conference on Computer Vision, ICCV 2015* (2015), pp. 1440–1448
20. J. Redmon, et al., You only look once: Unified, real-time object detection, in *Proceeding. IEEE Computer Social Conference Computer Visual Pattern Recognition* 2016-December, 779–788 (2016)
21. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in *Proceeding—30th IEEE Conference Computer Visual Pattern Recognition, CVPR 2017*. 2017-January, 6517–6525 (2017)
22. A. Asvadi et al., Multimodal vehicle detection: fusing 3D-LIDAR and color camera data. Pattern Recognit. Lett. **115**, 20–29 (2018)
23. H. Song, H. Liang, H. Li, Z. Dai, X. Yun, Vision-based vehicle detection and counting system using deep learning in highway scenes. Euro. Trans. Res. Rev. **11**(1) (2019)
24. H. Wang, et al., Real-time vehicle detection algorithm based on vision and LiDAR point cloud fusion. J. Sensors. 2019, (2019)
25. H. Nguyen, Improving faster R-CNN framework for fast vehicle detection. Math. Probl. Eng. 2019, (2019)
26. Z. Luo, Traffic analysis of low and ultra-low frame-rate videos, Doctoral dissertation. Universite de Sherbrooke (2018)
27. V.F. Kuzishchin, V.A. Dronov, Traffic-sign detection and classification in the wild. Therm. Eng. **48**(10), 835–841 (2001)
28. L. Yang, et al., A large-scale car dataset for fine-grained categorization and verification, in *Proceeding IEEE Computer Social Conference Computer Visual Pattern Recognition* (2015), pp. 3973–3981
29. B. Hicham, et al, Vehicle type classification using a semi supervised convolutional neural network. Colloq. Inf. Sci. Technol. Cist. (2018), pp. 313–316
30. R. Paredes, et al., Pattern recognition and image analysis, in *7th Iberian Conference, IbPRIA 2015 Santiago de Compostela, Spain, June 17–19, 2015 Proceedings. Lect. Notes Computer. Sci.* 9117, September, (2015)
31. Z. Wang et al., A robust vehicle detection scheme for intelligent traffic surveillance systems in smart cities. IEEE Access. **8**, 139299–139312 (2020). https://doi.org/10.1109/ACCESS.2020.3012995
32. K.V. Sakhare et al., Review of vehicle detection systems in advanced driver assistant systems. Arch. Comput. Methods Eng. **27**(2), 591–610 (2020)
33. X. Hu et al., SINet: a scale-insensitive convolutional neural network for fast vehicle detection. IEEE Trans. Intell. Transp. Syst. **20**(3), 1010–1019 (2019)
34. F. Bashir, F. Porikli, Performance evaluation of object detection and tracking systems, in *Proceedings 9th IEEE International Workshop on PETS* (2006), pp. 7–14
35. W.C. Chang, C.W. Cho, Online boosting for vehicle detection. IEEE Trans. Syst. Man, Cybern. Part B (Cybern.) **40**(3), 892–902 (2009)

Performance Comparison of Anomaly Detection Algorithms

Maya Manish Kumar and G. R. Ramya

Abstract In several studies, machine learning techniques that are repossessed in intrusion detection systems wangle wide-ranging acknowledgment by changing into a high-yielding domain and continues to be the main target of the researcher's vast significance. After several years of study, the intrusion detection community still faces difficult issues. During the process of detecting unexpected new attacks, reducing the high rate of false alarms remains an unanswered problem. Identification of anomalies is a key problem in malware detection in which the existence of planned or unintentional caused assaults, defects elsewhere is demonstrated by disturbances of normal conduct. This paper offers a top-level view of analysis directions for the utilization of tagged and untagged information to handle the difficulty of the identification of anomalies. By performance comparison of the available semi-supervised and unsupervised algorithms, you can select the best anomaly detection algorithm. The documented references would cowl the most theoretical issues, leading the researcher in new ways for study.

Keywords Anomaly detection · Performance · Comparison · Intrusion detection · Anomalies · Outliers · Tagged data · Untagged data · Semi-supervised · Unsupervised · Attacks · Intrusions

1 Introduction

Persons that act outstandingly or have odd properties are deviations or outliers [1]. The priority of finding these points or patterns is mentioned as the identification of abnormal data or anomalies. The elemental of the popularity of anomalies/abnormal

M. Manish Kumar (✉) · G. R. Ramya
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2cse19012@cb.students.amrita.edu

G. R. Ramya
e-mail: gr_ramya@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_58

information lies in varied domains, such as irregular mercantilism of day stocks in some days [2].

For about 20 years, intrusion detection has been researched. Intrusions are actions that breach the security policy of information technology, and intrusion detection is the method of detecting Intrusions. Detection of infiltration is based on the premise that the actions of the attacker would be substantially different from legitimate behaviors, which promotes and allows several non-authorized actions to be identified.

Network security systems are typically used as an inferior defense score to prevent info technology together with alternative security systems like network access and authorization. There are several explanations why the main components in the entire defense mechanism are detected by the intrusion. First of all, while attaching great importance to safety, many conventional services and equipment have been designed and developed. Second, in the design, computer systems and applications may have vulnerabilities, and intruders can use these vulnerabilities to target systems or applications. Therefore, prevention methods may not be that successful as being expected.

It is possible to separate fraudulent transactions or outliers into two big classes: outliers, large, and small global outlier could be a constant finding that includes a high or low price corresponding to different observations within the take a look at. A local outlier, on the other end, is a test point with a value beyond the average limits of the whole dataset, but it is normally high or low relative to neighboring points. An effective fraud detection system must be capable of reliably detecting fraud and changing its efficiency focused on adjustments in the actions of swindlers.

This analysis is focused on real-life user information from a foreign credit card, kdd cup99, and datasets of the stock market. Frauds are barely comparable to the overall volume of payments, so under-sampling methods were carried out to provide balanced classes for comparisons due to having a not labeled dataset [3]. In addition, while unsupervised algorithms, data labeling has been used in this study for the quality assessment of different approaches and does not include labeled data across the entire dataset.

2 Related Work

Machine learning methods are mainly techniques used to detect fraud. These approaches may be split into two classes: approaches that are semi-supervised and unsupervised. A model for a previous sample of suspicious and legitimate activities will be learned in supervised machine learning methods to identify new purchases as fake or valid. The labels are recognized when the mismatch of a transaction has been detected by cardholders, or an odd activity is detected by a card company and verified by a card issuer. The semi-supervised approaches have the drawback that these frameworks maybe cannot be able to recognize centered on the old assumption if scammers alter their patterns [4].

But on the other hand, data from previous blocks is acquired through unsupervised strategies and the discrepancies are based on modified transactions. Anomalies or irregular activities in unsupervised fake detection should be known as potential cases of suspicious purchases. The benefit of unsupervised learning is that the computer would not need to train itself on the awareness of the fake labels, and a decision will be taken based on the distribution of the transactions to define an activity as an outlier. For much of the unsupervised approaches, however, normal transaction labeling is needed so that machine learning techniques are learned on normal instances so that for financial events it can distinguish between natural and fake [5].

Research to compare the efficiency of various semi-supervised and unsupervised methods for studying the identification of credit card fraud [4]. As unsupervised techniques, 4 unsupervised abnormal detection approaches were used, one per SVM, limited Boltzmann computer, simplified adversarial system. The efficiency of various models was measured using the region under the curve (AUC).

To identify credit card fake, the hidden Markov model [6] is used. Initially, the classifier was tested on a cardholder's regular activity and after tested on arriving card payments. If a forthcoming card activity is not approved with a strong likelihood by the qualified HMM, it will be regarded as a fraud and to detect credit fraud [5], Semi-supervised and unsupervised machine learning anomaly detection techniques were mixed. The outputs manifest that the suggested approach is effective and could enhance prediction performance [7].

The lack of another SVM owing to its high degree of true alarm rate was discussed [8]. A new method called enhanced SVM was then proposed. The genetic algorithm (GA) has been used to extract customized information from the raw dataset as an attribute selection tool.

To identify abnormal issues, a single-class SVM was used [9]. As this technique is vulnerable to outliers, this paper implemented the ramp loss function to solve this problem. The aim of this role was to be performed by constructing a supervised, diffuse method. An increase in outlier identification was seen in the obtained results. Deterministic fraud detection (PFD) was contrasted with the method, and the findings show that PAD could have been surpassed. The algorithm of the SVM because of hieratic prior integration into the PAD algorithm. For the identification of credit card fraud [10], similar machine learning methods were compared, such as SVM, decision tree, and the paradigm of regression analysis. Based on different parameters, such as accuracy, the production of the model was contrasted, specificity, and accuracy. For the identification of unauthorized credit card transactions [11], two- and one-class SVMs have been used and contrasted. Using various kernels, these variables were compared and measured. The findings showed the supremacy of one-class SVM over 2-class SVM for the object tracking problem.

In the case of studies using autoencoder for outlier detection, autoencoder has been used as an object tracking tool in installing solar data based on the ensemble model [12]. A distinction was made between other autoencoder methods' ensembles. The limit of usual against abnormal findings was predicated on the premise that abnormal candidates account for 5 percent of the results. The detection of credit card theft has been suggested using periodic autoencoder and nonlinear autoencoder (VAE), defined as an autoencoder variant that uses a deterministic graph as a basis for the findings of anomalies [13]. The subsequent formation was used for the autoencoder and a restoration likelihood as an exceptional score. A simple periodic autoencoder was outperformed the VAE for card fake detection [12]. In this research, an autoencoder algorithm based on ensemble technique has been progressed to detect abnormal data or anomalies with various architectures stacked with several autoencoders and autoencoder-based ensembles. As a metric for the model assessment, root mean squared was used.

Mahalanobis or Cook's distance is among the order to accomplish multivariate outlier detection (MVO). Within the literature review, strong geometrician distance was used widely for anomaly detection. Once Mahalanobis is employed for MVO, the univariate outliers [14] are thought-about a broad (squared) Mahalanobis distance. However, because of traditional pure mathematics tolerance or sample correlation matrices resistance to isolated incidents [15], the geometrician distance is knowing about the existent of anomalies. The answer is also accomplished along with robustly approximating the average and correlation coefficients network, resisting the impact of peripheral observations [14]. As a consequence of obtaining a computationally fast algorithmic rule, the foremost usually used univariate positioning and scattering approximation technique is that the minimum variance determinant (MCD). This matrix is about the mistreatment of the measuring set of length h, which minimizes the variance matrices determinant.

In exploring geochemistry, multivariate outlier identification was used. The system was able to differentiate between the severe confidence interval values collected from the various distribution values. In this analysis, reliable estimates were used by Mahalanobis to reduce the effect of extreme values on the solution. A univariate outlier plot was implemented to facilitate the spatial representation of the isolated incidents on a map, to illustrate the significance of the distance from the central, various colors were also used. For the explanation of the multivariate outlier for the elemental dataset [16], the same approach was used. Before performing any research, the isometric logratio conversion was applied to the data.

3 Research Methodology

All outlier intrusion detection systems have a basic concept. The training stage attempts to use manual or automated methods to simulate the system. The database is a site for the client–server framework that keeps waiting for connection requests. The database can quickly add a socket when a relationship is formed between the sender and receiver, which will be used to initialize a controller object that runs on a different post. In combined methods, these handlers will be maintained.

Based on the method used, the activities described in the model can differ. Result date the address immediately with the selected user-defined information field in the training phase. To assess an irregular data instance [17], threshold parameters will be identified (Fig. 1).

Machine learning will automatically build an appropriate model based on some of the datasets provided. The availability of the specified training information is the

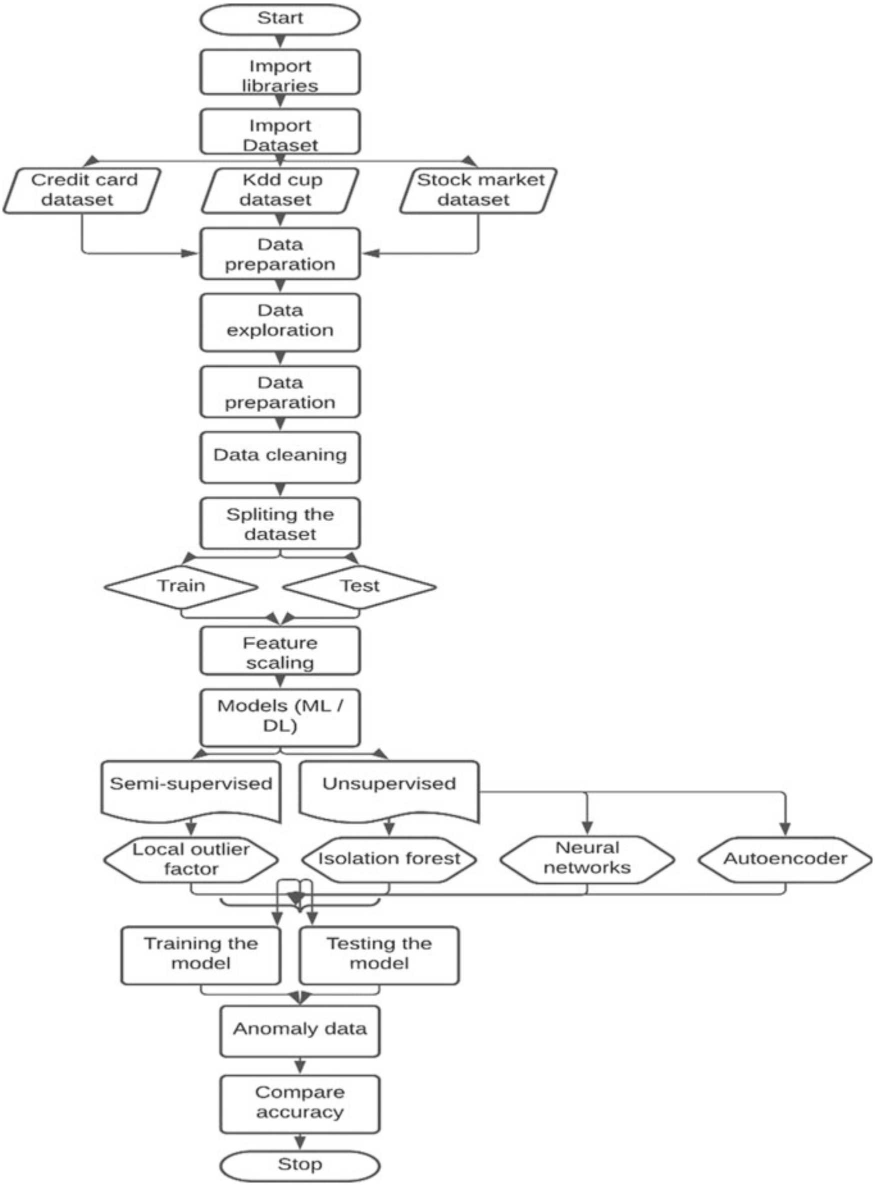


Fig. 1 Work analysis

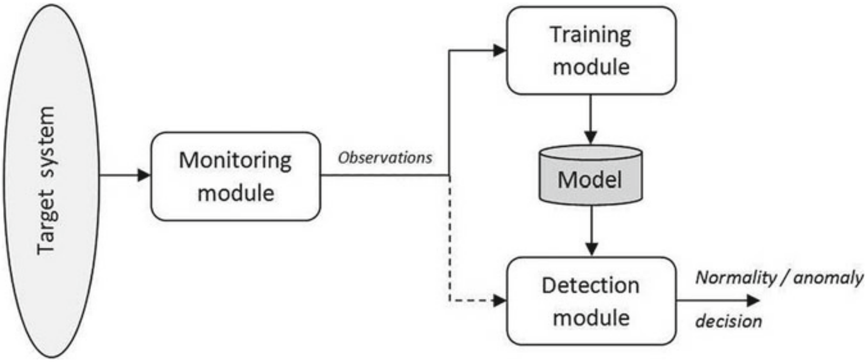


Fig. 2 Architecture of anomaly detection

reason for this strategy, or at the very least that it is easier to access than the work required to expressly outline the model. With the increase in complexity and, as a result, the emergence of various attacks, the only solution to comprehend the subsequent stage is machine learning technology, which allows building and conservation anomaly detection systems (ADS) with slighter individual participation (Fig. 2).

3.1 Semi-supervised Algorithm for Anomaly Detection

To build prediction models, a monitored development set consisting of each nonlinear partial example needed to be developed. Technically, supervised methods have been linked to increase sensing exactness in nursing compared to semi-supervised regression and unsupervised methods because they have access to a wealth of information. However, some technical issues make these methods seem less effective than they claim to be. The main downside is the lack of development data knowledge covering all the regions. The most important downside is that, moreover, getting correct labels is also a challenge, and therefore, the development usually involves certain noises that end up in a range of false-positive associates in nursing. In addition to these semi-supervised learning goals, the unlabeled data points are designated using knowledge gained from a small number of data points. The local outlier factor (LOF) [18] is the most common semi-supervised algorithm. LOF can also be treated as an unsupervised model of learning.

3.1.1 Local Outlier Factor (LOF) (Density-Based Technique)

A local density estimate-based detection of anomalies may be local outlier (LOF) detection. It can be summarized as three actions: First, concerning object o, stay in nursing object p the accessibility distance of the associate; second, which stays in the local reachability of the nursing object p [19]. Finally, the calculation of the local problem outlining object p intimately explains the tactics. Distance live will increase machine prices, notably for information at the high-dimensional house, under the initiative, estimating the accessibility distance.

We plan the instance in feature code using mounted bits to avoid high reason and communication costs and locally sensitive computing (LSH). The LSH algorithm rule is usually used for pattern matching and k nearest search. The LSH algorithmic rule is used to reduce spatial property for high-dimensional information. Connected instances of information projected to a small area fall within a single bucket with strong probability over many rounds of mapping. For all applications, Euclidean distances are used to find the next neighbor in one bucket that combines the high quality of the process and the consumption of the central processing unit. During this document, we tend to modify the first LSH algorithm to create acceptable WSN events.

3.2 *Unsupervised Algorithms for Anomaly Detection*

The first thing to be considered is unsupervised learning algorithms. The bulk-connected devices, however, are only an unregulated part of traffic. Next, it looks like unusual traffic is different from ordinary traffic. According to some three bases, info groups of identical occurrences which sometimes tend to be regular traffic measures are uncertain, whereas these changes occur very rarely from square measurements considered to be malicious in most instances. The foremost common unsupervised algorithms are isolation forest, neural networks, and autoencoder.

3.2.1 Isolation Forest (Density-Based Technique)

The isolation forest [20] is a variant of the random forest algorithm used to detect anomaly detection. Isolation forests rely entirely on random forests, which are a bunch of decision trees that can be used for regression and classification. The decision tree partitions the data and keeps partitioning the data one at a time until each partition is of the same type. On the other hand, isolation forest involves building isolation trees, where the idea is not to create homogeneous partitions but to create partitions to isolate each data point. Here, when we say “isolated,” it means that a particular partition only contains that data point. The whole intuition of the isolation forest is that regular points are much harder than abnormal points. In isolation, the number of partitions in the forest tells us whether a particular point is an abnormal point or a regular point. To build an isolation forest, two main steps need to be performed: We must randomly select a feature and randomly partition with the range. Repeat these steps until each point is isolated. Once a set of isolation trees is established, we must start to make predictions. The prediction process involves calculating anomaly scores for any given new points.

The main parameters of the isolation forest are the same number of iTrees as the random forest, the sampling size, and then the final parameters of pollution, which tell us how to classify points as anomalies based on the trained model.

3.2.2 Unsupervised Neural Networks (Neural Networks-Based Technique)

Neural networks [21] with manageable memory units allow the network to learn when hidden layer states can be forgotten and when new knowledge can be used to modify hidden states [17]. They are very useful in many competing long-term applications such as tools for translation, speech, and image classification. In general, long short-term memory (LSTM) iteratively converts the input data to a series of hidden states at the present level, and so the LSTM active learning should be sequential [22]. Replace the hidden unit with what is called LSTM and add a different connection from each cell that is called cell state, and this is now called LSTM recurrent neural network. Apart from the hidden state vector, LSTM is designed to mitigate a disappearing and exploding gradient concern that each LSTM cell holds a cell state vector, and at each step, the next LSTM can read, write, or reset the cell by explicit matching mechanism. Each unit has three same-shaped doors that look like binary doors [23].

The input gate controls the memory cell. The forget gate checks whether the memory cell remains at 0 and checks whether current cell status information is visible from the output. All of them have a sigmoid function. But why sigmoid its such theatrical nature is that they form smooth curves in the range from 0 to 1, and the model remains differentiable. We have another vector C bar beside these gates, which changes the cell state. It has a function to activate tanh. But why tanh here distributes the gradients quite well, the zero-centered range allows the cell state information to disappear without exploding.

3.3 Autoencoder (Neural Networks-Based Technique)

An autoencoder with two important changes is an interesting variant: First, the number of neurons on the input and the output is identical, therefore we can expect that the output is not just the same size as the input, but is the same image. Now it wouldn't usually make any sense why we would invent a neural network to do copying? The second part, therefore, goes here: In one of these layers, we have a bottleneck. That means the number of neurons in that layer is significantly lower than we normally would see so that it has to find a way to best represent such data with far smaller neurons. If you have a smaller budget, then we can't expect that the image will be the same but they are however very close, so let us go of all that fluff and focus on bare essentials. These autoencoders can create sparse input data displays and can thus be used for image compression. Autoencoders do not offer any tangible benefits over traditional compression algorithms such as jpeg. However,

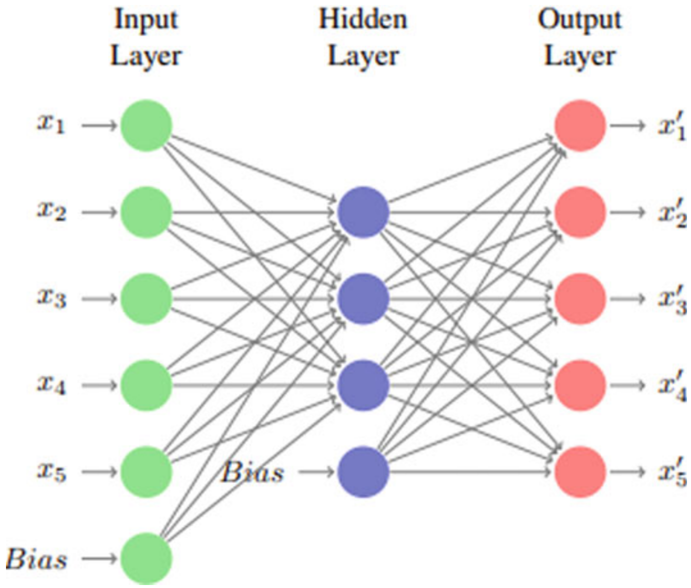


Fig. 3 Autoencoder architecture

many different versions serve other than compression as a comfort crumb. There are self-denoising encoders that can be presented with noisy images after learning these sparse representations. They may help denounce these pictures because they know more or less how this type of information would like. What is even better is a variant known as the variational autoencoder, which not only can learn these sparse images but also draw new images. For example, we can ask for new manuscript numbers that we can expect to make sense of results. Here, Fig. 3 represents an autoencoder.

4 Evaluation

Here we have evaluated semi-supervised/unsupervised techniques or density-based/neural-based techniques on three datasets are credit card, kdd cupp99, and stock market data [24].

4.1 Credit Card Dataset

The response is divided into two types: fraud and non-fraud, with one in the case of fraud and zero in the case of non-fraud. We could detect point anomalies in our credit card dataset where a particular data point is anomalous if it is much further ahead of all the others. Credit card fraud based on "amount spent" can be detected, for example. We can see the positive and negative features of the PCA transformed, and that the precision is highly floating. Aside from that, there is time and amount. Here, time refers to the amount of time that has elapsed between each deal, and hence the first deal in the dataset, which is measured in seconds and is unlikely to be helpful and can be removed when you come to the feature scaling and amount refers to the deal quantity. The difference in scale between the PCA variables and the dollar level indicates that the data scaling for these input variables should be used. Including the response variables, there were twenty-nine predictors in the dataset. Before incorporating all of the predictors into the tests, all of the distributions for deceptive and conventional transactions were forethought and contrasted to see if it was appropriate to use all of them. [25]. The total dataset, therefore, consists of 284,807 samples, 80% of which have been considered training data and 20% as test data. For comparison, assume the distribution of predictor variables, including time on the left and quantity spent on the opposite. The main one remaining is the time for each true or traditional dealings and deceitful or false for the response (class) whether or not a transaction was fraud or traditional [26].

From the figure on the left Fig. 4 can observe that the two transactions have identical allocations, pointing out that the time variable is not that helpful in predicting transaction types. As a consequence, the two quantities for the "amount" predictor are unique in various transactions, as expected [10].

In the case of fraud and normal, the answer is a binary class with 1 and 0. Before adding the predictors in the model, all distributions were mapped to determine whether all predictors were available in the scan. First, the left time, whether the transfer is fraud or not, is for an actual or natural transaction and fraud or false

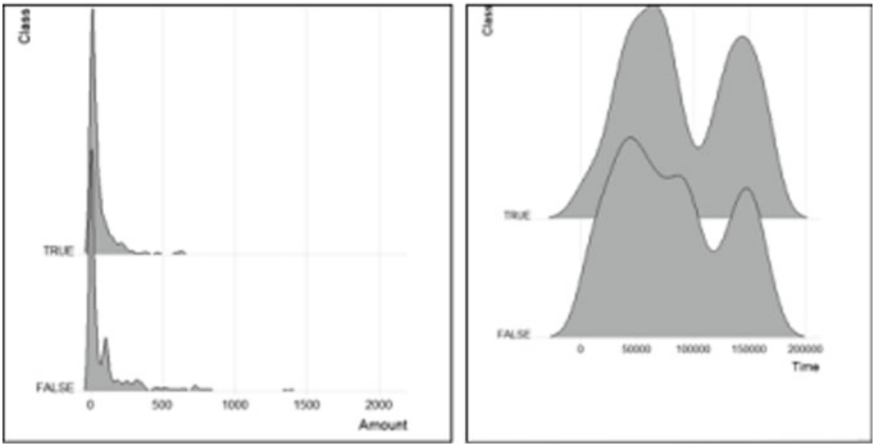


Fig. 4 Time and amount spent distributions as a transaction versus non-identical groups of a class as a response

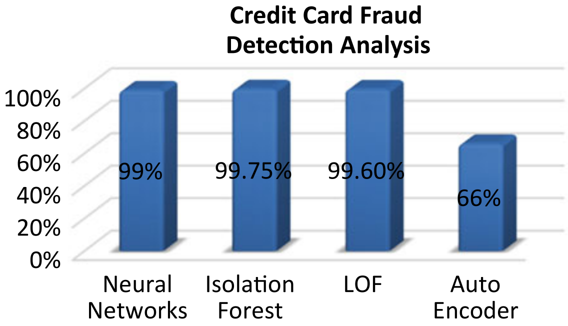


Fig. 5 Comparison of semi-supervised and unsupervised models

response (class). The dataset has been collected and analyzed from a research collaboration of worldline and the machine learning group (http://mlg.ulb.ac.be) of ULB on big data mining and fraud detection (Fig. 5).

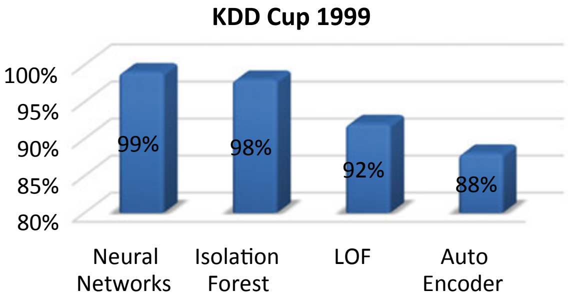
4.2 *KDD Cup Dataset*

In kdd cup99, we focus on collective anomalies, which collectively detect anomalies through a series of data circumstances. The KDD training dataset consists of 10% of the dataset, which consists of nearly 494,020 single link vectors, each of which comprises 41 characteristics and is tagged with one particular type of attack, i.e., either regular or attack. Each vector, with exactly one particular attack form, is classified as either normal or an assault. Deviations from "natural actions" are considered assaults, anything that is not normal. Attacks classified as usual are ordinary behavioral records. For memory-constrained machine learning approaches, a smaller 10% training dataset is also given. There are 19.69 percent natural and 80.31 percent attack connections to the learning dataset. The most common use of KDD CUP 99 has been in security threats. The simulated assault falls into one of the four groups below. The total dataset, therefore, consists of nearly 494,000 samples, 80% of which have been considered training data and 20% as test data. The link below gives a clear idea about the respective dataset http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (Fig. 6).

4.3 *Stock Market Dataset*

We use many datasets of S&P 500 components from various business sectors. Two different granularities of daily/monthly frequencies are used in these datasets. Market anomalies and price anomalies are two common types of anomalies in finance. This component contains the huge stock price stocks chosen by a team of regular and

Fig. 6 Comparison of semi-supervised and unsupervised models



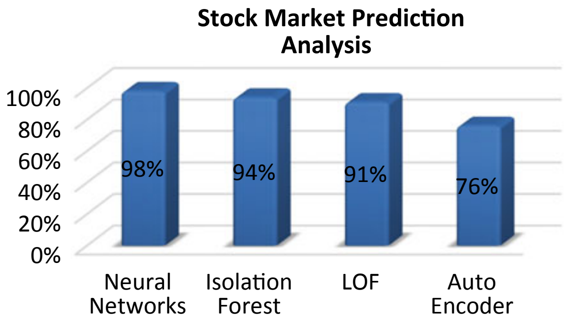
poor market analysts. The S&P 500 index represents the attributes of the top 500 biggest market caps and is also the leading tracker of US equity markets. We use 10 distinct datasets over four years, including 636-time series. Despite both size and diameter of time series in the databases, this analysis exceeds the past researches to the best of knowledge. Here, in the stock market dataset nearly 75% of which have been considered as training data and 25% as test data. This describes a list of datasets that we extracted from the Thompson Reuters database for experiments to study and validate our proposed method (the CSV files are available at www.ualberta.ca/~golmoham/DSAA2015/) (Fig. 7).

Accuracy Comparison of Anomaly Detection Algorithms on various datasets:

Algorithms/datasets	Credit card dataset (%)	Kdd cup99 dataset (%)	Stock market dataset (%)
Neural networks	99	99	98
Isolation forest	99.75	98	94
Local outlier factor	99.6	92	91
Autoencoder	66	88	76

Computational time differences (seconds) of Anomaly Detection Algorithms on various datasets:

Fig. 7 Comparison of semi-supervised and unsupervised models



Algorithms/datasets	Credit card dataset (s)	Kdd cup99 dataset (s)	Stock market dataset (s)
Neural networks	520	680	580
Isolation forest	180	210	194
Local outlier factor	250	298	271
Autoencoder	409	514	477

5 Anomaly Detection Algorithm Comparison

Diverse semi-supervised/unsupervised techniques or density-based/neural-based techniques are used for validating and performance comparison of each form. The individual methods of each algorithm have been analyzed. The previous description of everything was completely different supported. A comparison of every formula is provided in table one. Each algorithm has its advantages and disadvantages.

An algorithmic anomaly detection rule includes supervised, unsupervised, and semi-supervised techniques. Each algorithm rule has its advantages and disadvantages in several papers. In our paper, we tend to focus completely on unsupervised and semi-supervised techniques for detecting anomalies. Among all algorithms, it is obvious that semi-supervised algorithm techniques playacting well compared to unsupervised ones once target information hasn't any unknown attack. However, once it involves an unsupervised technique neural networks algorithmic rule was providing you with the simplest result among all.

Technique	Positives	Negatives
Isolation Forest (density-based technique algorithm)	# Using Isolation Forest, we will not solely sight anomalies quicker; however, we tend to additionally need less memory compared to different algorithms # Isolation Forest isolates anomalies within the data points rather than identification with traditional data points # This algorithmic rule works fine with a small amount of information	# It affects the precision of the algorithm as it detects local anomaly points

(continued)

(continued)

Technique	Positives	Negatives
Local outlier Factor (LOF) (density-based technique algorithm)	# LOF will discover the Outliers supported k-nearest distance and native reachability distance # LOF formula will apply to the wide selection of fields # LOF formula will beat several different algorithms	# It is not required that the LOF score perpetually finds the Outliers. It depends on the edge used on the score, and it varies with totally different datasets # LOF exactness can get affected if the information has high dimensions
Neural Networks (neural-based technique algorithm)	# Neural Networks performs pretty much still if you've got insufficient information # once part of the neural network fails, it will continue with no drawback with its parallel nature # A neural network learns and doesn't ought to be reprogrammed	# The neural network results in overfitting in some cases # It needs high process power and interval for giant neural networks
Autoencoder (neural-based technique algorithm)	# We will use long STM (LSTM) neural network cells in our autoencoder model # In several cases, autoencoders can extremely improve the performance # spatial property reduction is applied to victimization autoencoders # It outperforms alternative algorithms once it involves image, series variety of knowledge	# Slow convergence in some cases # Time-Consuming algorithm

6 Conclusion and Future Scope

Nowadays, machine learning technology receives an enormous response to its various options, including anomaly detection and outlier detection. In several papers, every algorithmic rule has its execs and cons. In our paper, we tend to focus fully on the unsupervised, semi-supervised detection of anomalies. Among all the algorithms, it is clear that semi-supervised algorithm techniques playacting well compared to unsupervised once target information hasn't any unknown attack. However, the neural networks algorithmic rule provided you with the simplest result of all once that involves an unsupervised technique. So finally in our paper we examined various algorithms like density-based techniques and neural-based techniques when density-based techniques are concerned if you want to detect anomalies with the very less computational time difference and also good performance isolation forest is the appropriate algorithm but the computational time difference is slightly higher with local outlier factor still LOF is the most appropriate algorithm for detecting abnormalities with high performance and when it comes to neural-based techniques, unsupervised neural network algorithm is the most appropriate algorithm still if the computational time difference of neural networks is very high. Overall, the local outlier factor and neural networks are the algorithms for detecting abnormalities if you want to choose one of the best suitable algorithms.

For further study, certain additional ensemble algorithms can be applied in particular to highly performing extreme learning machine algorithms.

References

1. F. Angiulli, R. Ben-Eliyahu-Zohary, L. Palopoli, *Outlier Detection Using Default Logic* (2003)
2. P.D.L. Ertöz, E. Eilertson, A. Lazarevic, P.-N. Tan, V. Kumar, J. *Srivastava.: Minds-Minnesota Intrusion Detection System* (Next Gener. MIT Press, 2004)
3. C. Mishra, T. Bagyammal, L. Parameswaran, An algorithm design for anomaly detection in thermal images, in innovations in electrical and electronic engineering, in ed. by M.N. Favorskaya, Mekhilef, S., Pandey, R. Kumar, N. Singh (Singapore: Springer Singapore, 2021)
4. X. Niu, L. Wang, X. Yang, A comparison study of credit card fraud detection: supervised versus unsupervised, arXiv Preprint arXiv:1904.10604 (2019)
5. F. Carcillo, et al.: Combining unsupervised and supervised learning in credit card fraud detection. Inf. Sci. (2019)
6. A. Srivastava et al., Credit card fraud detection using hidden Markov model. IEEE Trans. Dependable Secure Comput. **5**(1), 37–48 (2008)
7. A. Malathi, J. Amudha, P. Narayana, *A Prototype to Detect Anomalies Using Machine Learning Algorithms and Deep Neural Network, Lecture Notes in Computational Vision and Biomechanics*, vol. 28 (Springer Netherlands, 2018), pp. 1084–1094
8. T. Shon, J. Moon, A hybrid machine learning approach to network anomaly detection. Inf. Sci. **177**(18), 3799–3821 (2007)
9. Y. Tian et al., Ramp loss one-class support vector machine: a robust and effective approach to anomaly detection problems. Neurocomputing **310**, 223–235 (2018)
10. S. Bhattacharyya et al., Data mining for credit card fraud: a comparative study. Decis. Supp. Syst. **50**(3), 602–613 (2011)
11. M. Hejazi, Y.P. Singh., One-class support vector machines approach to anomaly detection. Appl. Artif. Intell. **27**(5), 351–366 (2013)
12. C. Fan et al., Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. Appl. Energy **211**, 1123–1135 (2018)
13. T. Sweers, T. Heskes, J. Krijthe, *Autoencoding Credit Card Fraud* (2018)
14. P. Filzmoser, R.G. Garrett, C. Reimann, Multivariate outlier detection in exploration geochemistry. Comput. Geosci. **31**(5), 579–587 (2005)
15. F.R. Hampel et al., Robust statistics: the approach based on influence functions, 2011196
16. P. Filzmoser, K. Hron, C. Reimann, Interpretation of multivariate outliers for compositional data. Comput. Geosci. **39**, 77–85 (2012)
17. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural. Comput. **9**, 1735–1780 (1997)
18. A. Ashok, S. Smitha, M.H.K. Krishna, Attribute reduction based anomaly detection scheme by clustering dependent oversampling PCA, in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1298–1304 (2016). https://doi.org/10.1109/ICACCI.2016.7732226

19. M.M. Breunig, H.-P. Kriegel, R.T. Ng, J. Sander, Lof: Identifying density-based local outliers, in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data* ed by W. Chen, J. F. Naughton, P. A. Bernstein May 16–18(Dallas, Texas, USA, ACM, 2000), pp. 93–104
20. H. Yong-xu, D. Lei, Q. Jiang-long et al., Parallel detection design based on isolation forest. J. Comput. Engi. Sci. **39**(2), 236–244 (2017)
21. K. Kavikuil, J. Amudha, Leveraging deep learning for anomaly detection in video surveillance. Adv. Intell. Syst. Comput. **815**, 239–247 (2019) (Springer Verlag)
22. R. Vinayakumar, K.P. Soman, P. Poornachandran, Long short-term memory based operation log anomaly detection, in *2017 International Conference on Advances in Computing, Communications, and Informatics (ICACCI)*
23. G.R. Ramya, P. Sivakumar, An incremental learning temporal influence model for identifying topical influencers on Twitter dataset. Soc. Netw. Anal. Min. **11** (2021) https://doi.org/10.1007/s13278-021-00732-4
24. A. Anandharaj, P.B. Sivakumar, Anomaly detection in time series data using hierarchical temporal memory model, in *2019 3rd International Conference on Electronics, Communication, and Aerospace Technology (ICECA)*, pp. 1287–1292 (2019). https://doi.org/10.1109/ICECA.2019.8821966
25. J. Van Hulse, T.M. Khoshgoftaar, A. Napolitano, Experimental perspectives on learning from imbalanced data, in *Proceedings of the 24th International Conference on Machine Learning* (2007)
26. M. Krivko, A hybrid model for plastic card fraud detection systems. Expert Syst. Appl. **37**(8), 6070–6076 (2010)
27. T. Garcıa, V. Dıaz, F. Macia, Vazquez, Anomaly-based network intrusion detection. Comput. Secur. **28**, 18–28 (2009)
28. T. Vijayakumar, Posed inverse problem rectification using novel deep convolutional neural network. J. Innovat. Image Proc. (JIIP) **2**(03), 121–127 (2020)
29. H. Wang, Sustainable development and management in consumer electronics using soft computation. J. Soft Comput. Parad. (JSCP) **1**(01), 56 (2019)

Hybrid Model Using Feature Selection and Classifier in Big data Healthcare Analytics

M. Kavitha, Singaraju Srinivasulu, Mulaka Madhava Reddy, Vellaturi Gopikrishna, Sindhe Phani Kumar, and S. Kavitha

Abstract The growing era of technology in healthcare results in a large amount of data generation termed big data. Mining this data generating valuable insights is crucial in the classification or prediction of disease at an early stage. Effective decision making in the domain of healthcare happens using advanced data mining techniques. In this article, a hybrid model using feature selection and classifier is examined. The feature selection method is applied to the heart disease dataset to select more appropriate features and results given to the classifier to predict the heart disease at an early stage. Three feature selection methods analysis of variance, Pearson's correlation coefficient, and mutual information gain applied on heart disease dataset and the performance of random forest classifier were examined over heart disease prediction. Experimental results showed that the feature selection approach increases the accuracy of the classifier.

Keywords Big data · Data mining · Feature selection · Healthcare · Machine learning

M. Kavitha (✉)
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur 522502, India
e-mail: mkavita@kluniversity.in

S. Srinivasulu · M. Madhava Reddy
Department of Information Technology, PACE Institute of Technology and Sciences, Ongole, Andrapradesh, India
e-mail: srinivasulu_s@pace.ac.in

M. Madhava Reddy
e-mail: madhavareddy_m@pace.ac.in

V. Gopikrishna · S. Phani Kumar
Department of Computer Science and Engineering, PACE Institute of Technology and Sciences, Ongole, Andrapradesh, India

S. Kavitha
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur 522502, India
e-mail: kavihabtech05@kluniversity.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_59

1 Introduction

Nowadays, a massive amount of data is generated in all the domains like medical, social sites, sensor networks, online transactions, and so on in a variety of forms like text, image, audio, video, and this data includes thousands of attributes (called it as big data). Effective processing of this massive high-dimensional data is a challenging task in this research era, and the removal of redundant attributes or irrelevant data from this is the crucial task to improve the performance and accuracy of the data mining approach. On these requirements, various authors proposed feature selection methods to improve performance in the data mining task. The knowledge extracted from these supports the classifiers to increase the quality of prediction [1].

In a big data environment, knowledge discovery is a multistep process that includes data preprocessing, data extracting, pattern assessment, and interpretation. Data preprocessing leads to remove the noise and makes the dataset ready for data extracting. Data combination, cleaning, transformation, and dataset reduction are various subtasks of data preprocessing. Data combination is a method of collecting data from various resources and storing them in a required format for further processing. Data cleaning is the process of handling missing values and removing noise from the dataset. Data transformation is the converting of data into the required format for further processing. Generalization and normalization are widely used data makeover mechanisms. Data reduction is the procedure of reducing the dimension of the dataset by removing irrelevant attributes from the dataset to improve the accuracy of the data mining task [2].

Data mining is the procedure of separating a required pattern from the preprocessed data and extracted patterns are interpreted for the generation of knowledge. Predictive data mining is one sort of data mining that finds the knowledge from the data with which the prediction of the class label from the given set of feature values of an instance is performed. Classification is the best example of a predictive-based data mining task. It is also termed as supervised learning approach, and it learns the labeled data of the dataset to build the classifier model. Generally, the dataset involves the set of rows called instances, a set of columns called features (attributes), and a target attribute related to each instance of the dataset [3–5].

The classification model predicts or classifies the label of the unlabeled instance. Some of the common classification procedures are naïve Bayes, decision tree, random forest, support vector machine, neural network, and so on are utilized to build the classifier prototype. The graphic representation of the classification procedure is shown in Fig. 1. The classifier algorithm utilizes a training dataset to build the classifier which includes input data and class labels. The testing dataset is utilized to authenticate the model [6–8]. Some datasets comprise many attributes, but some of them are not necessary for classification and prediction. Extracting those major features to build the classifier model is an essential step in machine learning procedures. So, there is a need to prioritize the attributes based on their relativeness. The complexity of the classification is reduced, and the accuracy of prediction is increased due to this required feature selection mechanism in the machine learning platform. Feature

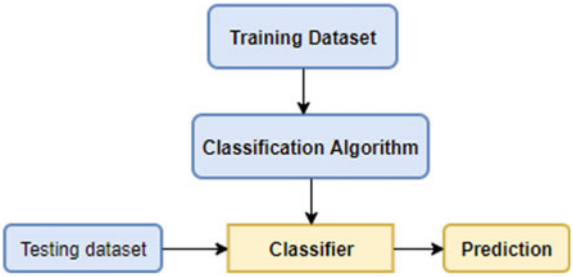


Fig. 1 Graphical representation of classification procedure

selection is the method to eliminate the redundant and unwanted features from a high-dimensional dataset and reduces the data size. The elimination of irrelevant and redundant features is the first objective of any feature selection algorithm followed by the selection of relevant features through various selection measures.

Section 1 discussed big data processing challenges, feature selection need, and data classification mechanism. In Sect. 2, various author's contributions in this identified research are discussed. Section 3 discussed research methodology. In Sect. 4, implementation details and result analysis are discussed. Finally, the paper concluded with the conclusion part.

2 Literature Review

Pavya et al. discussed the three-step process feature extraction, selection, and classification to diagnose thyroid disease. The authors discussed filter-based and wrapper-based feature selection methods and their role in thyroid disease identification and classification. They also analyzed the data using PCA dimensionality reduction approaches. The implementation of the recommended model is evaluated with the assistance of three metrics accuracy, sensitivity, and specificity [9].

Aich et al. recommended a hybrid classification prototype to predict Parkinson's infection using nonlinear feature elimination procedure and classification techniques. In their work, they attempted to differentiate Parkinson's disease alliance from healthy control group persons created on voice records with chosen features and various classification approaches. The performance of classification techniques is compared using specificity, sensitivity, positive and negative predictive metrics [10].

Liu et al. discussed an innovative model to diagnose breast cancer using feature selection and machine learning classifiers. Early-stage detection of malignant breast cancer can improve the survival rate of people through successful treatment. Relevant features from the breast biopsy image dataset are selected using the feature selection approach and machine understanding algorithms [11].

Koçak et al. discussed principal component analysis, relief feature selection approaches and neural network, nearest neighbor, support vector machine learning approaches in classification elder people's health conditions in the sensor-based remote healthcare system. The experimental results have shown that the proposed mechanism gives approximately 99% percent accuracy in the healthcare system for elderly people [12].

Krawczyk et al. recommended a collaborative classifier model to design an effective clinical support decision-creating system for breast tumor diagnosis. This method is proposed to detect the level of malignancy in different aged cancer patients. Highly discriminative attributes are selected using the attribute selection technique. The experimental results of this work prove that the designed model well classified the malignant level of breast cancer patients [13].

Waseem et al. discussed feature choice methods and classifiers to predict tumor. Reject option classifiers are utilized to enhance the predictive accuracy of tumor patients. The objective of the proposed work is to enhance 3 factors feature selection, studying mechanism, and denial rate for cancer forecast. Three different cancer datasets are investigated for cancer prediction using proposed work and revels that the predictive accuracy of deny option classifiers is distinct for various feature choice techniques [14].

Nnamoko et al. discussed the impact of feature selection on different ensemble approaches in diabetes prediction. The proposed approach is utilized on the VCI diabetes onset dataset and the exactness is compared over widely used classification algorithms [15].

Liu et al. discussed feature selections on healthcare data using principal component analysis and autoencoder-based approaches. Feature selection gives useful knowledge which is appropriate to specific health situations from the dataset. The relationships among the identified features support the clinicians to classify patient conditions in a better way. The proposed method tested six different healthcare datasets using three different classifiers and results evidence that it works well compared to standard approaches [16].

Kayastha et al. developed a novel model to evaluate biometric features for user authentication. Feature evaluation and selection are key steps to identify key features in building an accurate classifier for user identification. The developed model is tested using the data collected from IoT device wear on the user's wrist [17].

Pahwa et al. designed a hybrid approach with feature selection techniques and machine learning classifiers to predict heart disease. The feature choice techniques SVM-RFE and mutual information gain ratio are utilized to the dataset to select more weightage features for classification. Machine learning classification approaches are applied to selected features and the performance is evaluated in terms of model exactness and execution time [18].

Haq et al. suggested a model to predict heart disease with a sequential backward selection approach and machine learning classification techniques. The proposed model assists the physician to treat the patient at the early stage of the disease. Sequential backward selection algorithm selects the most suitable features to enhance the KNN and SVM techniques accuracy. The model performance is evaluated using a heart disease dataset from Cleveland resource [19].

Ghosh et al. proposed cardiovascular disease prediction system using relief feature selection and machine learning techniques mixture. Authors created a data resource using multiple datasets for implementation. Best features are selected using wrapper-based feature selection methods and that are given to various machine learning bagging methods. The results are compared in terms of accuracy, recall, f1-score, and precision [20].

Various authors' contributions in prediction of disease using feature selection and machine learning mixture are shown in Table 1.

3 Research Methodology

The main intention of this research is to examine the hybrid model in heart disease prediction. In this model, the feature selection approach is applied to the dataset and the selected features are given to the classifier to predict heart disease. The model performance was tested using three verities of feature selection approaches—the analysis of variance, Pearson's correlation coefficient, mutual information gain, and random forest classifier. Data preprocessing, feature selection, classification, and performance comparison are the major steps involved in this approach. The flow of steps in this model is shown in Fig. 2.

3.1 Dataset

A comprehensive heart disease dataset sample with 1195 instances is used in this work. This dataset is generated by combining five heart datasets (Cleveland, Hungarian, Switzerland, long beach VA, and Statlog) over eleven common features. This dataset is created exclusively to build a predictive model for early-stage detection of heart disease. Age, sex, type of chest pain, BP, cholesterol, resting electrocardiogram, fasting blood sugar, heart rate, exercise angina, old peak, ST-segment slope, and target are features in the dataset. The target is binary attribute means, class 1 refers to heart disease and class 0 refers to normal. The number of instances in each category is shown in Fig. 3.

3.2 Data Preprocessing

In data preprocessing, the data is converted into a more efficient format for operation. In this preprocessing step, the missing values in the age column are replaced with an average age of that column, and the instances having missing values other than the age column are deleted from the dataset.

Table 1 Presentation of feature selection and machine learning in prediction of disease

Article author	Healthcare application	Feature selection technique	Classification technique	Performance Metrics	Results
Pavya et al. [9]	Thyroid disease diagnosis	F-score and recursive feature elimination	MLP, NN, SVM	Accuracy, specificity, recall	Wrapper-based recursive feature elimination method enhances the performance of thyroid disease diagnosis
Aich et al. [10]	Parkinson's disease	Recursive feature elimination	SVM	Accuracy, specificity, recall, positive and negative predictive values	The mechanism supports the medical person to separate affected patients from the healthy group based on voice records
Liu et al. [11]	Breast tumor diagnosis	Most related features selection	DT, RF, SVM	Accuracy, recall, specificity	The predicted benign and malignancy of breast tumor helps the doctors to treat the patient in better way
[12]	Remote healthcare assistance for elder people	Relief, PCA	KNN, SVM, NN	Accuracy	Feature selection method enhances the performance of remote healthcare system
Waseem et al. (2019)	Cancer	Los Vegas filter, t-test, information gain	Reject option classifier	Accuracy	The predicted results assist the doctors to treat cancer patient at early stage
Haq et al., 2019	Heart disease	Sequential backward selection	KNN	Accuracy	The model separates the heart disease patients from healthy persons effectively

(continued)

Table 1 (continued)

Article author	Healthcare application	Feature selection technique	Classification technique	Performance Metrics	Results
Ghosh et al. (2021)	Cardiovascular disease	Relief feature selection	ML bagging methods	Accuracy, recall, precision, f1-score	The model predicts the cardiovascular abnormality at early stage

DT: Decision tree, *RF*: Random forest, *MLP*: Multilayer perception, *KNN*:*K* nearest neighbor, *SVM*: Support vector machine, *NN*: Neural network, *PCA*: Principal component analysis

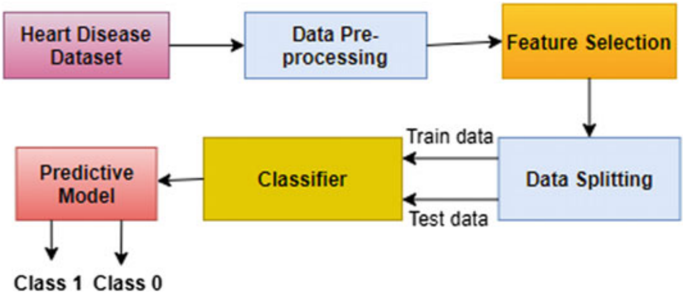
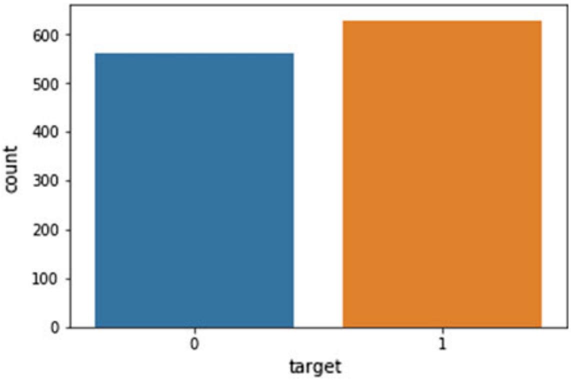


Fig. 2 Block diagram of hybrid model

Fig. 3 Two categories of observations in dataset



3.3 Feature Selection

Feature selection is one of the foremost data processing steps in many classification applications. It is the process of reducing the feature set by selecting appropriate features from the actual dataset based on some evaluation criteria [21]. Let A be the feature set with the number of features {a_1, a_2, a_3,…,a_n} where n is the number of features in the actual dataset. Feature selection function Fi is defined on A to select d most relevant feature out of D > d[1]. Fi generates a good feature subset that satisfies the evaluation criteria [22].

A good quality feature subgroup includes features highly forecasting of the class and not predictive of each other. This reduces the redundancy and increases improve the performance of the predictive model by decreasing the complexity in model building [23]. In this study, we used filter categorized correlation and statistical algorithms Pearson's correlation coefficient, ANOVA, and mutual information (entropy) gain for features selection.

3.4 Feature Selection Based on ANOVA

ANOVA is one of the statistical and ranking-based filter methods for feature selection, and its full form is an analysis of variance. The ANOVA elimination measure intends to lessen the size of the input highlight set and simultaneously to hold the class oppressive information for grouping problems. It selects the features based on the dependency of two variables. It assumes the linear relationship between variables and the target feature. F_classify() is one of the statistical methods for feature selection. It provides the ANOVA F-value for the provided sample. It takes two arguments X: {array of shape n -samples), Y: (array of shape n_samples), and returns F-statistic of each feature, P-values associated with the F-statistic. The approach is given below.

Step 1: Compare the variance between the groups or variance within the groups that means F-statistic values calculated.

F-statistic = variability between groups/variability within groups.

Step 2: The F-statistic score is converted into P-values associated with features. The largest F-statistic lead to small p-values. If p-value is small enough, then that hypothesis is rejected.

3.5 Feature Selection Based on Pearson's Correlation Coefficient

Pearson's correlation algorithm is used to find the relationship between the continuous features and the target feature. This is one of the useful feature selection methods in classification applications. In this, the strength of the linear relationship between data items varies between + 1 and -1.f_regression(), is a linear technique for assessing the specific impact of each of many regressors. This is one of the making functions used in the feature selection technique. It accepts three parameters as arguments such as P: {array with n samples}, q: {array with n samples}, center: bool, default = True and it returns F and pval both are array with n features. The approach is done in two steps.

Step 1: The correlation between each regressor and the target is calculated.
((P[:, i] - Mean(P[:, i])) * (q - Mean_q)) / (STD(P[:, i]) * STD(q)).
Step 2: It is transferred to an F-score then to a p-value.

3.6 Feature Selection Based on Mutual Information (Entropy) Gain

Mutual information (MI) gain is an extent of the proportion of information between two random components is symmetric and non-negative, and its worth is zero if the variables are free. The MI between two components $X = (x_1, x_2, x_3 \ldots x_k)$ and $Y = (y_1, y_2, y_3 \ldots y_d)$ is characterized as –

$$I(X, Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

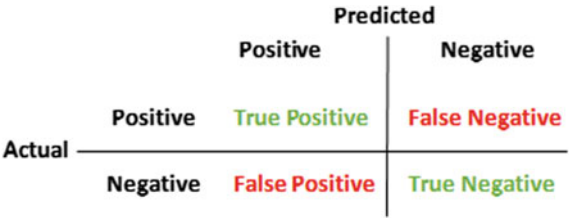
where $(x_1, x_2, x_3, \ldots, x_k)$ and $(y_1, y_2, y_3, \ldots, y_d)$ are the values of discrete variable X and Y and $p(X, Y)$ is a joint thickness work, and $p(x)$, $p(y)$ are the minor thickness capacities.

3.7 Dataset Splitting

Dataset is divided into two sets of data and called as train set and test set. The classification algorithm learns using the observations in the train set and observations in the test set are used to estimate the algorithm performance in the real world. Overfitting and underfitting balancing are done using this dataset splitting in classification applications.

3.8 Classifier

In this study, we used a random forest classifier for heart disease prediction. The classifier learns using the observations in the training dataset, and it predicts the outcome for new observations. Suppose, X and Y are two observations in the training dataset and the learning function is H: X &#xF0E0; Y, such that allotted examination X, $H(X)$ can calculate Y quantity. The outcome of classifies is either class 1: heart disease or class 0: normal. Random forest classifier takes the number of decision trees



		Predicted	
		Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Fig. 4 Confusion matrix

as input and takes its average to improve the prediction accuracy. Various subsets of original dataset are used to construct decision trees.

3.9 Performance Assessment

The performance of the model is assessed against the test dataset in words of machine learning assessment metrics confusion matrix, accuracy, precision, recall, f1-score, and wall time [24, 25].

3.9.1 Confusion Matrix

Confusion matrix is a $N * N$ table that sums up the prediction results of a classification model. One pivot of the matrix has the classes/labels anticipated by the model, while the other axis shows the actual classes [26]. Figure 4 shows confusion matrix. In a twofold problem, true positive means predicted positive and are positive, true negative means predicted negative and are really negative, false positive means predicted positive and are really negative, false negative means predicted negative and are positive [27].

Accuracy: Accuracy means the percentage of correct classifications. The higher value of accuracy refers the model is more accurate.

Precision: Precision refers the proportion of positive prediction results that are correct.

Recall: Recall refers the probability of actual positives that predicts correctly.

F1 Score: F1 score is the harmonic mean of precision and recall. The best model gives highest f1-score as 1.

4 Implementation and Result Discussion

In this study, the python programming language is used for the implementation of the model. Anaconda navigator Jupiter notebook is used to run the code, which provides a user-friendly environment to run the codes in Python. There are predefined functions for feature selections and to build classifiers on the dataset in the machine learning library. The step-by-step explanation is given in below:

Step 1: Load the dataset using the read.csv and preprocess steps applied.

Step 2: Call the train_test_split function to split the dataset in a 70:30 ratio.

Step 3: Applied feature selection method to select best features from a dataset.

Step 4: Called make_pipeline () method on the selected feature set and to the classifier.

Step 5: Called model. fit () function using the train part of the dataset.

Step 6: Called model. score () function on the test part of the dataset to get the accuracy of the model.

Step 7: Call classification report () function over train and predicted result of test dataset to get the model evolution report.

Step 8: The model results are compared using three feature selection methods over the prediction of heart disease.

We performed experiments on the heart disease dataset using three varieties of feature selection methods analysis of variance, Pearson's regression coefficient, and mutual information gain, and in all models, the selected best features are given to random forest classifier to validate the heart disease prediction. The random forest classifier computation comprising of numerous decision trees utilizes packing and feature randomness when constructing every individual tree to attempt to make an uncorrelated forest of trees whose forecast by panel is more exact than that of any individual tree. The dataset is divided into train sets and test set in the ratio of 70:30. Constant, duplicate, and quasi constant features are removed using transform operation. The feature selection techniques are applied, and the reduced feature set with six features was given to a random forest classifier and the classification report is observed for the applied model in terms of confusion matric, accuracy, precision, recall, and f1-score. The wall time of three models is different and given same accuracy. Table 2 indicates the wall time of three models. The performance of the model is evaluated against the general random forest classifier, support vector machine, and logistic regression models. In general model, the classifier takes all the features of training dataset into consideration while learning, and in same way, all the feature values are considered while classifying or predicting the observation result. The evaluation metrics results for both hybrid model and general model is shown in Table 3. The confusion matrix metric result for both hybrid model and general machine learning models is shown in Fig. 5. Figure 6 represent graphical

Table 2 Wall time of three models

Feature selection model	Wall time (ms)
Hybrid model using analysis of variance approach	413
Hybrid model using Pearson regression coefficient approach	512
Hybrid model using mutual information gain approach	619

Table 3 Performance report of hybrid model against other ML models

Model	Accuracy	Target class	Precision	Recall	F1-score
Hybrid model	0.9159	0	0.93	0.89	0.91
		1	0.91	0.94	0.92
Random forest	0.8571	0	0.87	082	0.84
		1	0.85	0.89	0.87
Support vector machine	0.8487	0	0.86	0.82	0.84
		1	0.84	0.87	0.85
Logistic regression	0.8151	0	0.84	0.78	0.81
		1	0.80	0.85	0.82

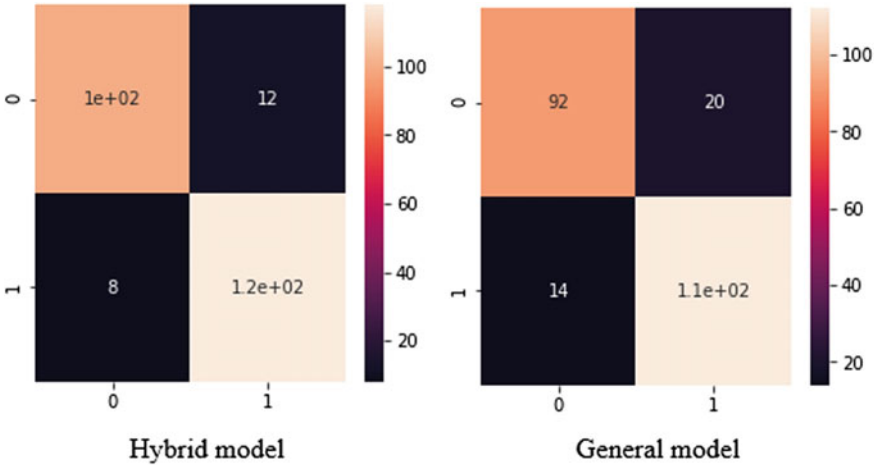
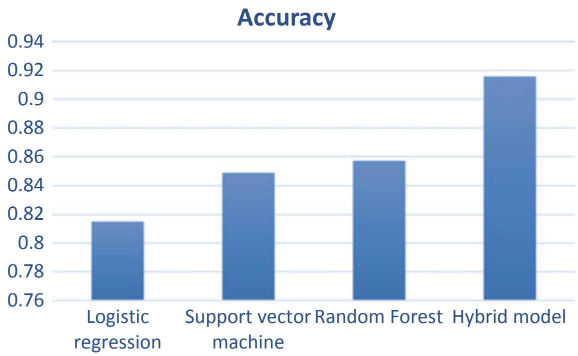


Fig. 5 Confusion matrix for hybrid model and general ML model

Fig. 6 Graphical representation of accuracy comparison



representation of accuracy comparison among hybrid and other machine learning techniques.

All these results showed that feature selection approach increases the accuracy of classification or prediction models in heart disease prediction.

5 Conclusions and Future Work

An efficient hybrid model to diagnose heart disease is examined in this work. In a hybrid model, the best features are selected from the dataset and are given to random forest classier. To analyze the performance of the hybrid model, three different feature techniques analysis of variance, Pearson's correlation coefficient, and mutual information gain are implemented over the heart disease dataset. To see the validity of the hybrid model, the dataset is given to general machine learning techniques random forest, support vector machine, and logistic regression classification algorithms. The performance of the hybrid model and the identified general machine learning models are examined in terms of confusion matrix, accuracy, precision, recall, f1-score, and wall time. The experimental results show that choosing appropriate features using a feature selection algorithm increases the performance of classifier in heart disease prediction. The hybrid model correctly classifies heart disease and healthy people with more accuracy.

In future work, we will collect a big dataset of various diseases and verify its prediction accuracy. We also planned to enhance the performance of hybrid model using other feature selection methods and deep learning mixture.

References

1. P. Sun, D. Wang, V.C.T. Mok, L. Shi, Comparison of feature selection methods and machine learning classifiers for radiomics analysis in glioma grading. IEEE Access **7**, 102010–102020 (2019)
2. A. Juneja, N. Narayan Das, Big data quality framework: pre-processing data in weather monitoring application, in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (IEEE, 2019), pp. 559–563
3. M. Keerthana, K.J.M. Meghana, S. Pravallika, M. Kavitha, An ensemble algorithm for crop yield prediction, in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. (IEEE, 2021), pp. 963–970
4. S.P. Potharaju, M. Sreedevi, A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi Univ. J. Sci. **31**(2), 456–470 (2018)
5. M. Kavitha, G. Gnaneswar, R. Dinesh, Y.R. Sai, R.S. Suraj, Heart disease prediction using hybrid machine learning model, in *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (IEEE, 2021), pp. 1329–1333
6. S.K. Jonnavithula, A.K. Jha, M. Kavitha, S. Srinivasulu, Role of machine learning algorithms over heart diseases prediction, in *AIP Conference Proceedings,* vol. 2292, No. 1 (AIP Publishing LLC, 2020), p. 040013

7. S. Anjali Devi, P. Sapkota, K. Rohit Kumar, S. Pooja, M.S. Sandeep, Comparison of classification algorithms on twitter data using sentiment analysis. Int. J. Adv. Trend. Comput. Sci. Eng. **9**(5), 8170–8173 (2020)
8. Patel, A.K., S. Chatterjee, A. K. Gorai, Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades. Earth Sci. Inf. **12**(2), 197–210 (2019)
9. K. Pavya, B. Srinivasan (2019) Feature selection algorithms to improve thyroid disease diagnosis, in *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)* (IEEE, 2017), pp. 1–5
10. S. Aich, M. Sain, J. Park, K.-W. Choi, H.-C. Kim, A mixed classification approach for the prediction of Parkinson's disease using nonlinear feature selection technique based on the voice recording, in *2017 International Conference on Inventive Computing and Informatics (ICICI)* (IEEE, 2017), pp. 959–962
11. B. Liu, X. Li, J. Li, Y. Li, J. Lang, R. Gu, F. Wang. Comparison of machine learning classifiers for breast cancer diagnosis based on feature selection, in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE, 2018), pp. 4399–4404
12. S. Koçak, T. Artuğ, G. Tulum, A preliminary study for remote healthcare system: activity classification for elder people with on body sensors, in *2018 6th International Conference on Control Engineering & Information Technology (CEIT)* (IEEE, 2018), pp. 1–3
13. B. Krawczyk, Ł. Jeleń, M. Woźniak (2014) Adaptive splitting and selection ensemble for breast cancer malignancy grading, in *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE)* (IEEE, 2014), pp. 104–111
14. M.H. Waseem, M. Sajjad, A. Nadeem, A. Abbas, A. Shaheen, W. Aziz, A. Anjum, U. Manzoor, M. A. Balubaid, S.-O. Shim, On the feature selection methods and reject option classifiers for robust cancer prediction. IEEE Access **7**, 141072–141082 (2019)
15. N. Nnamoko, A. Hussain, D. England, Predicting diabetes onset: an ensemble supervised learning approach, in *2018 IEEE Congress on Evolutionary Computation (CEC)* (IEEE, 2018), pp. 1–7
16. S. Liu, J. Yao, C. Zhou, M. Motani, Suri: feature selection based on unique relevant information for health data, in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE, 2018), pp. 687–692
17. Namrata, Kayastha, K. Sha, A novel and efficient approach to evaluate biometric features for user identification, in *2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)* (IEEE, 2019), , pp. 21–22
18. K. Pahwa, Kumar, R. Prediction of heart disease using hybrid technique for selecting features, in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)* (IEEE, 2017), pp. 500–504
19. A.U.l. Haq, Li, J., Memon, H., Memon, M., Khan, J., Munazza Marium, S. (2019) Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection, in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (IEEE, 2019). pp. 1–4
20. P. Ghosh, S. Azam, M. Jonkman, A. Karim, F.M. Javed Mehedi Shamrat, E. Ignatious, S. Shultana, A.R. Beeravolu, F. De Boer, Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access **9**, 19304–19326 (2021)
21. Blessie, E., Chandra, E. Karthikeyan, Sigmis: a feature selection algorithm using correlation based method. J. Algor. Comput. Technol. **6**(3), 385–394 (2012)
22. S.P, Potharaju, M. Sreedevi, Distributed feature selection (DFS) strategy for microarray gene expression data to improve the classification performance. Clin. Epidemiol. Global Health **7**(2), 171–176
23. E. Vamsidhar, B. Saichandana, J. Harikiran, A novel approach for feature selection and classifier optimization compressed medical retrieval using hybrid cuckoo search. Ind J. Electr. Eng. Inf. **6**(4), 410–417 (2018)

24. P. Tumuluru, C.P. Lakshmi, T. Sahaja, R. Prazna (2019) A review of machine learning techniques for breast cancer diagnosis in medical applications, in *Proceedings of the 3rd International Conference on I-SMAC IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2019*, pp. 618–623, 9032427 (2019)
25. Venubabu Rachapudi and Golagani Lavanya Devi, Feature selection for histopathological image classification using levy flight salp swarm optimizer. Recent Patents Comput. Sci. **12**, 329 (2019)
26. S. Hrushikesava Raju, L. Ramani Burra, S.F. Waris, S. Kavitha, IoT as a health guide tool, in *IOP Conference Series, Materials Science and Engineering*, 981,4, 10.
27. B. Dudi, V. Rajesh, Medicinal plant recognition based on CNN and machine learning. Int. J. Adv. Trends. Comput. Sci. Eng. **8**(4), 999–1003 (2019)

Performance Evaluation of Packet Injection and DOS Attack Controller Software (PDACS) Module

T. G. Keerthan Kumar, M. S. Srikanth, Vivek Sharma, and J. Anand Babu

Abstract Network management is a process where the networks are monitored and their performance is enhanced. Software defined networking (SDN) is one of the methodologies for network management which is used to obtain network configuration which is efficient programmatically so that the network performance can be monitored and improved. The industries make use of the technology based on software defined networking for fault tolerance and network enhancement. The security of the network is the only question which has been more challenging in research community, whenever attacks are performed by the malicious users. The service level will be compromised by the network. Due to such attacks, data plane resources are consumed, and it affects functionality of the control plane of software defined network. In order to maintain the software defined network's secure and efficient, We proposed a packet injection and DOS attack controller software (PDACS) module. PDACS module will protect the controller of the network when it is attacked by remote malicious users. In this proposed work, the attacks like packet injection and denial of service attack will be stopped, and thereby, controller performance is enhanced using PDACS module. The real-time implementation of the SDN in the proposed work is done using Zodiac FX switch. The hosts and controller system of the software defined network is connected using Zodiac FX switch. We created a test bed in order to carry out this work, we considered various host like attacker, controller with Zodiac FX switch, preventer. The attackers host flood huge number of packets to the controller to stop its functionality. When the controller is down, the PDACS module identifies the port and IP address of the attacker and blocks it permanently. Thus, the controller resumes its operation. It is guaranteed that there is no compromise in the performance of PDACS even though the network is exposed to

T. G. Keerthan Kumar (✉)
Department of Information Science and Engineering, Siddaganaga Institute of Technology, Tumakuru, India

M. S. Srikanth · V. Sharma
Department of Information Science and Engineering, Nagarjuna College of Engineering and Technology, Bengaluru, India

J. Anand Babu
Department of Information Science and Engineering, Malnad College of Engineering and Technology, Hassan, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_60

high packet injection and DOS attack as there is a threshold bit rate at every port of the switch. If threshold is exceeded, it is considered as attack and is blocked. Hence, we can say performance of PDACS module in detecting and preventing DOS and packet injection attacks in SDN is more efficient.

Keywords Software defined network (SDN) · Mininet · Zodiac FX switch · PDACS module · DOS attack · Performance · Packet injection attack

1 Introduction

Nowadays, software defined network (SDN) is among the trending technologies in the domain of network management. The urge of using software defined network has been increased for network providers and users [1]. This urge for software defined network eliminates most of the traditional approaches. The traditional architectures of the network are becoming inadequate day by day because of the tremendous development of technologies like cloud computing, big data, IOT, etc. These technologies increase the load on the networks. Thus, software defined network technology is developed tremendously in the decade, although the SDN can be compromised by various attacks like denial of service attack (DOS) [2], and packet injection attack [3]. There are certain methods to prevent packet injection attack [1] and DOS attack, and one of them is to detect the packet flood. We can state that, although SDN is vulnerable to attacks, it can recover itself.

The network management can be achieved by making use of software defined networks (SDNs). The software define network contains a control plane which acts as centralized control and network management on large scale is handled by the control plane [2]. The SDNs are most commonly used when there are static network nodes, and the speed of connectivity of these nodes is high. The SDNs are used in order to overcome the limitations of the traditional architecture. This SDN architecture can be used for developing ad hoc and cloud networks [4].

1.1 Motivation

Among the fundamental advantages of software defined networking is that it deals with the system traffic effectively. Each and every switch in the network will be controlled and monitored by software defined network so that network performance, and services will be satisfactory. SDN comprises of two unique planes to be specific data plane and control plane. Some advantages of SDN make the job of network administrator easy. One of the important roles of network administrator is to configure and manage the devices in the network for conveying strategies to the switches [5]. In SDN, the network administrator needs to monitor only the controller for circulating the approaches to the switches that are associated.

Even though there are many advantages of software defined network, the major issue is the security of the SDN network. The attacker attacks on the network so that efficiency of the network will be diminished. The denial of service attack (DOS) is one of the attacks where one attempts to overload the system by flooding the packets [6]. The service and performance of the network are denied, and certain websites will not be available for service after performing denial of service attack. Packet injection is a concept used by the malicious user which reduces the performance of the software defined network [7]. Packet injection interferes the communication of the network.

2 Literature Survey

The concept of inspector device is discussed in [2] which is used to obstruct packet injection in SDN in [2]. In this work, the controller is safeguarded from packet injection attack using inspector device. The attack is stopped by the inspector device which is implemented through simulations which enhances the productivity of the controller. The distributed DOS attack is simulated using ns-2 simulator in [3]. In this paper, the performance of querying algorithms in the router during the attack is analyzed, and the remedies to overcome DOS attack are addressed. The denial of service attack in application layer is addressed in [1]. Nowadays, the application layer DOS attack is becoming significant than standard network attack. In this paper, the latest attacks like Domain Name System, Slow Loris, etc., are mentioned, and properties of DDOS attack are discussed.

The various methods to perform DDOS attack and different methods of overcoming the DDOS attack is discussed and different tools to overcome that can be used to overcome these attacks are addressed in [7]. In view of this work, the distributed DOS attack is mitigated and its performance is assessed. The details of distributed denial of service attack and techniques to overcome the attack are explained in [5]. The presented work explains that a network can be attacked through various techniques such as distributed denial of service attack, flooding the packets to the network, and so on. This paper also explains how to handle these attacks. The method to defend packet injection attack in SDN is discussed in [4]. In this work, SDN is protected from packet injection using packet checker. Along with this, a new module is used to identify and reduce flooding of packets. The mechanism to defend packet flooding is carried out and is evaluated to obtain the results. According to the obtained results, the packet checker performs well in reducing the attack. The denial of service attacks (DOS) in OpenFlow networks are categorized into two types, and they are discussed in [8]. The mininet is used to implement the DOS attack, and analysis of the attack is done in the proposed work. The dependency of capability of switch on time out, estimation of flow rule, control panel, and bandwidth are clarified, and lastly, a portion of the methodologies to resolve the attacks are explained. A system to detect intrusion in software defined networking (SDN) is implemented in [6]. The intrusion detection system implemented will diminish the attack when it occurs so that it does not affect the network infrastructure. The system identifies DDOS attacks, and it notifies the controller about the DDOS attack, and then the system will download the traffic forwarding which is necessary to the network.

The distributed controller concept to identify DOS attack in SDN and the revival of controller which is attacked is discussed in [9]. Along with DOS attack, the paper explains other issues which come along this attack. The mininet is used to simulate this approach. One of the important drawbacks of this system is that mitigation is not taken and latency is high. The BFT protocol-based distributed controller architecture is discussed in paper [10]. In this work, an SDN structure in which single controller will not manage every device in the network like traditional way instead there will be multiple controllers which have dynamic and isolated occurrence given by the cloud. There is redundancy in the architecture proposed in this work which becomes one of the drawbacks of the system.

The distributed controller based on cluster is explained in the article [11]; in this paper, an OpenFlow controller is proposed. This controller achieves reliability and scalability even when there are more loads on datacenter. The drawbacks of this work are related to security issues. In according to the traffic loads, the controllers are automatically increased or decreased. These controllers are distributed in nature, and it is defined as elastic distributed controller. The concept of elastic distributed controller is discussed in [12]. The drawbacks are there will be more load on the switch and some other security related issues.

Software defined network provides different architectures like flat model, hierarchical model, leaderless model, etc. These architectural options are discussed in [13]. Each one of the designs will have different influence of SDN issues; i.e., they have different influence on issues like scalability, robustness, and privacy. The drawback of this work is that there is no experimental proof. In order to implement an SDN setup, a small test bed along with Zodiac FX switch is explained in [14, 15]. In this work, using Zodiac FX switch, the SDN is implemented and also explained about working of SDN and various types of planes. For the hosts that are connected in the SDN must be protected from DOS attack. For this purpose, a protection mechanism is proposed in [16, 15]. In this work, in-switch processing is used to identify and stop the DDOS attack. The idea proposed in this work is a good strategy to handle DOS attack. The security issues of the controller are not considered in this work which is one of the drawbacks.

3 Proposed Systems

The software defined networking (SDN) is subjected to packet injection and DOS attacks by the malicious user. This system is designed to identify and control the above-mentioned attacks. In order to achieve this, a software defined network has to be established. In order to create the network, a switch names as Zodiac FX is used for forming software defined network. In the proposed work, four systems are connected through ports with the help of the switch, and one among them will be considered as controller. The controller to the network has to be decided further. The system which will be connected to controller port will be treated as controller.

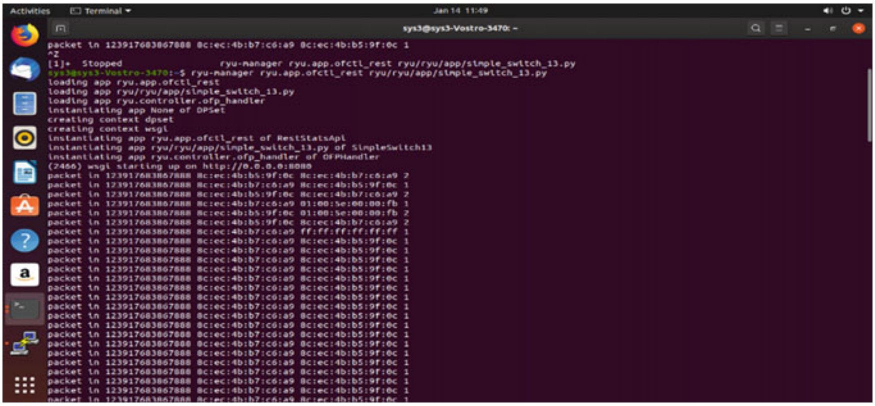
Here we use RYU controller acts as a framework for providing the software with application programming interfaces such as REST API (Representation State Transfer) that are required to play with OpenFlow concepts. Specifically, a python module named simple_switch_13 monitor the flows through the switch. When we start the RYU controller, it loads the required python modules as specified in the relative path given in the command. The transfer of packets from source to destination is monitored by controller. The controller will identify the destination of the packets when they arrive and then forwards the packets to the identified destinations. The intension of the attacker is to reduce the performance of the network. For achieving this, the attacker will infinitely number of packets to the network. When there is large number of packets, the controller will stop working. This situation is nothing but denial of service attack (DOS) [18]. This attack leads to great reduction in performance of the network or the disruption of the network.

In order to handle the denial of service attack (DOS), the attack must be identified and should be taken care of at early stages. Once the controller stops working, the system has to find the reason for the situation. The switch will be monitored in order to find the reason. If the bit rates are out of range from standard OpenFlow communication, then the occurrence of DOS attack will be confirmed. After identifying the attack, the job of the proposed system is to control the attack. The controller identifies the IP address of the attacker and blocks the attacker permanently. Thus, there will be no further attack from that attacker on the SDN network. Thus, using PDACS (Packet Injection and denial of service attack controller software) module, the proposed system identifies and stops the denial of service attack (DOS) attack.

The proposed system comprises of Zodiac FX switch, flow tables, control path, controller, and data paths. The packets which are sent from the sender will reach the controller first, and then it is forwarded to the destination from controller. During implementation proposed system, the message will be sent from one host system to other. In order to check the network connectivity, one can use ping messages. Ping messages will help to know if the network is broken. Once after checking network connectivity if network is not broken, the flows will get registered. The process of verifying whether the flows are registered correctly or not will be done by PUTTY command. After flows are registered properly, the system will be attacked by flooding the packets to establish the on the network. If the hosts become unreachable after flooding the packets, it confirms that attack is done successfully. After stopping the attack using PDACS, the hosts will again become reachable and flows will be normal as before attack.

4 System Implementation

The implementation is explained in four separate modules like setup, attack, detect, and prevent module.



```
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
^Z
[1]+  Stopped                 ryu-manager ryu.app.ofctl_rest ryu/ryu/app/simple_switch_13.py
sys3@sys3-Vostro-3470:~$ ryu-manager ryu.app.ofctl_rest ryu/ryu/app/simple_switch_13.py
loading app ryu.app.ofctl_rest
loading app ryu/ryu/app/simple_switch_13.py
loading app ryu.controller.ofp_handler
instantiating app None of DPSet
creating context dpset
creating context wsgi
instantiating app ryu.app.ofctl_rest of RestStatsApi
instantiating app ryu/ryu/app/simple_switch_13.py of SimpleSwitch13
instantiating app ryu.controller.ofp_handler of OFPHandler
(2466) wsgi starting up on http://0.0.0.0:8080
packet in 123917683867888 8c:ec:4b:b5:9f:0c 8c:ec:4b:b7:c6:a9 2
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b5:9f:0c 8c:ec:4b:b7:c6:a9 2
packet in 123917683867888 8c:ec:4b:b7:c6:a9 01:00:5e:00:00:fb 1
packet in 123917683867888 8c:ec:4b:b5:9f:0c 01:00:5e:00:00:fb 2
packet in 123917683867888 8c:ec:4b:b5:9f:0c 8c:ec:4b:b7:c6:a9 2
packet in 123917683867888 8c:ec:4b:b7:c6:a9 ff:ff:ff:ff:ff:ff 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
packet in 123917683867888 8c:ec:4b:b7:c6:a9 8c:ec:4b:b5:9f:0c 1
```

Fig. 1 Steps to installation

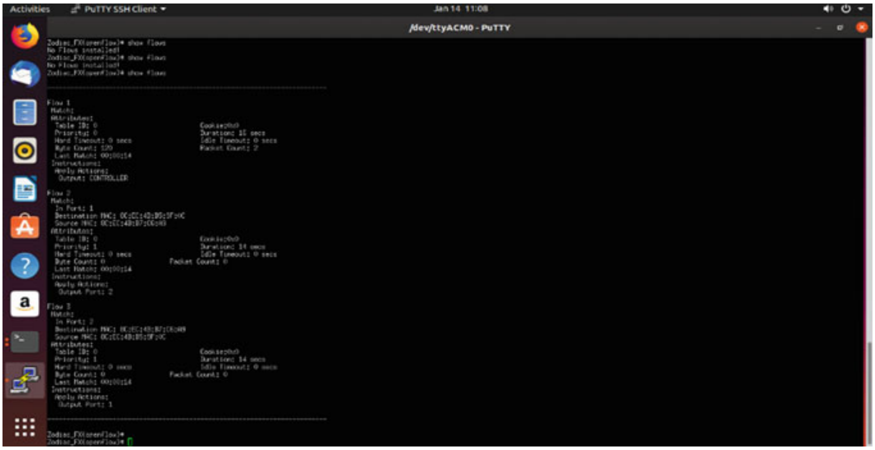
4.1 Setup

To create the test bed, we use four host system and the Zodiac FX Switch. All hosts are connected to ports of the switch. Let us name the hosts like host1, host2, host3, and host4, and the host-1 in the network is used to ping packets, host-2 acts as an attacker in the network, host-3 will receiver of the packets, and host 4 acts as controller in the network.

- **Steps to Installation**
  Open Flows to the network are installed as shown in Fig. 1. After installing the open flows, connection will be established between host and controller in the network.
- **Representation of Flows in the network**
  The installation of flows in the network is as depicted in Fig. 2. It gives the host, packet count, and destination addresses of the network.
- **Packets Transmission from host to destination**
  Packets are sent from source to destination in the network using ping command. The packet transmission takes place as shown Fig. 3.
- **To check the packet count**
  In order to check whether packets reached the destination, we can use 'show flow' command in PUTTY. This command shows that the packet count will be increased every second as shown in Fig. 4 if there is no problem in transmission of packets.

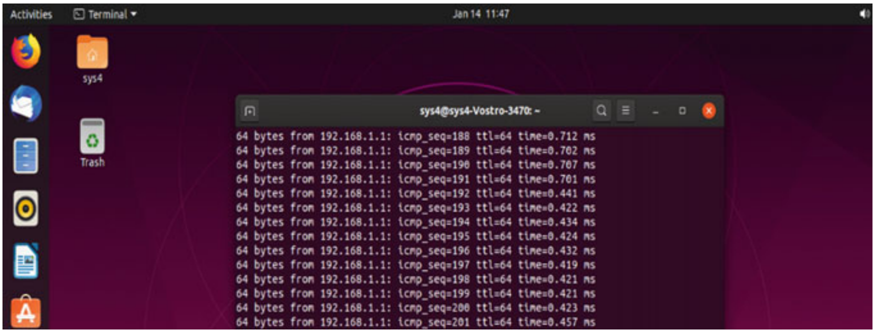
4.2 Attack

Here we show how to perform attacks on the SDN. As we are demonstrating the attack by pretending, we are the attackers, we use the existing host as attacker. The



```
Zodiac_FX(openflow)# show flows
No Flows installed!
Zodiac_FX(openflow)# show flows
No Flows installed!
Zodiac_FX(openflow)# show flows

-------------------------------------------------------------------------------

Flow 1
 Match:
 Attributes:
  Table ID: 0                             Cookie:0x0
  Priority: 0                             Duration: 15 secs
  Hard Timeout: 0 secs                    Idle Timeout: 0 secs
  Byte Count: 120                         Packet Count: 2
  Last Match: 00:00:14
 Instructions:
  Apply Actions:
   Output: CONTROLLER

Flow 2
 Match:
  In Port: 1
  Destination MAC: 0C:CC:4B:B5:5F:0C
  Source MAC: 0C:CC:4B:B7:C6:A9
 Attributes:
  Table ID: 0                             Cookie:0x0
  Priority: 1                             Duration: 14 secs
  Hard Timeout: 0 secs                    Idle Timeout: 0 secs
  Byte Count: 0                   Packet Count: 0
  Last Match: 00:00:14
 Instructions:
  Apply Actions:
   Output Port: 2

Flow 3
 Match:
  In Port: 2
  Destination MAC: 0C:CC:4B:B7:C6:A9
  Source MAC: 0C:CC:4B:B5:5F:0C
 Attributes:
  Table ID: 0                             Cookie:0x0
  Priority: 1                             Duration: 14 secs
  Hard Timeout: 0 secs                    Idle Timeout: 0 secs
  Byte Count: 0                   Packet Count: 0
  Last Match: 00:00:14
 Instructions:
  Apply Actions:
   Output Port: 1

-------------------------------------------------------------------------------

Zodiac_FX(openflow)#
Zodiac_FX(openflow)#
```

Fig. 2 Flows in the network



```
64 bytes from 192.168.1.1: icmp_seq=188 ttl=64 time=0.712 ms
64 bytes from 192.168.1.1: icmp_seq=189 ttl=64 time=0.702 ms
64 bytes from 192.168.1.1: icmp_seq=190 ttl=64 time=0.707 ms
64 bytes from 192.168.1.1: icmp_seq=191 ttl=64 time=0.701 ms
64 bytes from 192.168.1.1: icmp_seq=192 ttl=64 time=0.441 ms
64 bytes from 192.168.1.1: icmp_seq=193 ttl=64 time=0.422 ms
64 bytes from 192.168.1.1: icmp_seq=194 ttl=64 time=0.434 ms
64 bytes from 192.168.1.1: icmp_seq=195 ttl=64 time=0.424 ms
64 bytes from 192.168.1.1: icmp_seq=196 ttl=64 time=0.432 ms
64 bytes from 192.168.1.1: icmp_seq=197 ttl=64 time=0.419 ms
64 bytes from 192.168.1.1: icmp_seq=198 ttl=64 time=0.421 ms
64 bytes from 192.168.1.1: icmp_seq=199 ttl=64 time=0.421 ms
64 bytes from 192.168.1.1: icmp_seq=200 ttl=64 time=0.423 ms
64 bytes from 192.168.1.1: icmp_seq=201 ttl=64 time=0.457 ms
```

Fig. 3 Packets transmission from host to destination

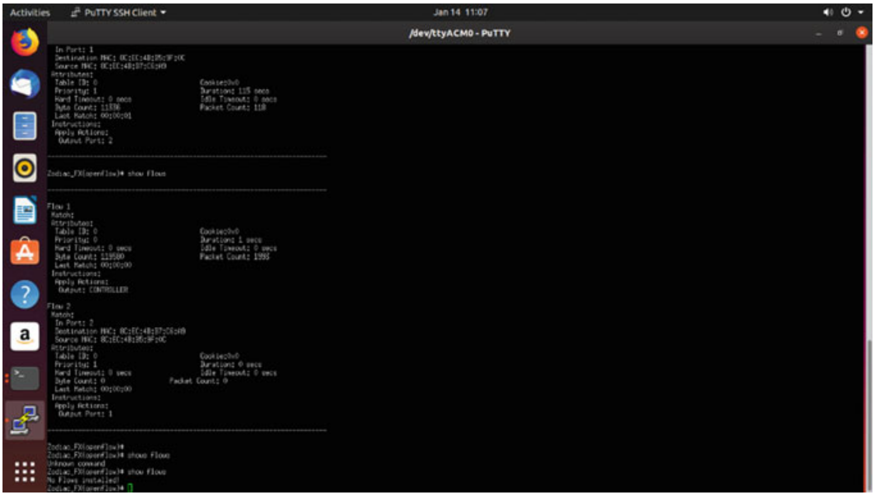
attacker can choose any port number and can perform the attack remotely without even disclosing his actual IPV4 address. The attacker can send multiple packets to one particular port and may stop the functionality of the controller (Fig. 5).

Procedure to perform attack on the network

To perform attack on network hping3 command is used. After performing hping3, the packets will be flooded which blocks the controller so that attack will be successful.

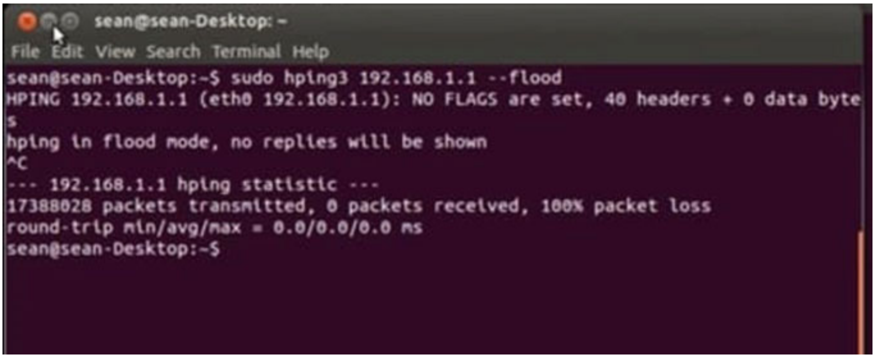
4.3 *Detect*

Here we shall look into the algorithm for detecting the attack by PDACS module.



```
  In Port: 1
  Destination MAC: 8C:EC:4B:B5:9F:0C
  Source MAC: 8C:EC:4B:B7:C6:A9
 Attributes:
  Table ID: 0                          Cookie:0x0
  Priority: 1                          Duration: 115 secs
  Hard Timeout: 0 secs                 Idle Timeout: 0 secs
  Byte Count: 11336                    Packet Count: 118
  Last Match: 00:00:01
 Instructions:
  Apply Actions:
   Output Port: 2

-------------------------------------------------------------------------------

Zodiac_FX(openflow)# show flows

-------------------------------------------------------------------------------

Flow 1
 Match:
 Attributes:
  Table ID: 0                          Cookie:0x0
  Priority: 0                          Duration: 1 secs
  Hard Timeout: 0 secs                 Idle Timeout: 0 secs
  Byte Count: 119580                   Packet Count: 1993
  Last Match: 00:00:00
 Instructions:
  Apply Actions:
   Output: CONTROLLER

Flow 2
 Match:
  In Port: 2
  Destination MAC: 8C:EC:4B:B7:C6:A9
  Source MAC: 8C:EC:4B:B5:9F:0C
 Attributes:
  Table ID: 0                          Cookie:0x0
  Priority: 1                          Duration: 0 secs
  Hard Timeout: 0 secs                 Idle Timeout: 0 secs
  Byte Count: 0               Packet Count: 0
  Last Match: 00:00:00
 Instructions:
  Apply Actions:
   Output Port: 1

-------------------------------------------------------------------------------

Zodiac_FX(openflow)#
Zodiac_FX(openflow)# show flows
Unknown command
Zodiac_FX(openflow)# show flows
No Flows installed!
Zodiac_FX(openflow)#
```

Fig. 4 To check the packet count



```
sean@sean-Desktop:~$ sudo hping3 192.168.1.1 --flood
HPING 192.168.1.1 (eth0 192.168.1.1): NO FLAGS are set, 40 headers + 0 data byte
s
hping in flood mode, no replies will be shown
^C
--- 192.168.1.1 hping statistic ---
17388028 packets transmitted, 0 packets received, 100% packet loss
round-trip min/avg/max = 0.0/0.0/0.0 ms
sean@sean-Desktop:~$
```

Fig. 5 Procedure to perform attack on the network

Algorithm 1 Detection of the Packet Injection Attack

```
INPUTS: Packet (packets incoming), Table (IP address and port no. mapping table), Switch
Output: A (action to be done)
for each Packet ∈ Switch do
    if Packet.IP ∈ Table then
    SET Table.IP = Packet.IP
    SET Table.Port = Packet.Port
    SET Table.ID = Packet.ID
    return continue
    end if
      if Packet.IP  ∈ Table and Table.Port = Packet.Port and Table.DPID =Packet.DPID  then
    return continue
    else
    return stop
    end if
end for
```

This means that, if there is an entry already present in the flow table just continue the process. If else then stop the process. The incoming packet's MAC address should match the address in the mapping table along with its port number. Let us understand this with pictorial reference.

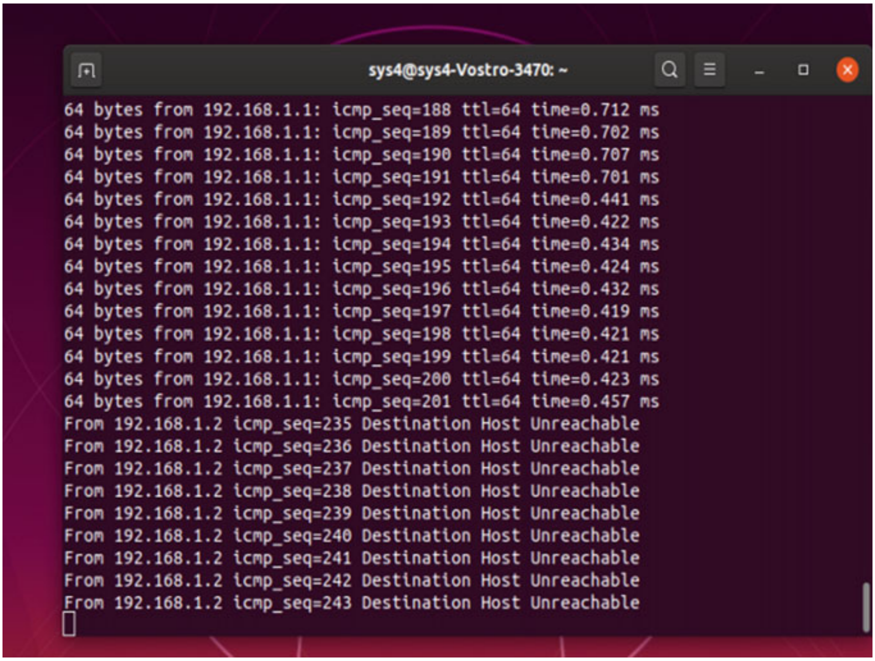
- **Dos attack in SDN**
  Once the attacker performs the attack, the packets which were reachable from host to destination becomes unreachable as shown in Fig. 6, and after the attacker attacks the network if the flows for connection establishment is checked once again, then it shows no flows installed as shown in Fig. 7.

4.4 *Prevent*

Now we shall see how our PDACS module will respond to the attack and how it will prevent and recover from the attack.

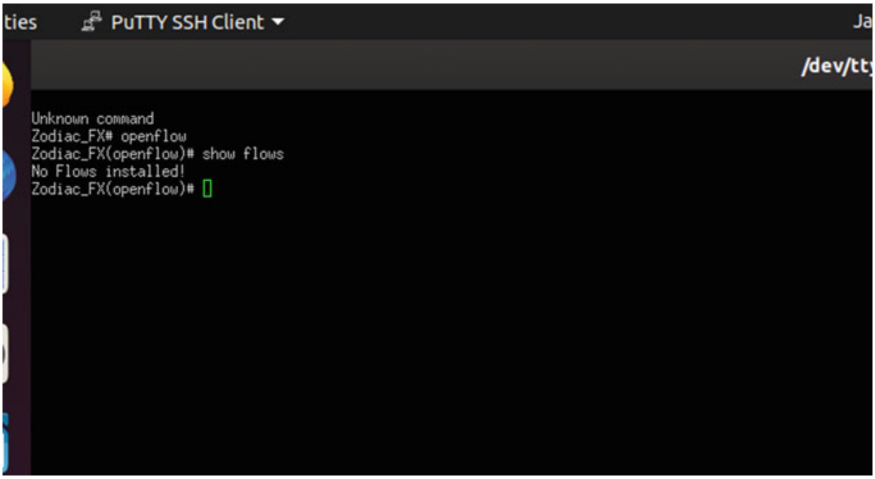
Algorithm 2 Prevention of the Packet Injection Attack

```
INPUT: Port (Port no.), B (Bit rate), IP_P (IPV4 address of a packet), P(packet), S(switch)
OUTPUT: C (Command)
SET Threshold_BR =200 Mbts/s
for each P in S :
      if P.B    <= Threshold_BR then

            return continue
      else
            SET Attacker_IP=P.IP_P
            SET Attacker_Port=P.Port
            BLOCK(Attacker)
            return continue
      end if
end for
```



```
64 bytes from 192.168.1.1: icmp_seq=188 ttl=64 time=0.712 ms
64 bytes from 192.168.1.1: icmp_seq=189 ttl=64 time=0.702 ms
64 bytes from 192.168.1.1: icmp_seq=190 ttl=64 time=0.707 ms
64 bytes from 192.168.1.1: icmp_seq=191 ttl=64 time=0.701 ms
64 bytes from 192.168.1.1: icmp_seq=192 ttl=64 time=0.441 ms
64 bytes from 192.168.1.1: icmp_seq=193 ttl=64 time=0.422 ms
64 bytes from 192.168.1.1: icmp_seq=194 ttl=64 time=0.434 ms
64 bytes from 192.168.1.1: icmp_seq=195 ttl=64 time=0.424 ms
64 bytes from 192.168.1.1: icmp_seq=196 ttl=64 time=0.432 ms
64 bytes from 192.168.1.1: icmp_seq=197 ttl=64 time=0.419 ms
64 bytes from 192.168.1.1: icmp_seq=198 ttl=64 time=0.421 ms
64 bytes from 192.168.1.1: icmp_seq=199 ttl=64 time=0.421 ms
64 bytes from 192.168.1.1: icmp_seq=200 ttl=64 time=0.423 ms
64 bytes from 192.168.1.1: icmp_seq=201 ttl=64 time=0.457 ms
From 192.168.1.2 icmp_seq=235 Destination Host Unreachable
From 192.168.1.2 icmp_seq=236 Destination Host Unreachable
From 192.168.1.2 icmp_seq=237 Destination Host Unreachable
From 192.168.1.2 icmp_seq=238 Destination Host Unreachable
From 192.168.1.2 icmp_seq=239 Destination Host Unreachable
From 192.168.1.2 icmp_seq=240 Destination Host Unreachable
From 192.168.1.2 icmp_seq=241 Destination Host Unreachable
From 192.168.1.2 icmp_seq=242 Destination Host Unreachable
From 192.168.1.2 icmp_seq=243 Destination Host Unreachable
```

Fig. 6 Dos attack in SDN



```
Unknown command
Zodiac_FX# openflow
Zodiac_FX(openflow)# show flows
No Flows installed!
Zodiac_FX(openflow)#
```

Fig. 7 Dos attack in SDN

What we can understand from this algorithm is that if the bit rate is higher than threshold at any port, the algorithm sets/assigns the number to the port and IP address of that packet as port number of the attacker and IP address of attacker and blocks it; otherwise, the packet is forwarded.

$$\text{BitRate}(B) = \frac{\text{Number of Packets}}{\text{unit time}}\text{Mbps} \tag{1}$$

If the attacker attacks SDN, he should flood packets in large number. So, the bandwidth (Eq. 1) will be definitely large (than 100Mbps). We wrote a python code to control and block all the bandwidth above the threshold. Thus, we stop attacks by PDACS module.

We have set the threshold bitrate as 200Mbits/s. If the rate of number of packets per second is more than the threshold, then PDACS module blocks the attacker.

Let us understand with pictorial description.

PDACS module to handle the attack

After attack is performed, PDACS module will be run as shown in Fig. 8. This module is used to overcome the DOS attack, and the connection between the hosts and destination will be re-established and it will in turn make destination is reachable from the connected other hosts.

To check the packet count after the Dos attack is handled

After Dos attack is handled, the packet count and connection establishment are rechecked using PUTTY as shown in Fig. 9.

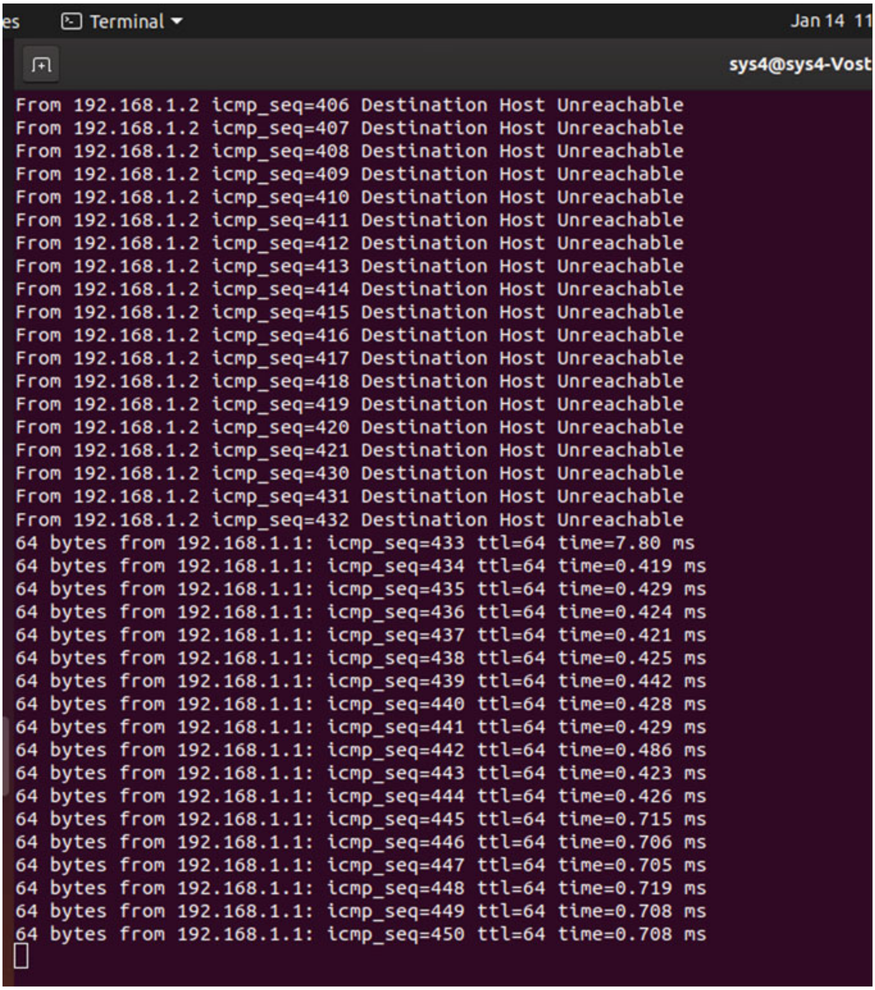
5 Simulation

The tool used to simulate our project is mininet. In mininet, we had written the python code to create our own topology add OpenFlow switches, hosts, controller, and links between them. We can use MINIEDIT to have the GUI experience rather a terminal look-up is depicted in Fig. 10.

The same procedure follows here too. First, we ping all the hosts and check reachability as shown in Fig. 11. If they are reachable, we now attack and stop the attack as mentioned in above case.

6 Results

Here we focus on the results of the PDACS module to check whether it can handle attacks in network. The results mainly indicate how well the PDACS module prevents the attack. We shall now evaluate the performance of the module.



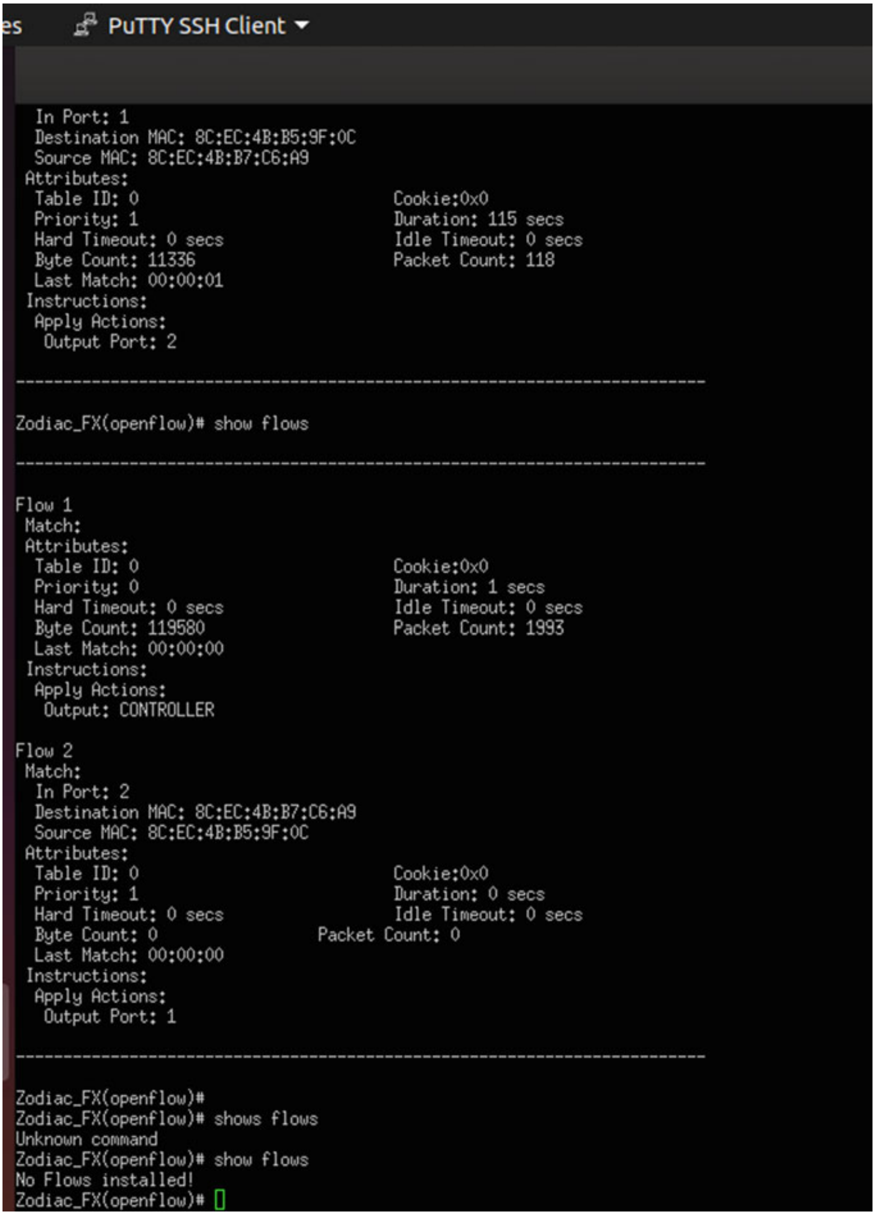
```
From 192.168.1.2 icmp_seq=406 Destination Host Unreachable
From 192.168.1.2 icmp_seq=407 Destination Host Unreachable
From 192.168.1.2 icmp_seq=408 Destination Host Unreachable
From 192.168.1.2 icmp_seq=409 Destination Host Unreachable
From 192.168.1.2 icmp_seq=410 Destination Host Unreachable
From 192.168.1.2 icmp_seq=411 Destination Host Unreachable
From 192.168.1.2 icmp_seq=412 Destination Host Unreachable
From 192.168.1.2 icmp_seq=413 Destination Host Unreachable
From 192.168.1.2 icmp_seq=414 Destination Host Unreachable
From 192.168.1.2 icmp_seq=415 Destination Host Unreachable
From 192.168.1.2 icmp_seq=416 Destination Host Unreachable
From 192.168.1.2 icmp_seq=417 Destination Host Unreachable
From 192.168.1.2 icmp_seq=418 Destination Host Unreachable
From 192.168.1.2 icmp_seq=419 Destination Host Unreachable
From 192.168.1.2 icmp_seq=420 Destination Host Unreachable
From 192.168.1.2 icmp_seq=421 Destination Host Unreachable
From 192.168.1.2 icmp_seq=430 Destination Host Unreachable
From 192.168.1.2 icmp_seq=431 Destination Host Unreachable
From 192.168.1.2 icmp_seq=432 Destination Host Unreachable
64 bytes from 192.168.1.1: icmp_seq=433 ttl=64 time=7.80 ms
64 bytes from 192.168.1.1: icmp_seq=434 ttl=64 time=0.419 ms
64 bytes from 192.168.1.1: icmp_seq=435 ttl=64 time=0.429 ms
64 bytes from 192.168.1.1: icmp_seq=436 ttl=64 time=0.424 ms
64 bytes from 192.168.1.1: icmp_seq=437 ttl=64 time=0.421 ms
64 bytes from 192.168.1.1: icmp_seq=438 ttl=64 time=0.425 ms
64 bytes from 192.168.1.1: icmp_seq=439 ttl=64 time=0.442 ms
64 bytes from 192.168.1.1: icmp_seq=440 ttl=64 time=0.428 ms
64 bytes from 192.168.1.1: icmp_seq=441 ttl=64 time=0.429 ms
64 bytes from 192.168.1.1: icmp_seq=442 ttl=64 time=0.486 ms
64 bytes from 192.168.1.1: icmp_seq=443 ttl=64 time=0.423 ms
64 bytes from 192.168.1.1: icmp_seq=444 ttl=64 time=0.426 ms
64 bytes from 192.168.1.1: icmp_seq=445 ttl=64 time=0.715 ms
64 bytes from 192.168.1.1: icmp_seq=446 ttl=64 time=0.706 ms
64 bytes from 192.168.1.1: icmp_seq=447 ttl=64 time=0.705 ms
64 bytes from 192.168.1.1: icmp_seq=448 ttl=64 time=0.719 ms
64 bytes from 192.168.1.1: icmp_seq=449 ttl=64 time=0.708 ms
64 bytes from 192.168.1.1: icmp_seq=450 ttl=64 time=0.708 ms
```

Fig. 8 PDACS module to handle the attack

6.1 Performance Evaluation

In this proposal, we are making controller to act as a self-repellent for the remote attacks whenever attacks are done. To achieve this, we use the choice to limit the bit rate. We identify all the packets having greater bandwidth than the threshold limit, and controller is able to block effectively all the packets having bandwidth greater the threshold. We can consider the bit rate, and it is the number of incoming packets at a particular time period as an attribute and analyze how attacker does the attack.

We plot a graph considering the data from Table 1 by making time in seconds as *X*-axis (from 0 to 70 s) and packet count (from 0 to 1700 packets) as *Y*-axis.



```
  In Port: 1
  Destination MAC: 8C:EC:4B:B5:9F:0C
  Source MAC: 8C:EC:4B:B7:C6:A9
 Attributes:
  Table ID: 0                          Cookie:0x0
  Priority: 1                          Duration: 115 secs
  Hard Timeout: 0 secs                 Idle Timeout: 0 secs
  Byte Count: 11336                    Packet Count: 118
  Last Match: 00:00:01
 Instructions:
  Apply Actions:
   Output Port: 2

-------------------------------------------------------------------------

Zodiac_FX(openflow)# show flows

-------------------------------------------------------------------------

Flow 1
 Match:
 Attributes:
  Table ID: 0                          Cookie:0x0
  Priority: 0                          Duration: 1 secs
  Hard Timeout: 0 secs                 Idle Timeout: 0 secs
  Byte Count: 119580                   Packet Count: 1993
  Last Match: 00:00:00
 Instructions:
  Apply Actions:
   Output: CONTROLLER

Flow 2
 Match:
  In Port: 2
  Destination MAC: 8C:EC:4B:B7:C6:A9
  Source MAC: 8C:EC:4B:B5:9F:0C
 Attributes:
  Table ID: 0                          Cookie:0x0
  Priority: 1                          Duration: 0 secs
  Hard Timeout: 0 secs                 Idle Timeout: 0 secs
  Byte Count: 0                Packet Count: 0
  Last Match: 00:00:00
 Instructions:
  Apply Actions:
   Output Port: 1

-------------------------------------------------------------------------

Zodiac_FX(openflow)#
Zodiac_FX(openflow)# shows flows
Unknown command
Zodiac_FX(openflow)# show flows
No Flows installed!
Zodiac_FX(openflow)#
```

Fig. 9 To check the packet count after the Dos attack is handled

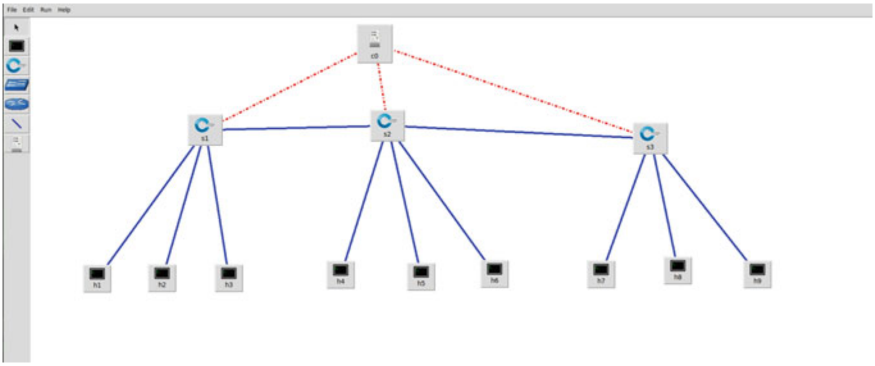
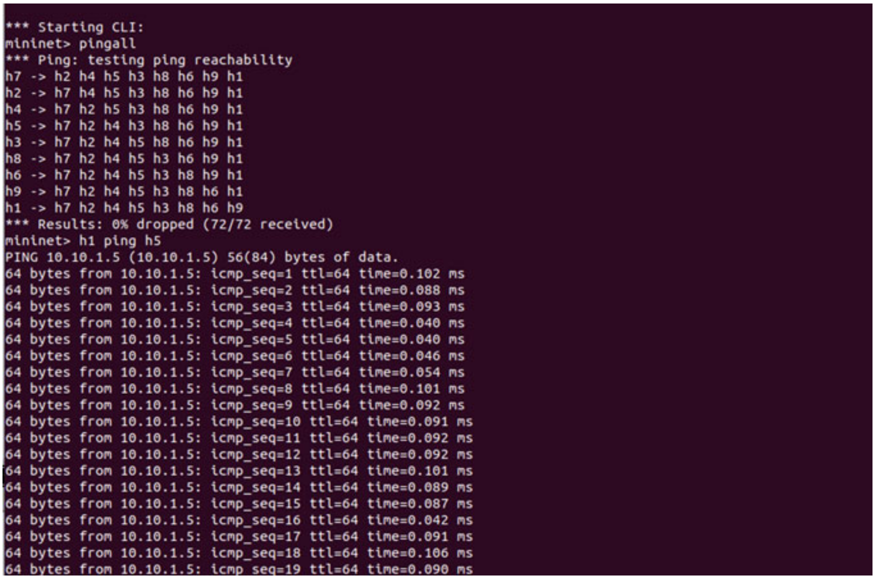


Fig. 10 MINIEDIT topology



```
*** Starting CLI:
mininet> pingall
*** Ping: testing ping reachability
h7 -> h2 h4 h5 h3 h8 h6 h9 h1
h2 -> h7 h4 h5 h3 h8 h6 h9 h1
h4 -> h7 h2 h5 h3 h8 h6 h9 h1
h5 -> h7 h2 h4 h3 h8 h6 h9 h1
h3 -> h7 h2 h4 h5 h8 h6 h9 h1
h8 -> h7 h2 h4 h5 h3 h6 h9 h1
h6 -> h7 h2 h4 h5 h3 h8 h9 h1
h9 -> h7 h2 h4 h5 h3 h8 h6 h1
h1 -> h7 h2 h4 h5 h3 h8 h6 h9
*** Results: 0% dropped (72/72 received)
mininet> h1 ping h5
PING 10.10.1.5 (10.10.1.5) 56(84) bytes of data.
64 bytes from 10.10.1.5: icmp_seq=1 ttl=64 time=0.102 ms
64 bytes from 10.10.1.5: icmp_seq=2 ttl=64 time=0.088 ms
64 bytes from 10.10.1.5: icmp_seq=3 ttl=64 time=0.093 ms
64 bytes from 10.10.1.5: icmp_seq=4 ttl=64 time=0.040 ms
64 bytes from 10.10.1.5: icmp_seq=5 ttl=64 time=0.040 ms
64 bytes from 10.10.1.5: icmp_seq=6 ttl=64 time=0.046 ms
64 bytes from 10.10.1.5: icmp_seq=7 ttl=64 time=0.054 ms
64 bytes from 10.10.1.5: icmp_seq=8 ttl=64 time=0.101 ms
64 bytes from 10.10.1.5: icmp_seq=9 ttl=64 time=0.092 ms
64 bytes from 10.10.1.5: icmp_seq=10 ttl=64 time=0.091 ms
64 bytes from 10.10.1.5: icmp_seq=11 ttl=64 time=0.092 ms
64 bytes from 10.10.1.5: icmp_seq=12 ttl=64 time=0.092 ms
64 bytes from 10.10.1.5: icmp_seq=13 ttl=64 time=0.101 ms
64 bytes from 10.10.1.5: icmp_seq=14 ttl=64 time=0.089 ms
64 bytes from 10.10.1.5: icmp_seq=15 ttl=64 time=0.087 ms
64 bytes from 10.10.1.5: icmp_seq=16 ttl=64 time=0.042 ms
64 bytes from 10.10.1.5: icmp_seq=17 ttl=64 time=0.091 ms
64 bytes from 10.10.1.5: icmp_seq=18 ttl=64 time=0.106 ms
64 bytes from 10.10.1.5: icmp_seq=19 ttl=64 time=0.090 ms
```

Fig. 11 Command line interface of ping command in mininet

Here we can observe in Fig. 12 that the attacker floods the packets in small interval of time and tries to collapse the controller. Now let us see how the PDACS module responds to this attack by plotting a graph taking time in seconds as *X*-axis (from 0 to 16 s) and packet count as *Y*- axis (0–2000 packets) as in Fig. 13.

We plotted the graphs using Wireshark tool. It produces the packet count per second, and we can plot it simultaneously. We initially plotted the graph when the attack was done. When the attack was initiated, there was an extreme rise in the packet count as you can see in Fig. 12. Right after a certain period of time, the controller stops working due to the large number of packets coming into the switch. It cannot

Table 1 Packet count after the attack

Sl.no	Time(s)	Packet Count
	0	0
	5	400
	10	750
	15	600
	20	800
	25	790
	30	900
	35	1000
	40	1300
	45	1490
	50	1630
	55	1800
	60	1700

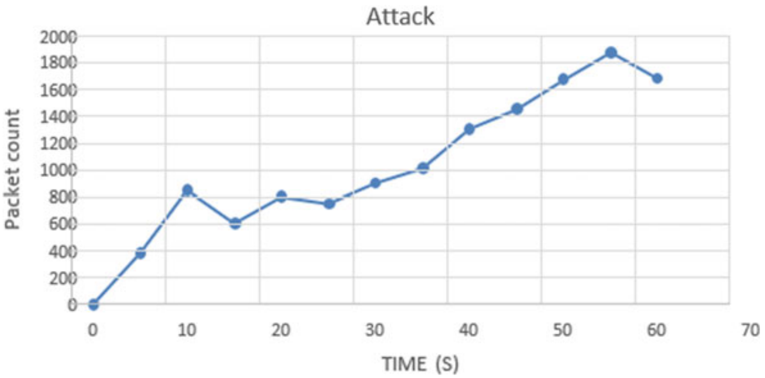


Fig. 12 Graph indicating attack

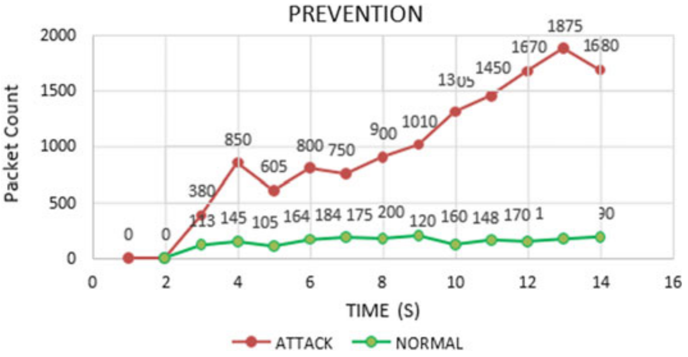


Fig. 13 PDACS preventing the attacks

process the large number of packets and the legitimate packets coming from other ports too. Now the attack continues, and the network is down. The PDACS responds to this attack and by checking the rate at which the packets are entering the port which is nothing but the bit rate. The PDACS module allows the packets having lesser bit (Green Line) rate than the threshold, and when the bit rate is high, it blocks them gradually as shown in Fig. 13 (Red Line) varying magnitude of packet count). Thus, only legitimate packets are processed further from the switch. The attack holds for short period of time. By the time the network is down, the module will check the packet's bit rate, and it blocks/allows the packet. Thus, the network is restored using PDACS. What we can incur from the graph is, only legitimate packets (whose bitrate is legitimate) are allowed and processed further and packets from attacker's end are blocked permanently.

7 Conclusion

In the proposed system, SDN network test bed is created using Zodiac FX switch. Zodiac, there are four ports in the switch, and one port will be treated as controller. Using this Dos attack will be simulated on the network and then controlled using PDACS module. This module focuses on the cut-off for bit rates, and it boosts the detection of the attack and also reduces the damage of entire network; i.e., if the bit rate is higher than threshold value, the packet is blocked. As of now, the project operates on single controller; but in coming days, we work with multiple controllers so that some of them can be used as a backup controller. This can be deployed if any one of the controllers are down. The two or more controllers shall work on a principle called load-balancing (dividing the packet load). Thus, we conclude that although SDN is exposed to numerous attacks, it can recover itself effectively by incorporating PDACS module.

References

1. M.J. Anagha, R. Lepakshi, V. Goutham, V. Thavish, T.G. Keerthan Kumar, Packet injection and dos attack controller software(PDACS) module to handle attacks in software defined network, in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (Erode, India, 2020), pp. 966–970. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000179
2. D. Kreutz, F.M. Ramos, P.E. Verissimo, C.E. Rothenberg, S. Azodolmolky, S. Uhlig,Software-defined networking: A comprehensive survey, in *Proceedings of the IEEE* Vol. 103, no. 1, pp. 14–76 (2014)
3. A.S. Alshra, J. Seitz, Using INSPECTOR device to stop packet injection attack in SDN. IEEE Commun. Letters **23** (2019)

4. T.G. Keerthan Kumar, H.K. Virupakshaiah, K.V. Nanda, Ensuring an online chat mechanism with accountability to sharing the non-downloadable file from the cloud, in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)* (2016), pp. 718–721. https://doi.org/10.1109/ICATCCT.2016.7912093
5. P. Manso, J. Moura, C. Serrao, SDN-based intrusion detection system for early detection and mitigation of DDoS attacks (2019)
6. M. Antikainen, T. Aura, Spook in your network: Attacking an SDN with a Compromised OpenFlow switch, in *19th Nordic Conference on Secure IT Systems* (Oct 2014)
7. S. Deng, X. Gao, Z. Lu, X. Gao, Packet injection attack and its defense in software-defined networks. IEEE Trans. Inf. Forensics Secur. **13** (2018)
8. D. Kreutz, F.M. Ramos, P. Verissimo, C.E. Rothenberg, S. Azodolmokly, S. Uhlig, Software-defined networking: A comprehensive survey. Computing Research Repository (CoRR), (June 2014). Accessed on 13 Nov 2014
9. D. Kreutz, F.M. Ramos, P. Verissimo, Towards secure and dependable software-defined networks, in *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, ser. HotSDN'13* (USA:ACM, NY, 2013), pp. 55–60
10. R. Kloti, V. Kotronis, P. Smith, OpenFlow: A security analysis, in *21st IEEE International Conference on Network Protocols* (2013)
11. N. Shastry, T.G. Keerthan Kumar, Enhancing the performance of software-defined wireless mesh network, in *International Conference on Communication, Computing and Electronics Systems. Lecture Notes in Electrical Engineering*, eds. by V. Bindhu, J. Chen, J. Tavares, vol. 637 (Springer, Singapore, 2020)
12. S. R. Mugunthan, Soft computing based autonomous low rate DDOS attack detection and security for cloud computing. J. Soft Comput. Paradigm (jscp) **1**(02), 80–90 (2019)
13. S. Smys, DDOS attack detection in telecommunication network using machine learning. J. Ubiquit. Comput. Commun. Technol. (ucct) **1**(01), 33-44 (2019)
14. CpQD OpenFlow 1.3 Software Switch, http://cpqd.github.io/ofsoftswitch13/. Accessed on 13 Nov 2014
15. K.S. Bhosale, M. Nenova, G. Iliev, The distributed denial of service attacks prevention mechanism on application layer, in *13th International Conference on Advanced Technologies, Systems and services in Telecommunications (TELSIKS)* (2017)
16. F. Lau, S.H. Smith, L. Taikovic, Distributed denial of service attacks, in *2000 IEEE International Conference on systems, man and cybernetics Smc 2000 Conference proceedings* (2002)
17. K.S. Vanitha, S.V. Uma, S.K. Mahidhar, Distributed denial of service: Attack techniques and mitigation, in *International Conference on Circuits, Controls, and Communications (CCUBE)* (2017)
18. K. Narasimha Mallikarjunan, K. Muthupriya, S. Mercy shalinie, A survey of distributed denial of service attack, in *10th International Conference on Intelligent Systems and Control (ISCO)* (2016)

SLAP-IoT: A Secure Lightweight Authentication Protocol for IoT Device

CH.N.S. Abhishek, Chungath Srinivasan, Lakshmy K.V., and P. Mohan Anand

Abstract Internet of Things (IoT) has evolved on a large scale and is widely being used across all the industries in various sectors. The IoT devices have a limited capacity in terms of memory and computational ability. Compared to other network applications, providing security for IoT device communication is a relatively more difficult task. The risk of getting prone to attacks can be minimized by implementing a robust authentication mechanism. To achieve it, we are proposing a lightweight authentication protocol. The security analysis was conducted using the Scyther tool, which proves that the mechanism proposed is secure against replay, session key disclosure and impersonation attacks. Moreover, the performance of the proposed protocol has been analysed and evaluated with other protocols in terms of communication cost.

Keywords IoT authentication · Scyther tool · Lightweight protocol · Two party protocol

CH.N.S. Abhishek (✉) · C. Srinivasan · L. K.V.
TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2cys19003@cb.students.amrita.edu

C. Srinivasan
e-mail: c_srinivasan@cb.amrita.edu

L. K.V.
e-mail: kv_lakshmy@cb.amrita.edu

P. M. Anand
Department of Computer science and Engineering, Indian Institute of Technology Kanpur, Kanpur, India
e-mail: pmohan20@iitk.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_61

1 Introduction

The use of IoT has expanded in popularity, and it has become part and parcel of our daily lives. As the number of IoT devices is increasing day by day, in 2016, there were around 4.6 billion devices, and by the end of 2021, the number would reach 13.8 billion. By 2025, it is estimated that approximately 30 billion IoT devices will exist [1]. When IoT was first introduced, less emphasis was placed on its security. As the Internet of Things is built on data that is mostly private and highly sensitive, it has the potential to be exploited and, thus, violating the user's privacy [2–4, 24]. As a result, the security of IoT devices is being prioritized. IoT devices can be secured by employing various mechanisms where authentication plays a vital role. It can help to reduce risk and guarantee that IoT devices are trustworthy. In authentication, identity of the devices is recognized and further validated. Authentication is performed in the initial phase so that communication between devices may begin, and each device can learn about the identity of the other. If authentication is not performed securely, attackers can gain access to the machines and can steal the data generated and transferred, which results in various attacks.

In IoT networks, authentication attacks are a major concern. The classification of attacks is grouped into various categories namely denial-of-service attack, masquerade attack, forging attack, man-in-the-middle attack, guessing attack, physical attack and routing attack [5]. There are various challenges with IoT authentication that must be addressed [6]. The first challenge is to cut energy costs during the process of authentication. ECC is an authentication mechanism that leverages implicit certificates to reduce energy usage and computing overhead [7] in sensor networks for distributed IoT applications. The second challenge [8] is the implementation of IoT-adapted authentication mechanisms. Distinct network architectures are based on different concepts of IoT, and authentication mechanisms need to be implemented to secure the communication [9]. Another challenge is devising an authentication mechanism capable of identifying users in their devices while avoiding persistent interaction between those components [10]. The authentication protocols in the IoT environment should essentially avail the limited amount of memory or several bits. If an enormous amount of memory is consumed, then abundant resources are utilized for implementation, and the computational cost of the protocol increases. The stand out feature in our proposed model is a lightweight authentication protocol that consumes 2948 bits.

This paper is formatted as follows. Section 2 primarily concentrates on the related works, and Sect. 3 gives a detailed overview of the protocol that is being proposed. Section 4 describes the outcomes and properties, and finally, Sect. 5 summarizes our conclusions and lists the possible future work that can be done in this area.

2 Related Work

Over the past years, many authentication mechanisms were proposed with varied architectures in the IoT environment. The main motto behind this was to develop a secure IoT system which is resilient against attacks. Kumari et al. proposed an ECC-based authentication system for IoT and cloud servers [11]. Automated Validation of Internet Security Protocols and Applications tool was employed to formally examine the security features of the suggested scheme. Security and performance review demonstrated that the proposed model is more effective, reliable and stable than existing models in the face of a variety of known attacks. A lightweight authentication protocol for IoT devices with a three-tiered architecture was proposed by Ali et al. [12]. In their mechanism, the number of positive and hostile acts was used to calculate the device's trust using a fuzzy method. The findings demonstrate the suggested protocol's advantage over other techniques in terms of attack resistance. Yang et al. proposed an authentication scheme for multi-server architecture using a smart card [20]. This mechanism combines the benefits of biometrics and password authentication. Session key disclosure attack is possible in this scheme. An authentication protocol for multi-server architecture was proposed by Li et al. [22]. Unfortunately, this protocol is vulnerable to replay and impersonation attacks. Dammak et al. proposed a decentralized mechanism for group key management employing one key distribution centre and various subkey distribution centres [25]. Totally eight algorithms are presented in this approach to address the scalability issues in group key management. Nafi et al. used a matrix based scheme for developing a lightweight key management system [26]. This mechanism is suitable for networks that contain limited resources.

For multi-server architecture, an identity-based authentication protocol is discussed using smart cards [13]. It is a dual server model which imposes varying levels of trust on both the servers, which are the service provider and the control servers. The verifier's information of the user is distributed between these two servers. They asserted that their protocol could withstand various attacks, ensure session key agreement and user anonymity. But the protocol is vulnerable to impersonation attack, stolen smart card attack and leak-of-verifier attack. The authors of [14] have proposed RSA-based two-way IoT authentication techniques using the Trusted Platform Module (TPM). The drawbacks of this mechanism are the significant key size of RSA and the large packet header. The authors of [15] proposed an elliptic curve and symmetric cryptography-based authentication and key management scheme. Additionally, it enables mutual authentication with the network control centre, besides its resistance to denial of service, replay and impersonation attacks. But this mechanism is inefficient in terms of communication and computation. Xue et al. [16] proposed a lightweight authentication and key agreement protocol for multi-server architectures based on dynamic pseudonym identity. But this protocol is vulnerable to Impersonation attack and Session key disclosure attack.

Some of the models that were proposed earlier have some gaps in them and thereby not satisfying the different attack vectors. By considering all of these mechanisms and their associated flow possibilities, in this paper, we have proposed an authentication protocol that is resistant to most of the attacks that were previously discussed in this section.

3 Proposed Mechanism

This section introduces a new authentication protocol that involves three entities: The IoT device, service provider and the trust centre. In this mechanism, various runs occur between IoT device, trust centre and service provider to establish a session key and authenticate each other. Here the trust centre holds the responsibility to authenticate the IoT device and the service provider. Later the trust centre generates a unique key for every session and can only be used by that particular device and the service provider.

As per the Fig. 1 in the first step, the IoT device sends its identity I_i, A_i, which is the hash value generated by the concatenation of password of the device P_i, Nonce N_i and the Nonce of IoT device N_i. All the values are encrypted by the public key of trust centre K_{pt}.

In Step 2, after receiving all the values from the IoT device, the trust centre decrypts them by using its private key. Also, the trust centre calculates M_i, C_i, D_i values. M_i = Hash ($T_i \parallel X$) is the hash value generated by concatenation of T_i and X, where is the secret number given by the trust centre for each IoT device. T_i is a hash value generated by concatenating I_i and N_i, T_i = Hash ($I_i \parallel N_i$). C_i is a hash value generated by the concatenation of T_i and A_i, C_i = Hash ($T_i \parallel A_i$). Finally, D_i = XOR(M_i, C_i) is generated by performing an XOR operation on M_i and C_i, and the trust centre will store this value. The trust centre will calculate the Hash of D_i, which is $D_i\prime$. The trust centre will send $D_i\prime$, registered device acknowledgement message, the nonce of trust centre N_T and nonce of IoT device N_i received in the previous step to IoT device by encrypting them with the public key of IoT device.

In the step 3, the service provider will register itself by sending its identity I_S, along with nonce N_S, and will concatenate these two values to generate S_i which is a hash value of I_S and N_S. It will send I_S, N_S, and S_i encrypted with the public key of the trust centre.

In Step 4, the trust centre will send the nonce of the service provider N_S and nonce of the trust centre N_T by encrypting them with the service provider's public key as described in the Fig. 1.

In Step 5, the IoT device will send a login request to the trust centre by sending its nonce N_i, $D_i\prime$, value, and the nonce of trust centre N_T, received in the previous step to the trust centre by encrypting them with the public key of trust centre.

In Step 6, the trust centre will use $D_i\prime$, values sent by the IoT device in step 5. This value is compared with Hash (D_i), this D_i which is previously calculated and stored by the trust centre. If both values are matched, then the trust centre will generate the

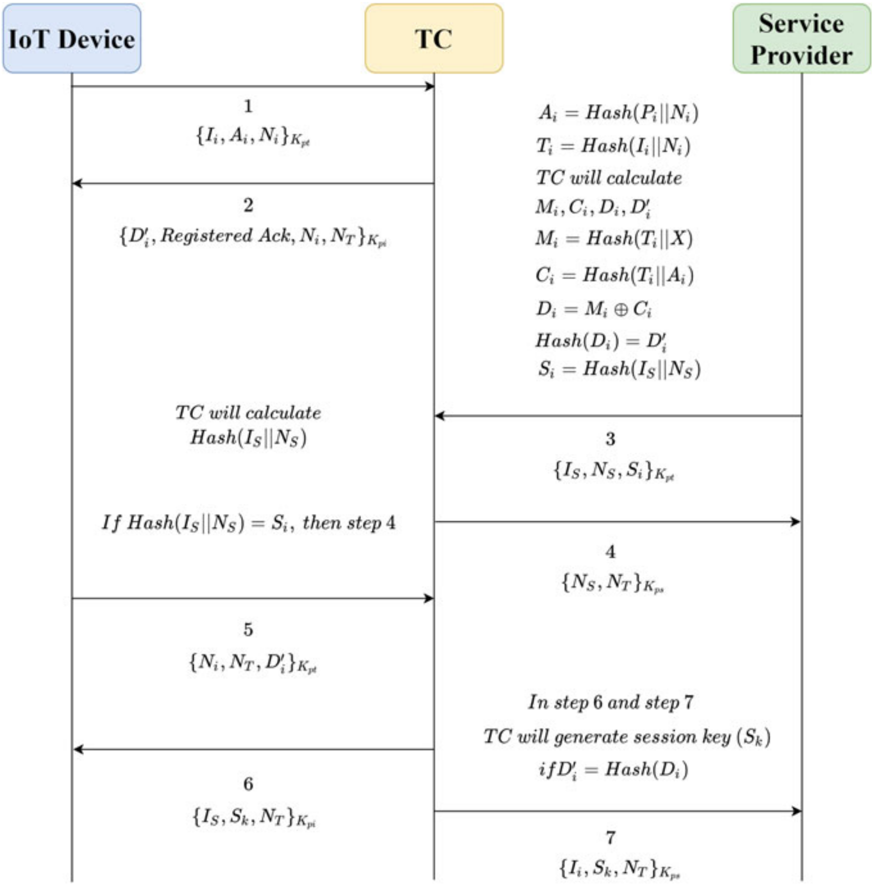


Fig. 1 Proposed protocol (SLAP-IoT)

session key S_k. Later, the session key S_k is sent to IoT device along with the nonce of trust centre N_T and identity of service provider I_S. These values are encrypted with the public key of the IoT device.

In the 7th step, the session key S_k is sent to the service provider along with the identity of IoT device I_i and nonce of trust centre N_T. These values are sent to the service provider by encrypting with the public key of the service provider.

The IoT device and service provider will decrypt the message which the trust centre sends by using their private keys and acquire the session key. This session key is used by the IoT device and service provider for the further transactions that occur in between them. The variables and their definitions used in the protocol are mentioned in Table 1.

Table 1 Variables and their definitions

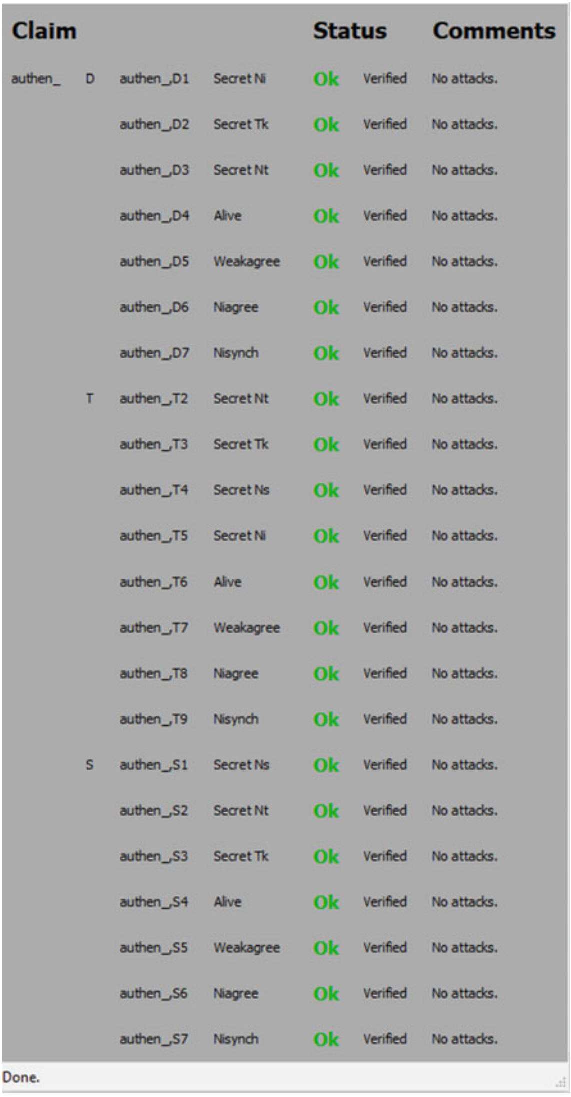
I_i	ID of i^{th} IoT Device
I_S	ID of Service provider
I_T	ID of Trust Centre
N_i	Nonce of IoT Device
N_S	Nonce of Service Provider
N_T	Nonce of Trust Centre
K_{pi}	Public key of IoT Device
K_{pt}	Public Key of Trust Centre
K_{ps}	Public Key of Service Provider
P_i	Password of IoT Device
X	Security number of device given by Trust Centre
A_i	Hash ($P_i \parallel N_i$)
T_i	Hash ($I_i \parallel N_i$)
M_i	Hash ($T_i \parallel X$)
C_i	Hash ($T_i \parallel A_i$)
S_i	Hash ($I_S \parallel N_S$)
D_i	XOR (M_i,C_i)
$D_i{\prime}$	Hash (D_i)
S_k	Session Key

4 Results

In this section, we will analyse the performance of the protocol and discuss the simulation results. For simulating this protocol, we have used a protocol verification tool called Scyther, which was developed in 2007 by Cas Cremersand [17]. The Scyther tool works on the adversary model proposed by Dolev-Yao [23]. It is very fast in terms of analysing the protocols formally and outperformed other state of art formal verification tools. In terms of protocol verification, this is a widely accepted tool. In Scyther [18], the verification of security properties can be done either by specifying the security properties as claims manually. If no claims are mentioned in the protocol, the tool can automatically generate the claims. The extension for protocol definition files is spdl (Security Protocol Description Language). In this tool, we can claim some security properties. After verifying the protocol, if the claims are not satisfied, then in the output console, we can see the status as Fail. Under the pattern section, scyther will generate various patterns describing the possibilities of an attack. If all the claims are satisfied, then the status is shown as OK, and the attack patterns are not generated. To model the intended security properties like Secret, Alive, Weakagree, Niagree, and Nisynch in Scyther, we use a keyword called claim.

Alive is a method of ensuring that an intended entity has completed certain acts [19]. Nisynch indicates that all messages received were sent by the communication

Fig. 2 Scyther output



Claim				Status		Comments
authen_	D	authen_,D1	Secret Ni	Ok	Verified	No attacks.
		authen_,D2	Secret Tk	Ok	Verified	No attacks.
		authen_,D3	Secret Nt	Ok	Verified	No attacks.
		authen_,D4	Alive	Ok	Verified	No attacks.
		authen_,D5	Weakagree	Ok	Verified	No attacks.
		authen_,D6	Niagree	Ok	Verified	No attacks.
		authen_,D7	Nisynch	Ok	Verified	No attacks.
	T	authen_,T2	Secret Nt	Ok	Verified	No attacks.
		authen_,T3	Secret Tk	Ok	Verified	No attacks.
		authen_,T4	Secret Ns	Ok	Verified	No attacks.
		authen_,T5	Secret Ni	Ok	Verified	No attacks.
		authen_,T6	Alive	Ok	Verified	No attacks.
		authen_,T7	Weakagree	Ok	Verified	No attacks.
		authen_,T8	Niagree	Ok	Verified	No attacks.
		authen_,T9	Nisynch	Ok	Verified	No attacks.
	S	authen_,S1	Secret Ns	Ok	Verified	No attacks.
		authen_,S2	Secret Nt	Ok	Verified	No attacks.
		authen_,S3	Secret Tk	Ok	Verified	No attacks.
		authen_,S4	Alive	Ok	Verified	No attacks.
		authen_,S5	Weakagree	Ok	Verified	No attacks.
		authen_,S6	Niagree	Ok	Verified	No attacks.
		authen_,S7	Nisynch	Ok	Verified	No attacks.

Done.

partner and received by another communication partner. We have implemented our proposed mechanism in the Scyther tool. As per the screenshot Fig. 2, we can see that this mechanism satisfies all the properties, and there is no scope for attacks. Here, we discuss some of the security features of the proposed protocol:

1. Resistance to Impersonation attack: An impersonation attack is not possible even if an attacker tampers the details of the IoT device because in the first step, identity I_i, A_i and nonce N_i are sent to the trust centre. Based on these values, the trust centre will calculate the values of M_i, C_i, and D_i. If the attacker tries to create an

Table 2 Proposed model results comparision with previous literature

Attack type	Yang et al. [19]	Sood et al. [12]	He et al. [20]	Xue et al. [15]	Li et al. [21]	Proposed
Impersonation attack	✓	✓	✓	✗	✗	✓
Replay attack	✓	✗	✗	✓	✗	✓
Session key disclosure attack	✗	✓	✓	✗	✗	✓

identical message as in step 1 and tries to send it to the trust centre, there will be a change in A_i, T_i, M_i, C_i, and D_i values. It will cause a mismatch so the attacker cannot impersonate a legitimate device. Also, in every step, nonce is being sent from one entity to other by encrypting them with public keys. In the next step, the concerned entity will echo the received nonce.

2. Resistance to Replay attack: In this attack, the attacker will forward the messages captured in the previous step. After validating the credentials, the IoT device will perform the login process. If the attacker tries to perform the replay attack after the IoT device is logged in, it will be of no use. Also, in this protocol, we are using nonce, which is a random number. As a result, the attacker cannot use the previously captured messages and pretend as a legitimate device because the nonce value will be updated at each step.
3. Resistance to Session key disclosure attack: Our proposed mechanism will generate the session key after validating the $D_i\prime$, value in the 6th Step. If the $D_i\prime$, value does not match, then the session key is not generated. If $D_i\prime$, value matches, then it is sent to the IoT device and the service provider by the trust centre after generating the session key. Before being sent, the session key is encrypted with the public keys of the IoT device and the service provider, respectively. The difficulty of the hash function and secret random nonces generated by the IoT device, trust centre and service provider, respectively, ensure the security of the session key in our protocol. So, session key disclosure attack is not possible in our proposed mechanism.

We conducted a comparison of our protocol's performance with other relevant publications such as Yang et al. [19], Sood et al. [12], He et al. [20], Xue et al. [15] and Li et al. [21] in terms of attack resistance and communication cost. From Table 2, it is clearly evident that our proposed protocol is better in terms of attack resistance when compared to related existing schemes where (✓) indicates that protocol resist the attack and (✗) does the opposite (Fig. 3).

We have calculated the login cost and authentication cost of our proposed protocol. In comparison with the related protocols, our proposed mechanism has more or less a similar communication cost. This is because, in our mechanism, each hash operation will consume 224 bits, considering the SHA-3 hashing algorithm, while the hash operation in other compared protocols utilizes 128 bits, and compromised hashing algorithms were used.

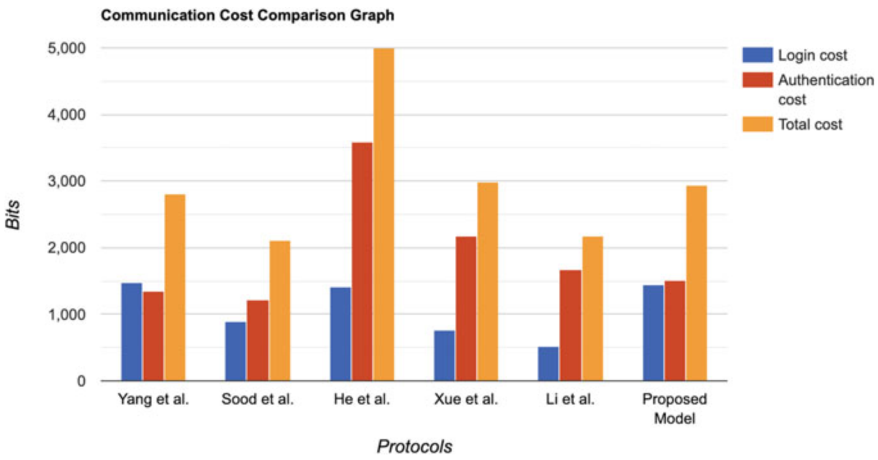


Fig. 3 Communication cost comparison graph

5 Conclusions and Future Work

In our work, we have implemented a protocol for authentication mechanism in IoT environment. The IoT environment contains devices, sensors that have limited resources in terms of memory and computation power, because of which they have a fragile security mechanism. In our approach, the trust centre plays a crucial role in authenticating, validating the IoT device, and establishing the session key between the IoT device and the service provider. The proposed protocol resists attacks like the session key disclosure, replay attack and impersonation attack. We have used a robust formal verification tool named "Scyther" to support our claim. In comparison with the related protocols, our proposed mechanism has more or less a similar communication cost satisfying the light weight property. Currently, the proposed protocol can be used for authentication when there are two parties involved. But, when it comes to group key agreement, our proposed approach is not applicable. So, as possible future work, we would like to extend this mechanism so that it can also be used in group key agreement protocols.

References

1. Statista: Internet of Things (IoT) and non-IoT active device connections worldwide from 2010 to 2025, https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/ Accessed on April 7, 2021
2. H. Chunduri, T. Gireesh Kumar, P.V.S. Charan, A Multi Class Classification for Detection of IoT Botnet Malware, in *Computing Science, Communication and Security. COMS2 2021. Communications in Computer and Information Science*, ed by N. Chaubey, S. Parikh, K. Amin, vol 1416 (Springer, Cham, 2021)

3. M. Shriny, A. Ajisha, C. Srinivasan, Design and implementation of the protocol for secure software-based remote attestation in IoT devices, in *International Conference on Soft Computing and Signal Processing* (Springer, Singapore, 2019)
4. S.K.B. Hemanth, K.V. Lakshmy, Enhanced attach procedure for prevention of authentication synchronisation failure attack, in *Soft Computing and Signal Processing. ICSCSP 2019. Advances in Intelligent Systems and Computing*, ed by V. Reddy, V. Prasad, J. Wang, K. Reddy, vol 1118 (Springer, Singapore, 2020)
5. M. El-Hajj et al., A survey of internet of things (IoT) authentication schemes. Sensors **19**(5), 1141 (2019)
6. E.D.O. Silva, et al. Authentication and the internet of things: a survey based on a systematic mapping, in *International Conference on Software Engineering Advances* (2017)
7. H. Khemissa, D. Tandjaoui, A lightweight authentication scheme for E-health applications in the context of internet of things, in *2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies* (IEEE, 2015)
8. H. Khemissa, D. Tandjaoui, A novel lightweight authentication scheme for heterogeneous wireless sensor networks in the context of internet of things, in *2016 Wireless Telecommunications Symposium (WTS)* (IEEE, 2016) https://ieeexplore.ieee.org/abstract/document/9242592
9. M. Shahzad, M.P. Singh, Continuous authentication and authorization for the internet of things. IEEE Internet Comput. **21**(2), 86–90 (2017)
10. A.P. Haripriya, K. Kulothungan, ECC based self-certified key management scheme for mutual authentication in Internet of Things, in *2016 International Conference on Emerging Technological Trends (ICETT)* (IEEE, 2016)
11. S. Kumari et al., A secure authentication scheme based on elliptic curve cryptography for IoT and cloud servers. J. Supercomput. **74**(12), 6428–6453 (2018)
12. A. Shahidinejad et al., *Light-Edge: A Lightweight Authentication Protocol for IoT Devices in an Edge-Cloud Environment* (IEEE Consumer Electron, Magaz, 2021)
13. S.K. Sood, Dynamic identity based authentication protocol for two-server architecture. J. Inf. Secur. **3**(04), 326 (2012)
14. T. Kothmayr, et al., A DTLS based end-to-end security architecture for the Internet of Things with two-way authentication, **37th Annual IEEE Conference on Local Computer Networks-Workshops** (IEEE, 2012)
15. M. Qi, J. Chen, Y. Chen, A secure authentication with key agreement scheme using ECC for satellite communication systems. Int. J. Satell. Commun. Netw. **37**(3), 234–244 (2019)
16. K. Xue, P. Hong, C. Ma, A lightweight dynamic pseudonym identity based authentication and key agreement protocol without verification tables for multi-server architecture. J. Comput. Syst. Sci. **80**(1), 195–206 (2014)
17. C.J.F. Cremers, The scyther tool: verification, falsification, and analysis of security protocols, in *International conference on computer aided verification* (Springer, Berlin, Heidelberg, 2008)
18. C. Cremers, The scyther tool, www.cs.ox.ac.uk/people/cas.cremers/scyther/ [Online; Accessed on March 10, 2021]
19. G. Lowe, A hierarchy of authentication specifications, in *Proceedings 10th Computer Security Foundations Workshop* (IEEE, 1997)
20. D. Yang, B. Yang, A biometric password-based multi-server authentication scheme with smartcard, in *International Conference On Computer Design and Appliations*, vol. 5 (ICCDA, 2010), pp. 554–559
21. D. He, S. Wu, Security flaws in a smartcard based authentication scheme for multi-server environment. Wirel. Pers. Commun. **70**(1), 323–329 (2013)
22. X. Li, Y.P. Xiong, J. Ma, W.D. Wang, An efficient and security dynamic identity based authentication protocol for multi-server architecture using smartcards. J. Netw. Comput. Appl. **35**(2), 763–769 (2012)
23. D. Dolev, A. Yao, On the security of public key protocols. IEEE Trans. Inf. Theo. **29**(2), 198–208 (1983)
24. A. Bashar, Sensor cloud based architecture with efficient data computation and security implantation for Internet of Things application. J. ISMAC **2**(02), 96–105 (2020)

25. M. Dammak, et al. Decentralized lightweight group key management for dynamic access control in IoT environments. IEEE Trans. Netw. Serv. Manage. **17**(3 (2020): 1742-1757
26. M. Nafi, S. Bouzefrane, M. Omar, Matrix-based key management scheme for IoT networks. Ad Hoc Netw. **97**, 102003 (2020)

Medical Service Chatbot for Basic Medication

P. Sai Kiran Reddy, Y. Risheet, M. Naga Sivaram, Md. Roshan Tanveer, G. Chandra Sekhar, and G. Krishna Kishore

Abstract The main motto of the health care chatbot is to avoid unnecessary point of contact with the doctors, surgeons, and medical advisors. The chatbots can connect people with text or through voice interface which will make it as an application of artificial conversation and machine learning. As the chatbots are fully operational and an easy tool to communicate, it is fully accessible at all times as there is no requirement of any human for operating. For a basic medication arriving at the hospitals after traveling, a long distance from source to destination is not suggested. Therefore, as a replacement chatbots can fill that requirement in an effective manner in dealing with the complex problems. The whole application is developed using the python web framework Django and the database used is postgreSql. The underlying principle for predicting the disease is decision tree.

Keywords Medical chatbots · Health care · Decision tree classifies Django (Python web based framework) · Confidence · Entropy · Information Gain · postgreSQL

1 Introduction

Chatbots are automated systems which imitate the conversations between two individuals. They provide a simulating platform for effective and smart communications with the user in extracting the information. They are used and available in many domains some of them are as follows: business, marketing, stock markets, customer care, healthcare, counseling, recommendation systems, entertainment, journalism, and shopping.

The aim of this paper is to address the need and usage of chatbots in the healthcare domain. The problem with these chatbots is that they just provide monotonous

P. Sai Kiran Reddy (✉) · Y. Risheet · M. Naga Sivaram · Md. Roshan Tanveer ·
G. Chandra Sekhar · G. Krishna Kishore
Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

G. Krishna Kishore
e-mail: gkk@vrsiddhartha.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_62

answers to user's questions. The paper aims at proposing a chatbot system, which is capable of establishing a smart communication. This can be achieved by implementing a smart and a responsive chatbot which is able to communicate with the user just as another human can communicate. This technique makes the system more communicative in the natural language, proves fruitful for counseling, and can also be modeled for prediction of diseases.

The project recognizes the need of chatbot in healthcare domain and address the issues like distance to be traveled and time to be waited for getting an appointment which is not nominal. Therefore, the project addresses all these issues. The flow of the model begins with registration for users, doctors, and admin.

The users should register with the chatbot before they want to access the chatbot. By filling the details, the account for a particular user is created and is simply stored in the postgreSQL database. Then, after logging in the user will be able to access the chatbot. After going through a set of symptoms the user will has to select a symptom from the set which is suffering with and then has to choose it by clicking. If there are multiple symptoms, there is a provision available to select multiple symptoms from the given set. After application of the Decision tree algorithm to the queries obtained from the user the best possible disease from the prognosis column based on the parameters like confidence, entropy and information the disease will be shown. If there is possibility in some rare cases multiple diseases will be shown by the bot. After the prediction, the user is provided with some basic supplements, and some more information about the disease will be provided. The chatbot connects the victims with the doctors. The doctors are supposed to register in this chatbot by filling out some details. Then, the chatbot prescribes the doctors to the user after prediction so that a chat window is created for victim and doctor where they can chat and share the information about the problem faced by the victim so that necessary actions are prescribed by the doctor to the victim in the chat window.

Therefore, the chatbots are fully operational at all times. Because of the growth in population and also the medical charge for basic medication is too high it will be a hectic task for future generations to consult the doctors and book an appointment. All this work can simply be replaced by a chatbot. After answering certain number of questions, we will be able to get solution to the problem, we are facing within minutes.

Chatbots are getting prominent in every field they are introduced. As chatbots are easy to operate people may find them as their first option to consult. Moreover, chatbots can also be used to book or schedule appointments with doctors easily without any consultation fee. There is no need of waiting in the queue for long hours, just for minutes of solution. Chatbots will also prescribe the preventive measures in order to tackle the problems easily in an effective style and also help in scheduling the appointment with the prominent doctors in the field without the involvement of consultancy efficiently.

2 Literature Survey

Dinesh Kalla, Fnu Samaah proposed a thematic analysis on the qualitative data to identify the common trends and patterns. The query has been processed by dividing and breaking up strings into tokens, which are small words or units that can be used further [1]. Flora Amato, Stefano Marrone, Vincenzo Moscato, Gabriele Piantadosi, Antonio Picariello, and Carlo Sansone proposed the computational cluster providing storage, computing and machine learning skills to deduce the given query into token and to identify the accurate result for the given number of symptoms [2]. Lekha Athota, Vinod Kumar Shukla, Nitin Pandey, Ajay Rana proposed the way in which the client inputs the question in the UI as the text. The words are then tokenized using Tokenization. The stop words are removed to extract important keyword. Feature based on *N*-gram TF-IDF algorithm [3].

Nivedita Bhirud, Subhash Tataale, Sayali Randive, Shubham Nahar induced a system to implement word segmentation (tokenization). Then, implement POS tagging the process includes creation of a dependency parser. Detection of disease and extraction of information. To implement Decision or ML Engine [4]. Mrs. Rashmi Dharwadkar, Dr. Mrs. Neeta A. Deshpande used a google app for maintaining a basic and formal conversation. Complaints are identified using NLP. Porter stemming algorithm is used for calculating the threshold and word similarity. Support vector machine is used for classification [5].

J. Jinu Sophia, D. Arun Kumar, M. Arutselvan, S. Barath Ram constructed a JSON file for creating intent and training data and converting into bag of words. Usage of NLTK and Tflearn for model building. Lancaster stemming is used for stemming the words [6].

The above given literature surveys have given a module based on the predefined rules. They are developed using the NLP Library or some other API's or some other text mining techniques. The main motto of the chatbots is to provide some basic medication and also to establish some communication with the doctors in the locality. But the above discussed chatbots does not belong to this sort. Therefore, not only the prediction of disease but also finding and cure should be the end result.

This is obtained through the model developed using Django framework. The chatbot we are bringing is it not only predicts the disease but also provide basic medication. Establishing a fair communication with doctors will be done through the developed model Django framework and information is stored in the postgreSQL.

3 System Description

The aim of the project is to make the people's life a lot easier by developing a website that helps patients to identify their disease provided with proper symptoms using the machine learning and also provides a platform to the user to contact with the specialist doctors and have a quick chat with them and have an appointment when required.

The key part of this is to identify the disease of the user based on the sequence of symptoms that are entered by the user. The model was developed using the dataset available on the internet using the decision tree algorithm. The front-end web pages are used to extract the input symptoms from the user and predict the disease using the trained model and stores back them to the back-end database postgreSQL using the patient id. The user has to sign up in the application so that a separate record is created for every user and will be assigning an id that is unique for every user that helps in storing the complete information of the user. Then, the user has to sign into the application to access the features of the designed model.

The complete information about the user will be seen in the dashboard. Then, user will be provided with the set of all the symptoms, and the user has to choose the subset of symptoms from that. Once all the symptoms are noted the model will try to predict the disease and also provide you with the extra information related to the disease that is available in the internet. Then, user will move on to the consultation part.

The mentioned steps are needed to be followed in order to access the functionalities:

A. **Sign up**
The user has to sign up on the website by providing the basic information about their name, age, mobile number.

B. **Sign in**
The user has to sign into the application in order to access the features. The user has to sign in with the credentials given during sign up. A separate record was created for every user in the database.

C. **Disease prediction**
The model was built with a predefined set of symptoms that are available. The same sets of symptoms are listed on the user interface so that user has an option to choose the symptoms they were facing.
After having a note of all the symptoms, they are given as input to the trained model. The model predicts the disease and displays back to the user with the accuracy scores.

D. **Consultation**
Based on the output disease predicted they are provided with the list of available specialist doctors. Then, the user will be taken to user interface where they can chat with doctor and can provide you with the basic medication. If necessary, user will be provided with the appointment where he can visit the hospital and have basic tests.

E. **Feedback**

After the consultation with a specialist doctor, user has to fill the form containing the review and rating to the doctor based on the treatment and the way he responded back.

The below given figure explains about how the internal process of prediction and offering the services to the victims (Fig. 1).

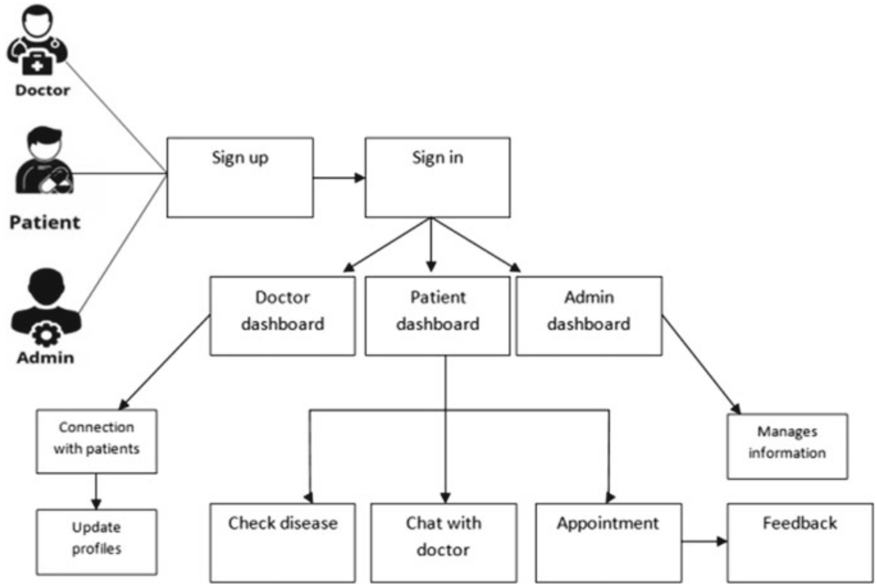


Fig. 1 Block diagram

4 Methodology

The proposed model for the chatbot is the eager decision tree algorithm for prediction. The decision tree algorithm is best suitable for this type of classification problems. Inputs given are symptoms with prognosis column which will be efficient to solve the phrases easily given by the user. This forms the dataset. Information gain of the disease helps in even more division of symptoms to get the required result. The decision tree algorithm is a learning algorithm for classification which identifies a finest classifying point and classifies based on it. Information gain and entropy are used as measures to bring out the best attribute, i.e., symptoms to predict the disease here might be cases where data which cannot be distinguished exists. In those cases, the input is mapped to high-dimensional attribute space such that they can be separated by a hyper plane. This projection is carried out well by kernels.

A. **Data Preprocessing**:

The dataset is obtained from Kaggle. It specifies the disease with number of symptoms which are related to the disease, which are useful in providing a better accuracy in prediction.

Given below is the brief overview of the dataset.

The stored information contains the symptom related to particular disease in the prognosis column on the basis of which we can predict the disease also some doctor related information.

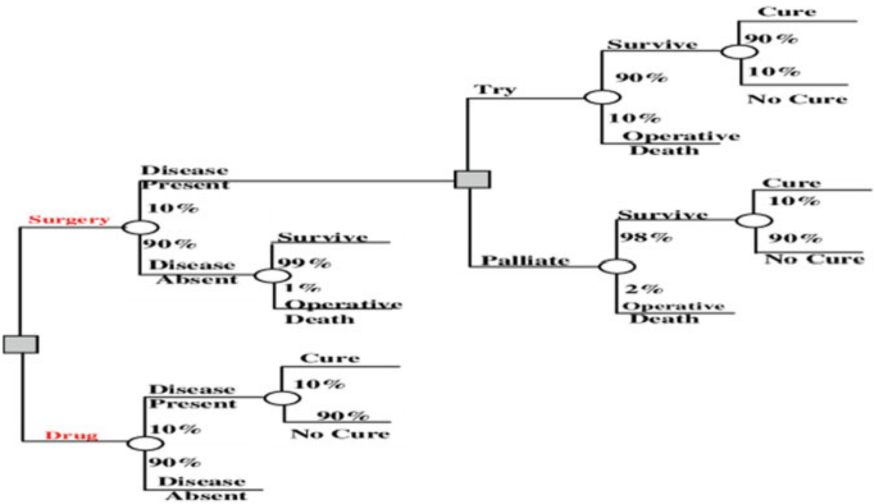


Fig. 2 Figure showing the classification structure of the algorithm through which the rules are defined

Based on the conversation form the user and the chatbot, this will specify the precautions for the disease predicted. Given below is the brief overview about the precautions for the given disease.

These are the phases of executing the algorithm in the application:-

- Consider the training dataset containing the diseases
- Preprocess the dataset for the model development
- Split the data into training and testing datasets
- Build the model and train the model against training data (Fig. 2).

B. **Building the model**:

The decision tree is constructed based on the class labels which are used for prediction. The parameters for the construction of tree are based on each symptom which is obtained from the user. Based on the high information gain, the decision tree will be constructed.

After the construction of decision tree, the model is trained on the training dataset available on the internet. After, the model is tested on the testing dataset the model provides a good acceptable accuracy score. Then, the model is stored in file and then deployed with the application. Then, the model will be ready to be used in real life by answering the disease back to user based on the symptoms.

1. If user says a yes, the system traces back its path to the root node to check for the disease.
2. This keeps on iterating for all users and the tree keeps getting updated for new entries for which it traces the path available. Through which the values obtained from the testing and training phase are evaluated.

C. **Integrating the model with the web**:

The web pages are built using the front-end technologies like html, cuss, and javascript. These web pages help in signing up of the user with the application and helps in taking the input symptoms from the user and store them in back-end database and then displaying the results back to the user.

These pages provide an interactive environment for the user. The back-end technology used is the Django framework in python. This framework uses the MVT architecture for building the web application. The model view template architecture which is similar to model view controller architecture. It provides the separate files for every module.

The model contains the actual logic to connect with the back-end database postgreSQL. The view file contains the actual business logic of the project. The template section contains the front-end web pages that the user actually interacts with. The Django is an open-source framework available in python, and it is more concerned in automating the working areas.

The postgreSQl database is used to store the data in the backend. It is an open source and object-relational database. In these advanced relational databases, the data of each user is stored as a separate record. The pgAdmin is used for visualizing and managing the data in tables by providing the user interface to interact with admin. Separate tables are created for users and doctor. The users are provided with an id which is unique within the application. So, the complete information about the user will be stored under this id.

Concepts used in Decision tree, while constructing the decision tree to frame the rules are entropy and information gain.

Formula for evaluating entropy:

$$\text{Entropy } p(i) = -p(a) \times \log(p(a)) - p(b) \times \log(p(b))$$

Formula for evaluating information gain:

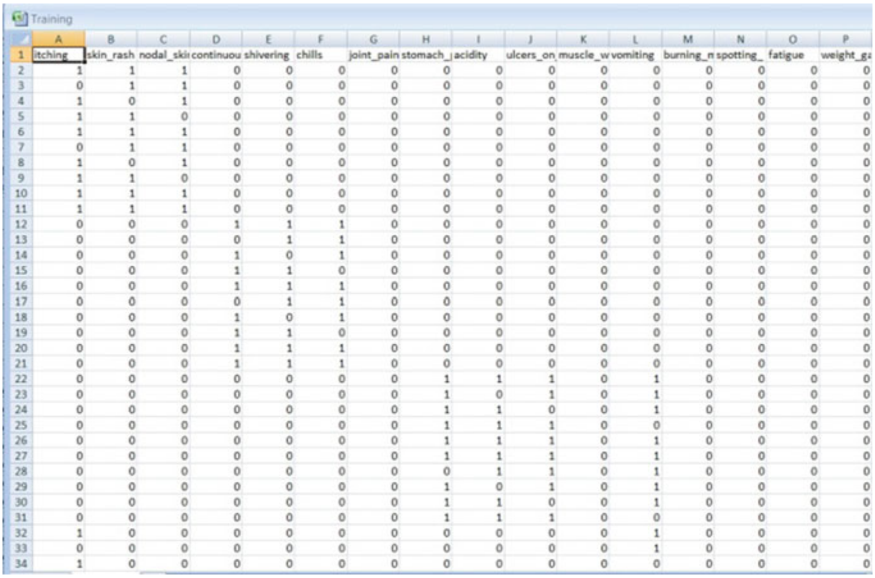
$$I = -\Sigma\, p_i \log(p_i)$$

5 Results

Chatbot will be asking the user about the symptoms and also the duration of the symptom he has been suffering with and predicts the disease. Based on the symptoms given, the chatbot will be providing the basic information about the disease and also the appointment links of the hospitals from the nearby location of the user.

By training the dataset of different count, we get the class wise distribution of the record. Below figure shows the training record in percentage. The confidence of each disease will also be shown on the given user's prediction (Fig. 3).

The above given dataset is obtained from Kaggle website where it has 140 symptoms and 40 diseases. As the first step of execution the dataset will be preprocessed,



Training

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	itching	skin_rash	nodal_ski	continuou	shivering	chills	joint_pain	stomach_	acidity	ulcers_on	muscle_w	vomiting	burning_n	spotting_	fatigue	weight_g
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
14	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
15	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
18	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
19	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
21	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0
23	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0
24	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
25	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
26	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0
27	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0
28	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0
29	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0
30	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
31	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
32	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
34	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 3 Dataset containing symptoms in row1 and disease in prognosis column

and the information will be classified using the data frames. The Decision tree Algorithm is executed on the above dataset to obtain the accurate result or disease suffered by the user (Fig. 4).

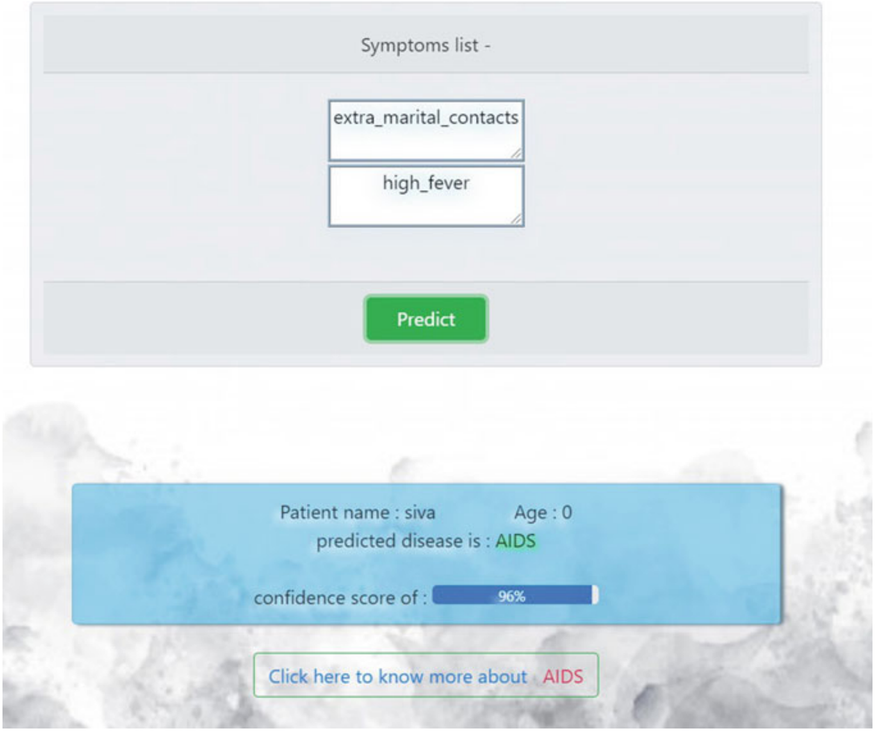
The above-mentioned symptoms are extra martial contacts and high fever which are the main reasons for the AIDS; therefore, the confidence score of the disease was 96% (Fig. 5).

The above predicted message is heart attack as breathlessness and chest pain are not only the main symptoms of the heart attack; therefore, the confidence is only 69% as there is a minute chance for some other disease (Fig. 6).

As we can observe the confidence score of the disease is entirely based on the number of combinations of symptoms chosen by the user. Therefore, from the symptoms entered the best fit is chicken pox which is displayed (Fig. 7).

6 Conclusion

Therefore, the application has given exact results for the disease faced by the victims and also established a communication medium like chat window with the doctor. The graphical user interface is user-friendly and attractive enough so as to make the users visualize the output. The application has also displayed the confidence of each after the prediction. The results obtained here shows that the problem faced by the user can be eradicated by basic medication. The effective means of communication



Symptoms list -

extra_marital_contacts

high_fever

Patient name : siva Age : 0

predicted disease is : AIDS

confidence score of : 96%

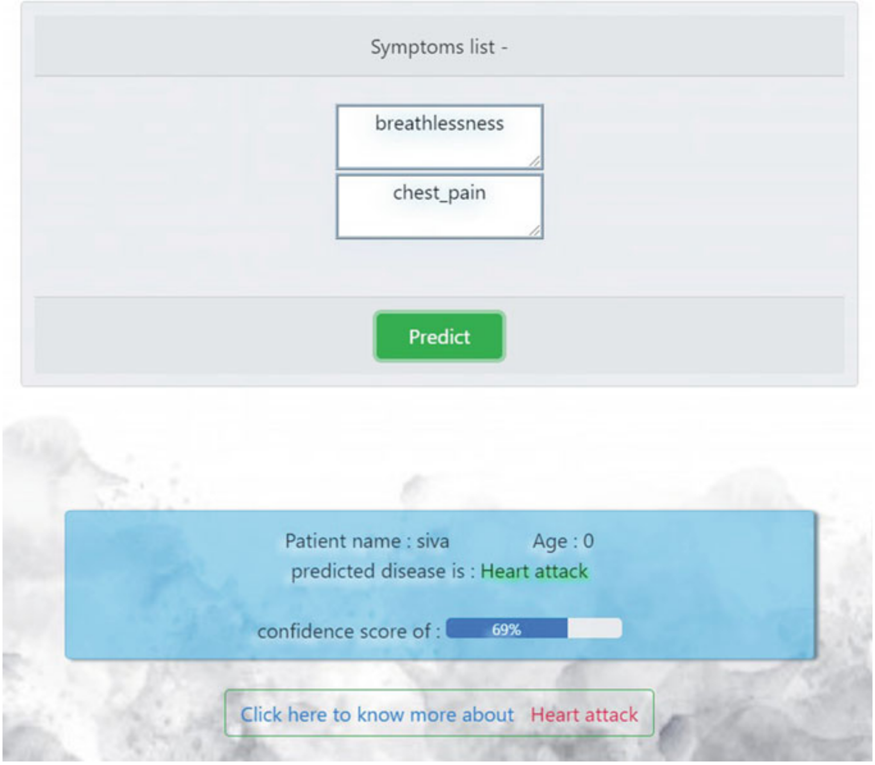
Fig. 4 Disease predicted is AIDS from the given symptoms

is also established with the user by the doctor so as to improve the interaction without booking an appointment.

7 Future Scope

Chatbots are being in use only in particular domains and in future we will see them a lot more in action. There are some ample amounts of innovations yet to be uncovered in healthcare domain using these bots. These bots will be available for everyone at every time and helps in curing the small-scale diseases and if needed they helps to consult to the specialists.

This application helps in having all the data related to the user and there is not much paperwork involved. These data will be useful for the doctors in emergency conditions. There is a lot of scope to add the location-based services where user can find the best hospitals near him. There is a possibility of having video conferencing services for having virtual meetings between with the patient and the doctor. The



Symptoms list -

breathlessness

chest_pain

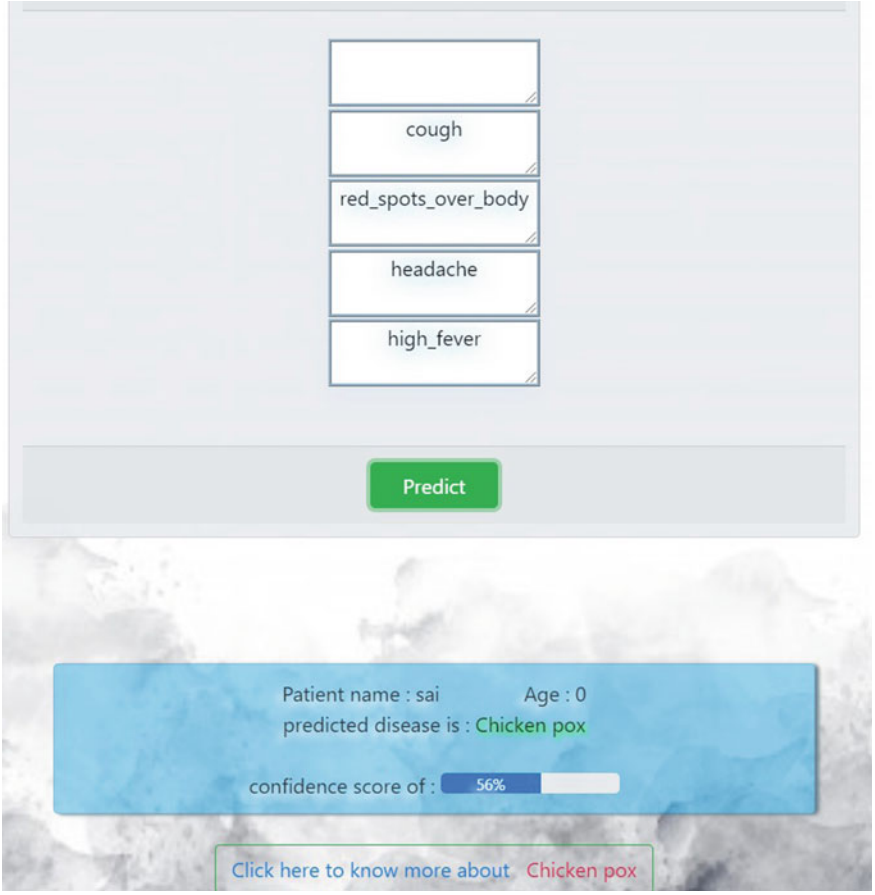
Patient name : siva Age : 0

predicted disease is : Heart attack

confidence score of : 69%

Fig. 5 Disease predicted is heart attack from the symptoms obtained by the user

chatbots will act as a first option for the people of the upcoming decades for basic medication services.



cough

red_spots_over_body

headache

high_fever

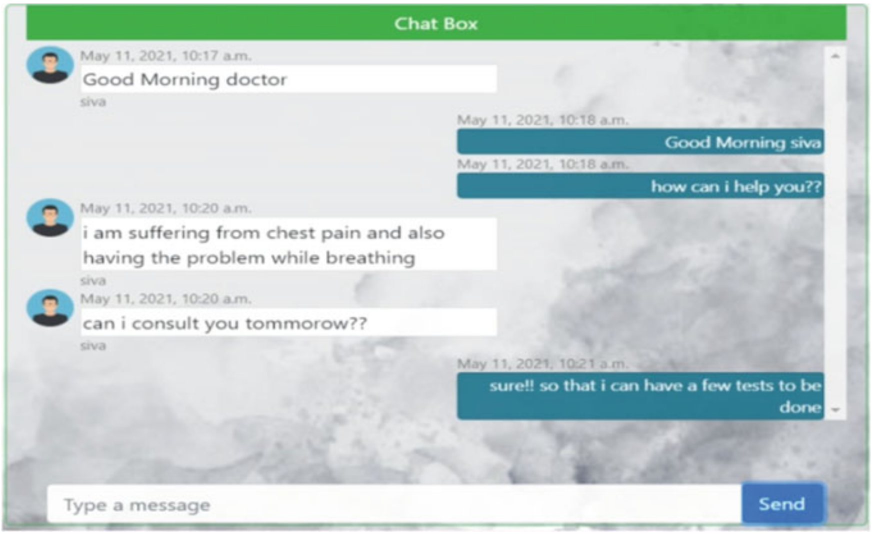
Patient name : sai Age : 0

predicted disease is : Chicken pox

confidence score of : 56%

Click here to know more about Chicken pox

Fig. 6 Disease predicted is chicken pox from the symptoms obtained from the user



May 11, 2021, 10:17 a.m.
Good Morning doctor
siva

May 11, 2021, 10:18 a.m.
Good Morning siva

May 11, 2021, 10:18 a.m.
how can i help you??

May 11, 2021, 10:20 a.m.
i am suffering from chest pain and also having the problem while breathing
siva

May 11, 2021, 10:20 a.m.
can i consult you tommorow??
siva

May 11, 2021, 10:21 a.m.
sure!! so that i can have a few tests to be done

Fig. 7 Chatting with the respective doctor prescribed by the bot

References

1. D. Kalla, F. Samaah, Chatbot for medical treatment using NLTK Lib. Int. Organ. Sci. Res. (IOSR) (2018)
2. F. Amato, S. Marrone, V. Moscato, G. Piantadosi, A. Picariello, C. Sansone, Chatbots meet eHealth: automatizing healthcare. DIETI University of Naples Federico II, via Claudio (2015)
3. L. Athota, V.K. Shukla, N. Pandey, A. Rana, Chatbot for healthcare system using artificial intelligence, in *Infocom Technologies and Optimization (Trends and Future Directions) ITO (IEEE)*, June 4–5, 2020
4. N. Bhirud, S. Tataale, S. Randive, S. A literature review on chatbots in healthcare domain. Int. J. Sci. Technol. (July 2019)
5. R. Dharwadkar, N.A. Deshpande, A medical chatbot. Int. J. Comput. Sci. Technol. (2018)
6. J. Jinu Sophia, D. Arun Kumar, M. Arutselvan, S. Barath Ram, A survey on chatbot implementation in health care using NLTK. Int. J. Comput. Sci. Mob. Comput. (3rd March 2020)

FPGA Implementation and Comparative Analysis of DES and AES Encryption Algorithms Using Verilog File I/O Operations

Akshada Muneshwar, B. K. N. Srinivasarao, and Ajay Chunduri

Abstract Data communicated electronically is prone to attacks. Due to communication systems and wireless communications technology advancements, providing data security plays a vital role. Expanded interest for information security is an evident reality. The need to protect the critical network infrastructure and information systems has become important. Towards accomplishing higher security, cryptographic algorithms assume a significant part in the safeguarding of information from unapproved utilization. Due to the ability to provide faster and more customizable solutions, the field-programmable gate arrays (FPGAs) are suitable option to implement the cryptographic algorithms. In this paper, we have implemented the data encryption standard (DES) and advanced encryption standard (AES) encryption algorithms. Here, the input plain text data and the key input are read from a text file, and the encrypted output data is captured in a text file. The mentioned design is implemented in Verilog language using Libero SoC Design Suite, targeting the PolarFire FPGA family and the MPF500T-1FCG1152I device.

Keywords Cryptography · Encryption · DES · AES · FPGA · Verilog · Key · Round · File.

1 Introduction

With the rapid development of data communication and Internet technology, the secure exchange of information over these communication channels plays a crucial role. The widespread use of computer and communication systems by various organizations and industries has increased the risk of theft of data. In the case of applications such as bank transactions, electronic mail, audio/video conferencing,

A. Muneshwar (✉) · B. K. N. Srinivasarao
National Institute of Technology, Warangal, Telangana, India
e-mail: srinu.bkn@nitw.ac.in

A. Chunduri
Microchip Technology Pvt. Ltd., Hyderabad, India
e-mail: ajay.chunduri@microchip.com

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_63

intellectual property and customer data, secure communication is essential. To prevent an unauthorized person from accessing sensitive information over the unsecured medium, a mechanism to provide data security is needed. Thus, to accomplish the task of information security, cryptography is essential [1].

Cryptography is the practice and study of hiding information. It allows us to store sensitive information or transmit it across the insecure channel so that unauthorized persons cannot read it [2]. It is a science of converting plain understandable data into obscure data and again transforming that message into its original understandable form. The initial input message is called the plain text, while the encrypted message is known as the ciphertext. The process of converting from plain text to ciphertext is known as encryption. The process of converting from ciphertext to plain text is known as decryption. A key is a variable applied in cryptography to calculate a block of plain input text to deliver encoded text or retrieve back unencrypted data from the encrypted information. Cryptographic algorithms are classified in two types depending on the key used: symmetric and asymmetric algorithms. Symmetric systems use an identical key to both encrypt the message text and decrypt the ciphertext (encrypted plain text). Data encryption standard (DES), 3DES and AES are symmetric cryptosystems. Asymmetric systems use different keys for encryption and decryption [2]. Rivest–Shamir–Adelman and elliptic curve are asymmetric cryptosystems. Symmetric cryptosystems are more suitable to encrypt an enormous amount of data at high speed.

Based on the design by Horst Feistel, the data encryption standard (DES) algorithm was developed at IBM around the 1970s and was later adopted by the National Institute of Standards and Technology (NIST). However, DES was found to be falling short in handling the new challenges in encryption with advancement in technology with time and a new encryption algorithm—advanced encryption standard (AES) was announced [3]. On 2 January 1997, NIST asked for proposal for encryption standard to follow DES and advanced encryption standard (AES) was approved based on the level of security offered and feasibility in various environments. Existing systems are explained in Sect. 2. The overview of the data encryption standard (DES) is discussed in Sect. 3, and advanced encryption standard (AES) is discussed in Sect. 4. The Verilog file I/O operations are described in Sect. 5. Section 6 explains the implementation process. Results are discussed in Sect. 7, and Sect. 8 concludes the paper.

2 Literature Review

The analysis is done between the standard DES algorithm comprising the 16-round pipeline and DES algorithm with 8-round pipeline [4]. The DES operation was implemented on Xilinx XC3ES500E field-programmable gate array (FPGA) using VHSIC Hardware Description Language (VHDL). It compared the resource utilization in both the implementation types, and the average resource utility is found to be 9.7% for DES 8-round algorithm and 21.2% for DES 16-round implementation.

The AES encryption algorithm is implemented on Xilinx Spartan-3 XC3S400 FPGA using the VHDL language [5]. 128-bit plain text is used as input, and key size used is 128 bits in size. Xilinx EDA tool is used to execute the VHDL code for the algorithm.

The AES encryption algorithm is implemented on Xilinx Spartan-3 XC3S400-4PQ208 device FPGA using the VHDL language [6]. 128-bit plain text is used as input, and the key size used is 128 bits. The simulation, synthesis, design resource utilization and timing results are discussed.

A compact FPGA architecture for advanced encryption standard (AES) algorithm was evaluated [7]. Lesser area consumption and good performance were achieved using folded architecture. Xilinx Spartan II XC2S30 FPGA was used for implementation. The encryption speed and the cost involved were found to be useful for wireless communication and low-cost embedded applications.

Advanced encryption standard (AES) algorithm, which aims at low complexity architecture, is implemented targeting the FPGA [8]. It attains low latency and high throughput. Various performance observations and simulation results are compared with similar existing systems. The design was implemented using VHDL language. Xilinx ISE and ModelSim software were used for synthesis and simulation process, respectively.

Comparative study of advanced encryption standard with the key length of 128 bits and 256 bits is carried out [9]. The time required during encryption and decryption process is discussed. It is noted that AES decryption operation with 256-bit key is robust as it takes more time to decrypt the data.

3 Overview of DES Encryption

3.1 Description of DES Encryption Operation

In DES encryption operation, the input data and the key are 64 bits in size. During various steps of encryption, it uses 48-bit key. DES encryption process utilizes operations such as permutation, XOR and substitution [10]. It includes multiple rounds of processing. Figure 1 shows the steps involved in DES encryption are: initial permutation (IP), 16 rounds of complex key-dependent calculation and the final permutation, which is the inverse of initial permutation. The 64 bits of input data are permuted in the initial permutation (IP) step. These 64 bits are then divided into two parts: left and right blocks, each of length 32 bits. Now, for the 16 iterations of processing, function f is performed using the 32-bit right block of input data and the key as shown in Fig. 2. The output obtained from the f function is then XORed with the left block of input data. This XORed result which is 32 bits in size is then swapped with the original right block of input data [10]. The sixteen keys are obtained using a key generation process where each key of size 48 bits is obtained using the 64-bit key that was provided at the start of encryption. After the 16 rounds of processing, the left and right blocks are swapped. Then, the inverse permutation operation is carried out, which outputs the encrypted data.

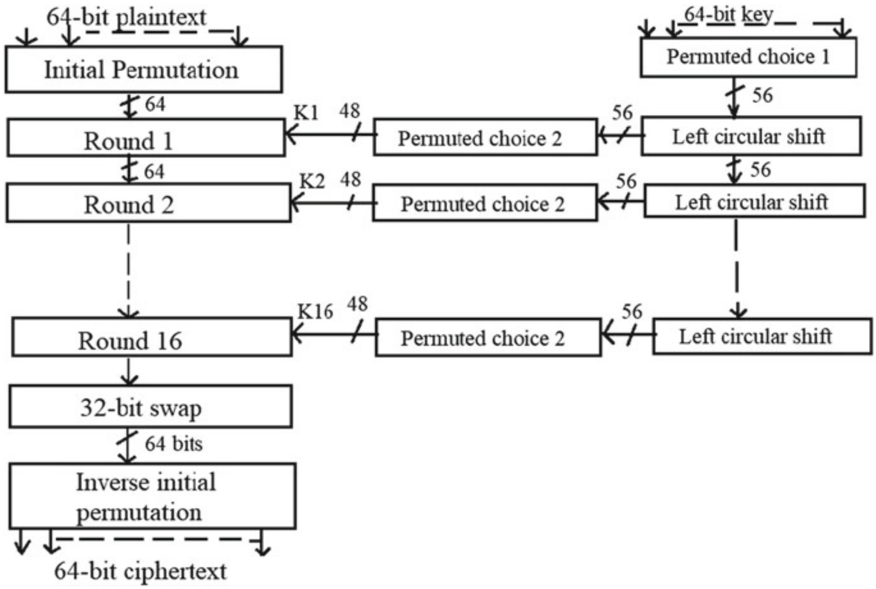


Fig. 1 Structure of DES encryption algorithm

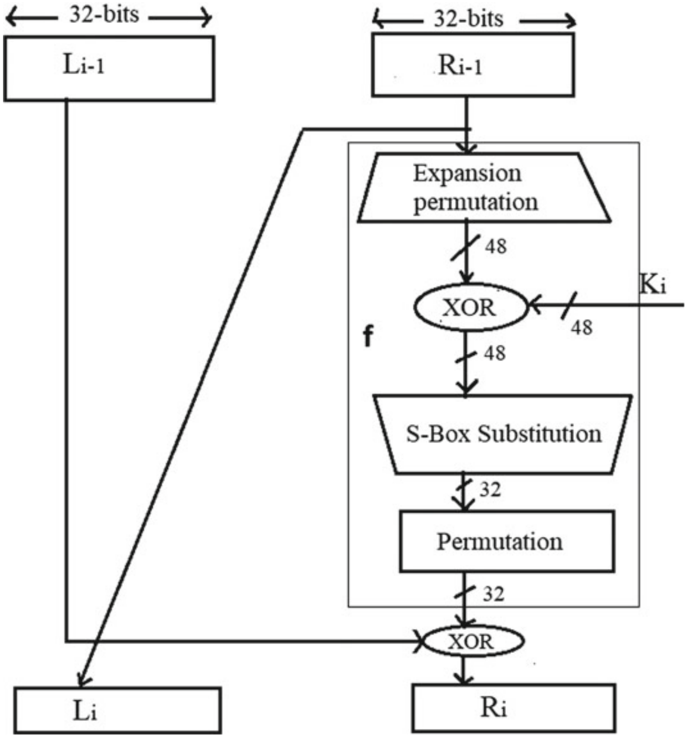
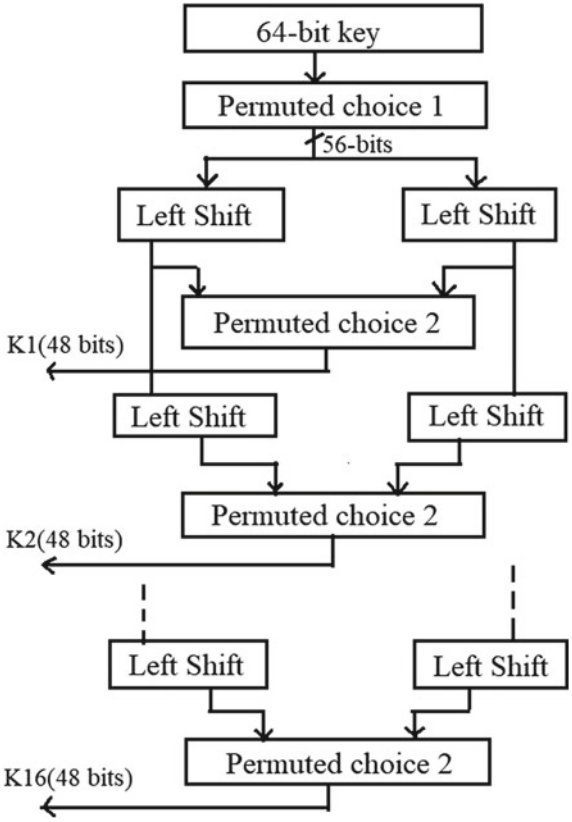


Fig. 2 DES round function

Fig. 3 DES Key Schedule



3.2 Key Schedule Process

The key generation schedule, as shown in Fig. 3, is used to obtain the 16 keys that will be required during the 16 rounds [11]. In the case of the first iteration round, the permutation method is used to obtain the 56-bit key from the 64-bit input key. This 56-bit key is divided into two halves, i.e. left and right blocks. The keys required for the subsequent rounds are obtained by cyclically shifting the 56-bit keys of the previous round. The 48-bit round key is obtained by permuting the cyclically shifted 56-bit key.

3.3 Decryption Operation

The steps involved in the decryption process are the same as the steps involved in encryption process. The subkeys applied during each round are in the reverse order that of the encryption process [12].

4 Overview of AES Encryption

4.1 Description of AES Encryption Operation

Advanced encryption standard (AES) is a symmetric encryption algorithm that implies that the same key is utilized during the encryption and decryption processes. The input plain text data is 128 bits in size, and key lengths can be 128, 192 or 256 bits. AES is an iterative cipher. In one iteration, it encrypts 128 bits. Thus, for each round the 128-bit data and 128-/192-/256-bit key are required [13]. Depending on the length of the key used, the number of rounds during encryption varies, as shown in Table 1. There are ten rounds for a 128-bit key, twelve rounds for a 192-bit key and fourteen rounds for a 256-bit key.

Using the original input key, different keys are generated for each round iteration. As shown in Fig. 4, each round consists of four steps; they are substitution of bytes, shifting the rows, mixing the columns and adding the round key. The 128 bits of plain text input data are converted into 16 blocks of bytes. These 16 blocks are arranged into a matrix with four rows and four columns [13]. The steps involved in all rounds are similar except the last round. During the encryption process, initially, the data input is XORed with the key; i.e. the input data matrix is XORed with the four words of the key given, where each word is a 32-bit size block. The steps involved in each round are described as below:

- **Substitution of bytes:** Using a pre-defined table (S-box), the 16 input bytes are substituted.
- **Shifting of rows**: The rows of the 4 × 4 matrix except the first row are shifted to the left depending on the row location. The second row is circular left-shifted by one-(byte) position. The third row is circular left-shifted by two-(byte) position. The fourth row is circular left-shifted by three-(byte) position.
- **Mixing the columns**: Each column of the data matrix is treated as polynomial, and a certain function is performed on it, and new column is obtained. Thus, all columns of the matrix are evaluated in such a manner. This step acts as a diffusion factor of the AES encryption process.
- **Adding the round keys**: The 16 bytes of the matrix are treated as 128 bits and are XORed with the 128-bit key generated in each round using the key schedule [14].

Table 1 Key lengths and number of rounds for AES

AES	Structure of AES		
Type	Input data size	Key lengths	nr = Number of rounds
AES-128	128 bits	128 bits	10
AES-192	128 bits	192 bits	12
AES-256	128 bits	256 bits	14

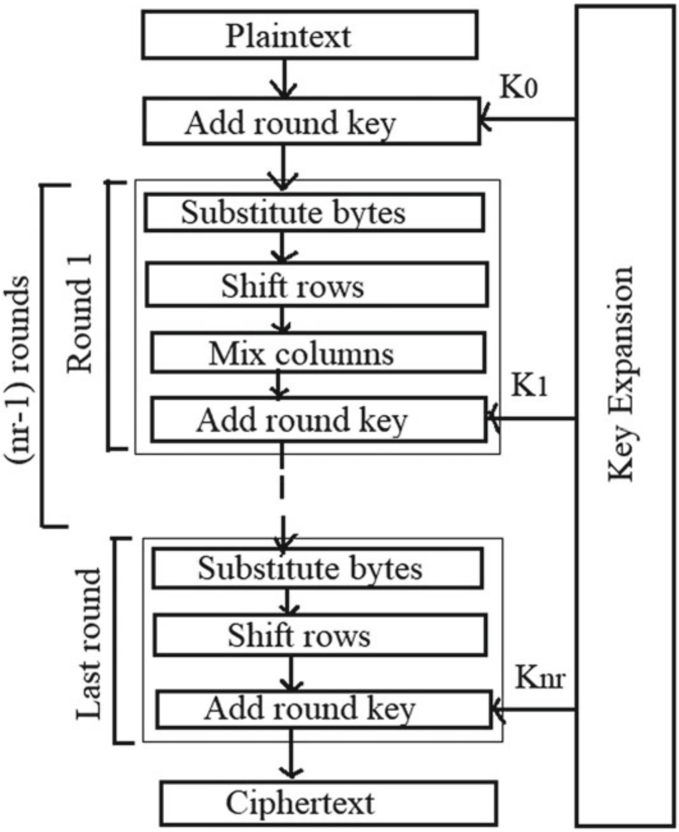


Fig. 4 Structure of AES encryption algorithm

4.2 Key Schedule Process

Before the rounds in the encryption process start, the data input is XORed with the key. Thus, for the given key length the number of subkeys required is equal to one plus the number of rounds. Thus, for encryption process with 128-bit key length, it requires 11 subkeys. For 192-bit key length, it requires 13 subkeys, and for 256-bit key length, it requires 15 subkeys [15]. The key that is used during the operation in each round is of length 128 bits. So, key schedule operation is carried out to obtain the required 128-bit key for each round. The AES key generation schedule is carried out in terms of one word where one word equals 32 bits. Furthermore, the subkey generation is tracked in terms of the words. Key used in each round consists of 4 words (128 bits). Thus, for encryption where input is 128-bit key, considering requirement of 11 subkeys throughout the encryption process, 44 words are generated during the key generation operation. Similarly, 52 words and 60 words are generated during the key generation operation for input key length of 192 bits and 256 bits, respectively [15]. The key generation has three steps: byte substitution, rotation and XOR with round constant (RCON).

4.3 Decryption Operation

The steps involved during decryption can be obtained by directly reversing the order of the steps in the encryption process [16]. The steps involved in each round are also reversed in this case. The encrypted data is given as the input during the decryption process.

5 File Operations in Verilog

To perform operations by the means of files, Verilog offers various system tasks and functions [17]. They facilitate the user to open and close files, assign a value to a variable by reading value from a file and capture the output into a file.

The function $fopen is utilized to open the input files, by passing appropriate filename as the argument and specifying type argument as ‘r’ which implies that the specified file is opened in read mode. This function returns a unique identifier value for a file which is stored in a variable. The function $fscanf is used to analyse the text read from the file as stated by the format and stores the result in the argument specified. In this case, the value read is stored in an argument, which is the input for the Verilog testbench. It reads until the end of file is reached, by utilizing the function $feof, which examines the end of file. Lastly, whenever the output value is evaluated, it is captured in a file using $fmonitor task [17].

6 Implementation

The proposed design is realized in Verilog HDL using Microchip’s Libero SoC Design Suite v2021.1 targeting its PolarFire family with device MPF500T-1FCG1 152I. Libero system on chip (SoC) is the FPGA design suite offered by Microchip Technology. The tools used for simulation and synthesis are ModelSim ME 2020.4 Pro and Synplify R202009MSP1-1, respectively. They are integrated with the Libero Design Suite.

In the Verilog code implementation of DES encryption algorithm, the data and key input are provided as input to the top-level DES module and it provides encrypted output. The top-level DES main module instantiates the key generation module, which generates the required number of keys, to be used for each round iteration processing. A generate block is used to instantiate the f function for the sixteen rounds of iteration. Also, the initial permutation module and inverse initial permutation module are instantiated in the top-level DES module to perform initial permutation operation and inverse initial permutation operation, respectively.

In the Verilog code implementation of AES encryption algorithm, the data and key input are provided as input to the top-level AES module and it provides encrypted output. The top-level AES module instantiates the key generation module, which generates the required number of keys, to be used for each round iteration processing. To perform each round of processing, the module named round is instantiated as number of times depending on the number of rounds required for the given key length. For the AES-128, module round is instantiated nine times, in the top-level module. For AES-192, module round is instantiated eleven times. For AES-256, module round is instantiated thirteen times. The round module then instantiates the modules—subbytes, shiftrow, mixcolumn and addroundkey, which then performs the required operations. The module named roundlast is instantiated for the last iteration, which does not perform the mixcolumn operation.

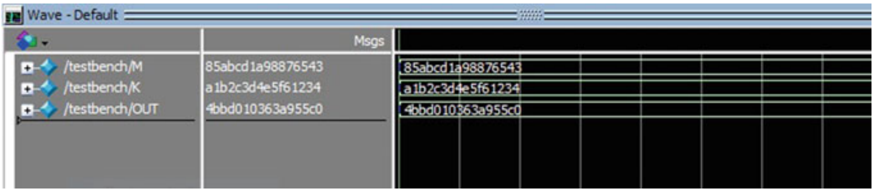
In the proposed design, files are used in the Verilog testbench, to read the input values and to capture the output value of simulation. The file named Data_in_file is used to store the input data, and the file named Key_file is used to store the key value. These files are stored in the simulation folder of the project folder location, and the output file named Output_file is generated in the same location after the simulation step completion.

After the place and route step, the reports section in the Libero tool contains the log file which comprises the resource usage report that lists the type and percentage of resource used for each resource type relative to the total resources available for the chip. And verify power step during the design flow allows the user to analyse the power consumption within the device, and the report for the same is available in the reports section in the Libero tool.

7 Results and Discussion

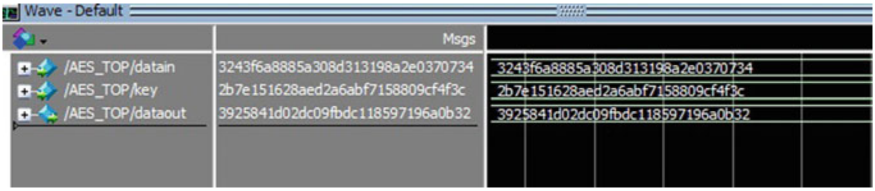
Libero SoC Design Suite provides tools to perform steps associated with FPGA design flow. Thus, for the Verilog codes for DES and AES encryption algorithms, have carried out steps of FPGA design flow such as Design entry, Simulation, Synthesis, Place and route, timing and power analysis.

ModelSim simulator offered by Mentor Graphics acts as a design functionality verification tool and allows users to verify HDL source code. Libero SoC Design Suite has integrated the ModelSim Pro tool in its environment. Figure 5 shows the ModelSim graphical user interface (GUI) that is launched when pre-synthesis simulation step is performed for DES encryption operation. Figure 6 shows the simulation waveform when the data input is 128 bits and the key is 128 bits in size. For AES encryption Verilog code, Fig. 7 shows the simulation waveform when the data input is 128 bits and the key is 192 bits in size. Figure 8 shows the simulation waveform



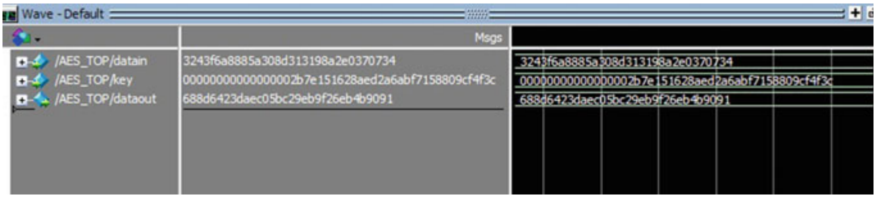
```
/testbench/M     85abcd1a98876543
/testbench/K     a1b2c3d4e5f61234
/testbench/OUT   4bbd010363a955c0
```

Fig. 5 Simulation output of DES encryption



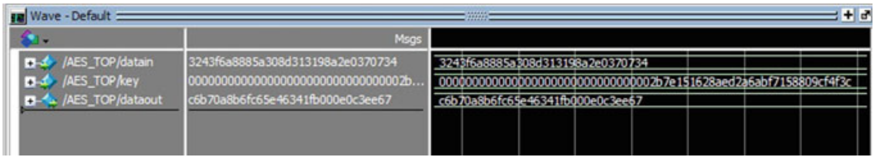
```
/AES_TOP/datain    3243f6a8885a308d313198a2e0370734
/AES_TOP/key       2b7e151628aed2a6abf7158809cf4f3c
/AES_TOP/dataout   3925841d02dc09fbdc118597196a0b32
```

Fig. 6 Simulation output of AES-128 encryption



```
/AES_TOP/datain    3243f6a8885a308d313198a2e0370734
/AES_TOP/key       00000000000000002b7e151628aed2a6abf7158809cf4f3c
/AES_TOP/dataout   688d6423daec05bc29eb9f26eb4b9091
```

Fig. 7 Simulation output of AES-192 encryption



```
/AES_TOP/datain    3243f6a8885a308d313198a2e0370734
/AES_TOP/key       000000000000000000000000000000002b...
/AES_TOP/dataout   c6b70a8b6fc65e46341fb000e0c3ee67
```

```
00000000000000000000000000000000002b7e151628aed2a6abf7158809cf4f3c
```

Fig. 8 Simulation output of AES-256 encryption

when the data input is 128 bits and the key is 256 bits in size. The output of the AES encryption operation is 128 bits in length here.

The resource usage report of the device is available after the place and route operation. Each type of resource available on the device lists the amount and percentage of the resource used relative to the total resource available on the chip. For example, Table 2 shows the resource utilized for the DES and the AES encryption techniques. Furthermore, Table 3 depicts the percentage resource utilization. The verify power step in the Libero design flow allows us to foresee the total power consumed within the device for the given design overall and also in depth. Table 4 shows the total power consumed for the given design on the device targeted.

Table 2 Resource usage

Type	Total	Used			
		DES	AES-128	AES-192	AES-256
4LUT	481272	5081	56123	62849	76885
DFF	481272	0	0	0	0
I/O register	582	0	0	0	0
Logic element	481272	5081	56123	62849	76885

Table 3 Percentage resource usage

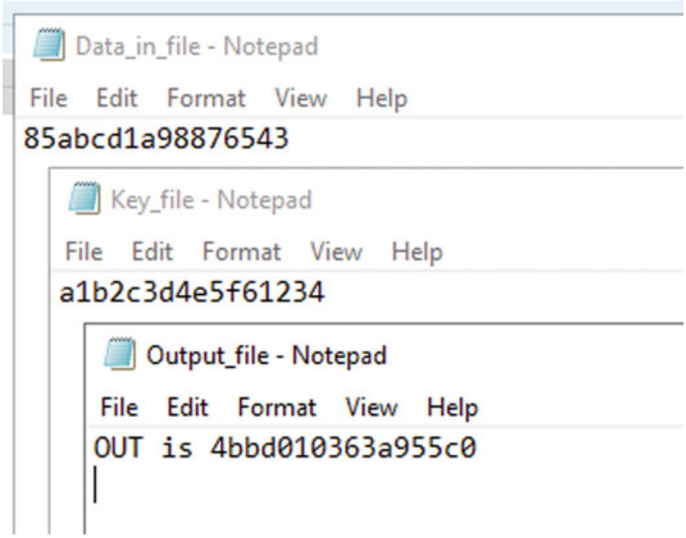
Type	Resource usage (%)			
	DES	AES-128	AES-192	AES-256
4LUT	1.06	11.66	13.06	15.98
DFF	0	0	0	0
I/O register	0	0	0	0
Logic element	1.06	11.66	13.06	15.98

Table 4 Power summary

Algorithm	Total power (mW)
DES	157.102
AES-128	176.183
AES-192	177.437
AES-256	181.456

Figure 9 shows the input and the output files in the case of DES encryption algorithm. Figure 10 shows the input and the output files, when the input is 128 bits and the key is 128 bits in size. For AES encryption Verilog code, Fig. 11 shows the input and the output files when the data input is 128 bits and the key is 192 bits in size. Figure 12 shows the input and the output files, when the data input is 128 bits and the key is 256 bits in size.

During the mid-1970s to the mid-1990s, DES was the symmetric encryption algorithm that was used widely. However, the 56-bit keys were no longer secure and could be broken using exhaustive key search. The advancement in the technology aided in breaking DES with less cost and less time involved. Thus, AES was suggested around the mid-1990s, which supported three key lengths of 128, 192 and 256 bits and also provided security against brute-force attacks. Also, AES was found to be faster than DES.

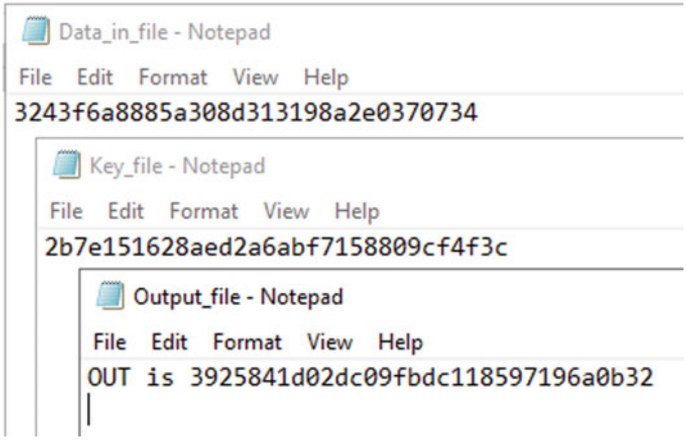

Data_in_file - Notepad
File Edit Format View Help
Key_file - Notepad
File Edit Format View Help
Output_file - Notepad
File Edit Format View Help

```
85abcd1a98876543
```

```
a1b2c3d4e5f61234
```

```
OUT is 4bbd010363a955c0
```

Fig. 9 Input and output files of DES encryption


Data_in_file - Notepad
File Edit Format View Help
Key_file - Notepad
File Edit Format View Help
Output_file - Notepad
File Edit Format View Help

```
3243f6a8885a308d313198a2e0370734
```

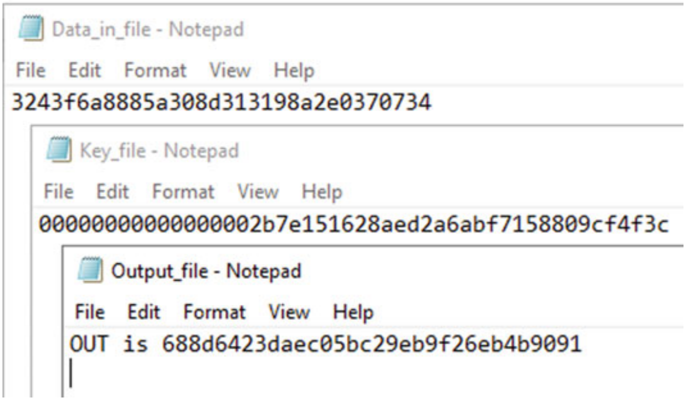
```
2b7e151628aed2a6abf7158809cf4f3c
```

```
OUT is 3925841d02dc09fbdc118597196a0b32
```

Fig. 10 Input and output files of AES-128 encryption

8 Conclusion

The data encryption standard (DES) algorithm is a symmetric block cipher that uses 64-bit data input and 64-bit key. The advanced encryption standard (AES) algorithm is a symmetric block cipher that uses 128-bit input data and a key length of 128, 192 and 256 bits. Verilog code is implemented for DES and AES encryption algorithms on the Libero SoC Design Suite, by providing the data input and the key for the

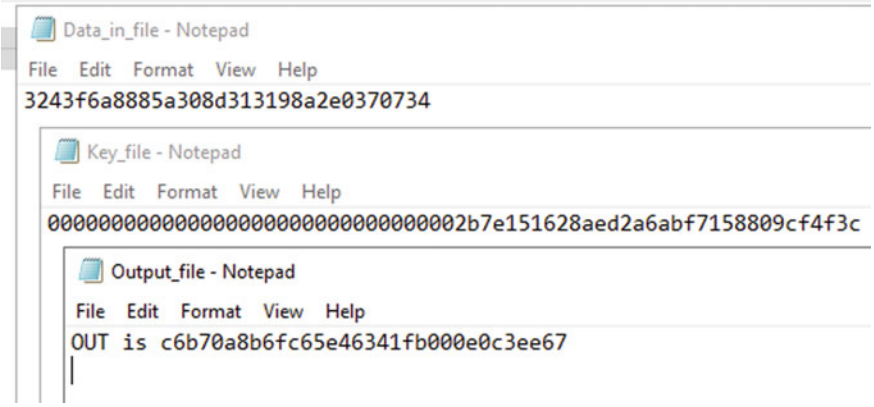

Data_in_file - Notepad
File Edit Format View Help
Key_file - Notepad
File Edit Format View Help
Output_file - Notepad
File Edit Format View Help

```
3243f6a8885a308d313198a2e0370734
```

```
00000000000000002b7e151628aed2a6abf7158809cf4f3c
```

```
OUT is 688d6423daec05bc29eb9f26eb4b9091
```

Fig. 11 Input and output files of AES-192 encryption


Data_in_file - Notepad
File Edit Format View Help
Key_file - Notepad
File Edit Format View Help
Output_file - Notepad
File Edit Format View Help

```
3243f6a8885a308d313198a2e0370734
```

```
000000000000000000000000000000002b7e151628aed2a6abf7158809cf4f3c
```

```
OUT is c6b70a8b6fc65e46341fb000e0c3ee67
```

Fig. 12 Input and output files of AES-256 encryption

Verilog testbench through the text files, and attained the ciphertext, i.e. encrypted output in a file. Simulation waveform result is observed. Also, resource utilization and power usage reports are obtained for both the encryption algorithms.

References

1. C. Paar, J. Pelzl, *Understanding Cryptography, A Textbook for Students and Practitioners* (Springer, Heidelberg, 2010)
2. A. Azad, Efficient VLSI implementation of DES and triple DES algorithm with cipher block chaining concept using Verilog and FPGA. Int. J. Comput. Appl. **44**(16), 0975–8887 (2012)
3. M. Dworkin, E. Barker, J. Nechvatal, J. Foti, L. Bassham, E. Roback, J. Dray, Advanced Encryption Standard (AES), in *Federal Information Processing Standards Publication 197* (2001)

4. V. Kristianti, E. Wibowo, A. Pertiwi, H. Afandi, B. Soerowirdjo, Finding an efficient FPGA implementation of the DES algorithm to support the processor chip on smartcard, in *The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)* (2018)
5. M. Yewale, M. Sayyad, Implementation of AES on FPGA. IOSR J. VLSI Signal Process. (IOSR-JVSP) **4**(5), 65–69 (2014)
6. R. Patil, V. Shete, Implementation of advanced encryption standard on FPGA. Int. J. Eng. Res. Technol. (IJERT) **2**(8) (2013)
7. P. Chodowiec, K. Gaj, Very compact FPGA implementation of the AES algorithm, in *International Workshop on Cryptographic Hardware and Embedded Systems, CHES* (2003)
8. Jadhav, A., Choudhari, P., Gherade, T., Wagaj, S.: Implementation of Advanced Encryption Standard (AES) on FPGA. International Journal of Engineering Science and Computing, volume. 7, Issue. 5, May 2017
9. T. Cedric, R. Suchithra, A comparative study on AES 128 BIT AND AES 256 BIT. Int. J. Sci. Res. Comput. Sci. Eng. **6**(4), 30–33 (2018). E-ISSN: 2320-7639
10. F. Khan, R. Shams, A. Hasan, N. Hasan, Implementation of data encryption standard (DES) on FPGA. J. Comput. Sci. Newports Inst. Commun. Econ. **5** (2014). ISSN: 2226-3683
11. A. Singh, M. Marwaha, B. Singh, S. Singh, Comparative study of DES, 3DES, AES and RSA. Int. J. Comput. Technol. **9**(3) (2013)
12. S. Zeebaree, DES encryption and decryption algorithm implementation based on FPGA. Indo. J. Electr. Eng. Comput. Sci. **18**(2), 774–781 (2020)
13. F. Noorbasha, Y. Divya, M. Poojitha, K. Navya, A. Bhavishya, K. Koteswara Rao, K. Hari Kishore, FPGA Design and implementation of modified AES based encryption and decryption algorithm. Int. J. Innov. Technol. Explor. Eng. (IJITEE) **8**(6S) (2019). ISSN: 2278-3075
14. S. Dalal, K. Kasat, FPGA implementation of AES algorithm using cryptography. Int. J. Electron. Commun. Soft Comput. Sci. Eng. (2017). ISSN: 2277-9477. (IETE Zonal Seminar—Recent Trends in Engineering and Technology)
15. M. Pitchaiah, P. Daniel, Praveen, Implementation of advanced encryption standard algorithm. Int. J. Sci. Eng. Res. **3**(3) (2012). ISSN 2229-5518
16. N. Srinivas, M. Akramuddin, FPGA based hardware implementation of AES Rijndael Algorithm for encryption and decryption, in *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (2016)
17. IEEE Std 1364-2001, *IEEE Standard Verilog Hardware Description Language* (The Institute of Electrical and Electronics Engineers, New York, 2001)

Smart Lung Cancer Detector Using a Novel Hybrid for Early Detection of Lung Cancer

Praveen Tumuluru, S. Hrushikesava Raju, M. V. B. T. Santhi, G. Subba Rao, P. Seetha Rama Krishna, and Ashok Koujalagi

Abstract Nowadays, most people are dying because of late identification of lung cancer. Hence, to extend the survival rate, there are mechanisms used to detect lung cancer in the early stages so that proper treatment to be taken to get cured. One of the important factors that cause lung cancer is smoking. Maybe present smokers or quitted smokers may be affected by the chance of getting lung cancer. There are stages used in the early detection of lung cancer such as image acquisition, preprocessing, segmentation, lobe separation, feature extraction, and classification. The enhanced method is defined as a combination of Chronic Obstructive Pulmonary Disease (COPD) and lung cancer is made detected enhanced Low Dose Computed Tomography (LDCT) along with guaranteed convergence particle swarm optimization (GCPSO). The advantage of the proposed approach is fully automated and that led to consuming less time for analysis and processing. This systematic approach along with the necessary sensors and their integration in order to get feedback about the images obtained from enhanced LDCT scans. The sensors are defined in the Internet of Things (IoT) would be capable to analyze the scanned images and efficiently generate a report that states the person is cancer affected or not.

Keywords Lung cancer · Enhanced LDCT · GCPSO · Early prediction · Internet of Things · Hybrid framework · Accuracy · Scanned images · Reports

1 Introduction

There are a number of studies on lung cancer and its impact in the society. One of major significant factors considered is smoking. There are many people because of their habits may be attacked with lung cancer. Not only old people who are at ages 50–60 group or above 60 ages are fall in lung cancer victims although there are young generation people are there in less case. In order to make such victims survive further, there is a need to detect the cancer in advance that means early stage detection.

P. Tumuluru (✉) · S. Hrushikesava Raju (✉) · M. V. B. T. Santhi · G. Subba Rao ·
P. Seetha Rama Krishna · A. Koujalagi
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_64

There are many approaches used such as chest X-ray, CT scans, variations of screenings such as k-means clustering, k-median clustering, particle swarm optimization, inertia-weighted particle swarm optimization, and guaranteed convergence particle swarm optimization (GCPSO), biopsies, and many other approaches. Among them, LDCT is the best technique to reduce the lung cancer as well as GCPSO is the best analyzation approach. By considering many people lives into mind, the context of early detection of lung cancer is determined through modified and enhanced LDCT Technique. This novel technique is an extension of LDCT with few more features and aims to achieve more accuracy than LDCT. Hence, the proposed approach will analyze the LDCT and enhanced LDCT through the results obtained over the many samples.

There are few states such as Haryana, Delhi, Kerala, Mizoram, and Karnataka, the cancer persons when take a sample of 10,000 samples are more on the average of 120–137. Hence, in order to detect the lung cancer early, the enhanced LDCT would reduce such number of lung cancer cases. When people are healthy in India, the families are healthy and that leads to regions are healthy. Its result directly leads to healthy India and is free from lung cancer. As compared to LDCT, the enhanced LDCT approach would the benefit the society in determining the lung cancer in early stages only when applied to the smokers or smokers who quitted. The hybrid approach involves the following steps.

(1) The purpose of COPD with LDCT is described and provides the advantages of it.
(2) The additional GCPSO is used in the background and its purpose also demonstrated.
(3) The combined effect is studied and its impact, and consequences are noted.
(4) The results are monitored using sensors of Internet of Technology.
(5) The accuracy is observed when using IoT when compared with LDCT.
(6) The pictorial representation of graph must depict the performance of hybrid approach vs LDCT traditional approach.

2 Literature Review

There are many advances in the technology for early detection. Among these listed studies as below in which most studies are publications, but one listed at last is a project on analysis of cancer over few samples and determining the best early detection approach.

According to Li et al. referred in [1], identification of lung-cancer-related genes with the shortest path approach in a protein–protein interaction network. The method proposed is computational approach which extracts cancer cells from PPI network, which is constructed based on STRING which helps to select genes of type NSCLC and SCLC. It also considers the genes that have significant genes value. The kind of cancer related genes ESR1, FDXR, ABCA1, IRS1, HSP90AA1, FOXM1, and IGBP1 is identified using shortest path algorithm such as Dijkstra's algorithm. The candidate genes identified during this PPI network have more cancer cell weightage than those compared with gene expression profiles. Hence, the more efficiency is obtained when compared with existing approaches in the cancer research.

According to Sutedja et al. referred in [2], new techniques for early detection of lung cancer. There are few methods used for early detection of cancer which are such as sputum cytology which guarantee lives with rate 34–37 only. Another method in this category is randomized clinical trials which guarantee the survival rate is 63–67%. Another method in this cohort is intraluminal bronchoscope treatment which is alternative to discussed methods in order to achieve efficiency of survival rate using parameters accuracy and minimal invasiveness. Among these methods, bronchoscope although accurate would be costly treatment when compared.

According to Tantraworasin et al. mentioned in [3], lung cancer: Diagnosis and treatment approach. There are few multi-modality approaches are used such as PET, CT, EBUS, and VATS which play a key role in lung cancer treatment along with treatments such as surgery, chemotherapy, radiotherapy, targeted therapy, and immunotherapy. Advanced and target therapies are used for lung cancer treatment with less side effects.

According to Thakur et al. mentioned in [4], lung cancer identification: a review on detection and classification. The method defined in lung cancer detection and their classification is demonstrated such as a greater number of CT scan images are taking would help to get more information about nodules and accurate assessment. In this, the various approaches in CAD are analyzed and their studies would present about nodules.

According to Judice and Geetha specified in [5], a novel assessment of various bio-imaging methods for lung tumor detection and treatment by using 4-D and 2-D CT images. The machine learning method called Hidden Markov Model is presented that takes less processing time and is automated compared with traditional methods. The preprocessing stage uses histogram and mean error square techniques to remove noise from the cancer CT images of type 2D and 4D. The automation process would decrease the diagnosis time.

According to Latimer and Mott demonstrated in [7], MD, Naval Hospital Pensacola, Pensacola, Florida, Lung Cancer: Diagnosis, treatment principles, and screening. This report would provide the symptoms of lung cancer are cough, systemic symptoms such as weight loss and anorexia, hemoptysis, and dyspnea. The techniques CT and PET are performed includes three stage simultaneous steps such as tissue diagnosis, staging, and functional evaluation. The treatment includes new targeted molecular therapies align with a cluster of experts for monitoring and analysis. The one reason for getting lung cancer is smoking habit and past history of it even in case of smoking quitted persons.

According to Senthil Kumar et al. detail in [8], lung cancer detection using image segmentation by means of various evolutionary algorithms, here are approaches defined in case of lung cancer detection which are such as k-means clustering, k-median clustering, particle swarm optimization, inertia-weighted particle swarm optimization, and guaranteed convergence particle swarm optimization (GCPSO). Among these, GCPSO has proved that it is providing more accuracy than others when tested in MATLAB environment. According to Henschke et al. mentioned in [10], early lung cancer action project: overall design and findings from baseline screening, many methods are analyzed over a 1000 samples and concluded that biopsies, CT scan and X-ray scans when applied, LDCT will be approved best method to make the severity to the normal.

According to Rampinelli et al. mentioned in [11], this demonstrates about LDCT method for early detection lung cancer. This method improves diagnosis quality based on parameters such as tube velocity, current, and speed rotation. The double reading and CAD system won't increase the cancer detection but leads to detect a greater number of pulmonary nodules from 50 to 76%. The COPD along with LDCT provides a sensitivity and specificity of 63% and 88%, respectively.

In the respect of sources specified in [6, 9] resemble the discussion where former describes on LDCT technique is more better when compared to chest X-ray scan and would lower the people from dying from lung cancer, when this technique is to be taken every year till 74 years if the person is having good health that may extend the life of the person, and latter depicts on screening technique which is a careful refinement set of activities in order to detect the lung cancer in the beginning state itself and screening technique's advantages and disadvantages are both discussed in detail and are pinned in this study. In the depicted study focused in [12], the purpose of LDCT is explained in terms of it takes very less time to identify the abnormalities in the lung, so that proper treatment is initiated and that extends survival of the person when compared with traditional X-ray scan. With respect to source specified in [13], there AI and DL are used to detect the even small dots related to lung cancer and are handled in efficient manner. The following is the procedure of steps involved in this study.

The sequence of tasks will go in the fashion mentioned in Fig. 1 that leads output of one task as input to other task that are in relation. This leads to make prediction or decisions based on the quality report generated from DL and radiomics.

From the perspective of source given in [14], the LDCT used here detects new subsolid nodules from never smokers' database but not from smokers' database, those would have less probability of lung cancer and have more probability to be cured by spontaneous treatment. In the aspect of demonstration mentioned in [15], the LDCT technique is applied over non-smokers as well as smokers of both the sexes in a trial manner based on NELSON dataset, results early detection of lung cancer which avoids the growth in the death rates. With the regard of source mentioned in [16], the X-ray images are taken from online data repositories for study, the hybrid framework of few methodologies such as VGG, STN, and CNN are combined and is termed as VDSNet is worked based on metrics such as F-score, precision, accuracy, and recall. The hybrid approach proved 73% accuracy compared to existing approaches. With respect to the [17], the LESH is proposed along with sensitivity analysis (SA) would produce better results after analyzing the quality images to accurately predict the lung cancer. In the aspect of source specified in [18], logit boost classifier is used to classify the samples that are lung cancer affected or normal with 99.23% accuracy. The methods such as filtering and noise removal are done in the preprocessing step. In this, 13 nodules are extracted after applying LIDC-IDRI, in which 4 are significant

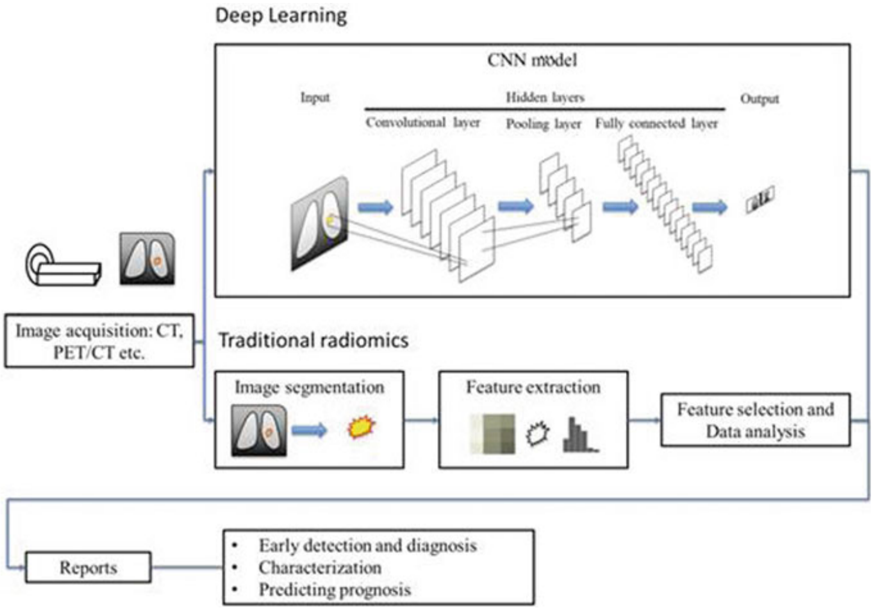


Fig. 1 Steps involved in the AI and radiomics for the lung cancer

by the PCA. In the process of source specified in [19], the proposed approach focus on images that are segmented with varying contrast and improved the contrast and reduce the false-positives in a better manner using discriminative texture features. As per the source specified in [20], the combination of GPU and multiple CPUs would act like a tool that automatically distinguish the affected nodules and produce better results than other existing approaches in this domain. It uses reading, normalization, segmentation, extraction, and classification as steps.

Here, the IOT is also used in a way that would contain the functionality of LDCT and GCPSO, which could generate reports that could be efficiently analyzed for better prediction and minimization of loss rate during the analysis process. There were few IOT studies are demonstrated but their intended theme is used somehow in the processing of the task.

According to the information in [21], the current pandemic is to be limited by wearing a digital mask that reports the virus in the location in which the user is now staying. The created mask will provide statistics about the things in the current environment. Regarding the source mentioned in [22], the IoT is utilized to determine the location and automatically changes its currency into the user's currency. In this scenario, user flexibility is provided.

According to the study cited in [23], the IOT is employed in power banks and portable devices to interchange charging power in a user-friendly environment. The personalized charging method is accomplished through the use of a specially built software and IOT technology. According to [24], the IOT is employed in communicating the weighted objects falling to the other devices in order to collect it and gently send it to the earth using an automated net.

According to [25], the IOT is utilized in industries where the amount of gas is monitored and detects leaks if any are detected during the passage of gas across the pipes that are placed from the source to the destination. This detection prevents dangerous accidents from happening to humans.

According to the source indicated in [26], the IOT is useful over the users in such a way that users' health bulletins are monitored and a guidance to maintain fitness based on food diet is provided. According to the source provided in [27–30], the IOT and GSM are employed in determining the popular destinations when a user wishes to go around the world. The top places and ranked places in those cities, as well as a route map, will be presented.

As per the source mentioned in [31], the COPD patients should be more in the LDCT screening tests and they have air particles in the scanned images and would suffer from respiratory issues. The LDCT scans would process the problem in less time and would result in quality images for further analyzation. Regarding to the description of [32], the traffic in the social multimedia is analyzed and focused on QoS requirements like low latency and high bandwidth, in turn reduce the energy incurred over profile cloning, identity theft, and gathering sensitive data. As per description given in [33], the lights in the streets are taking more energy and also negligence in saving of power made a novel work of saving the energy, as well as reducing the cost of streetlights energy.

In all the mentioned tasks, the significance of IOT is demonstrated in a way that would read data, process such data and sends a report to the communication center or authorized gadget or a person's smart device. The usage of IOT is automation become very fast and would lead to perform the activities in order without human intervention. The IOT based applications can be considered where sensors to be incorporated based on the purpose of the application. Not only few references that are mentioned but also many IOT based applications are applicable to this domain but all those are differentiated based on architectures, measures, and the kind of sensors that are adapted to finish the task.

3 Proposed Approach

In this, the enhanced LDCT would work in the stages such as image acquisition, preprocessing, segmentation, lobe separation, feature extraction, and classification. Among these, first stage image acquisition deals with extracting 3D and 4D images of lungs for further analysis. The second stage preprocessing deals with reducing the noise in CT scanned images. The third stage called segmentation deals with partition the preprocessed CT scanned images into multiple regions. The fourth stage called lobe separation deals with is the angle in camshaft degrees between the maximum lift points, or centerlines, of the intake and exhaust lobes and it affects the

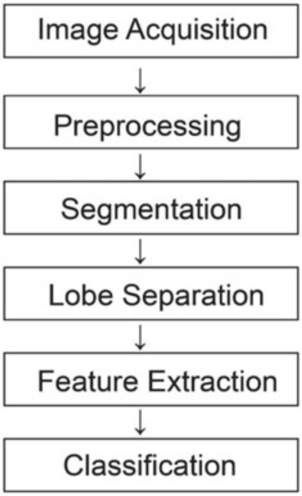


Fig. 2 Stages in the cancer detection

amount of valve overlap; that is the brief period of time when both the intake and exhaust valves are open. The fifth stage called feature extraction deals with physical dimensional measures and gray-level co-occurrence matrix (GLCM) method are used. The last stage classification deals the sample under TNM-8 where T for the extent of the primary tumor, N for lymph node involvement, and M for metastatic disease. T-classification is performed using CT, the N- and M-classification using CT and PET-CT.

The above steps are also taken in other manner such as reading and normalization, segmentation, extraction and classification but the output achieved is similar to the steps described as in Fig. 2.

The proposed approach is hybrid framework which is a combination of COPD with LDCT and GCPSO that would process the person efficiently and is automatic approach. This automatic nature would speed up the task to the extent possible and generate more quality images of persons. Hence, the hybrid framework is divided into 4 modules.

(A) COPD with LDCT—Also called enhanced LDCT
(B) GCPSO
(C) IOT for analyzation and automatic prediction
(D) Recommendations_based_Output.

The pseudo procedure of the hybrid framework is as follows:

Step 1: Call COPD_with_LDCT() where patients with COPD have less survival rate, where COPD stands for chronic obstructive pulmonary disease and LDCT stands for Low Dose Computed Tomography.
Step 2: Call GCPSO() for accurate prediction of samples in terms of guaranteed to local extreme. Generate the report that consists of images with nodules, where GCPSO stands guaranteed convergence particle swarm optimization.

```
Step 3: Call IOT as virtual doctor and as an expert—reading the report, Where IOT stands for Internet of Things.
Step 4: Provide recommendations.
```

The pseudo procedure of the enhanced LDCT which is a combination of COPD and LDCT as follows:

```
Pseudo_Procedure COPD_with_LDCT(NLST_based_database[][]):

Step 1: Identifies the itches such as FEV (Forced Expiratory Volume) and FVC (Forced Vital Capacity) ratio should be less than 0.70.
Step 2: Additional itches such as emphysema and chronic bronchitis also to be identified.
Step 3: Testing LDCT for COPD patients increase the risk from 2 to 4-fold when compared against without airflow obstruction.
Step 4: Verify the scores such as Lung Cancer Screening Score (LUSS) and Diffusing Capacity for Carbon Monoxide (DLCO) and reduce such scores in order to increase patient survival rate.
Step 5: While LDCT screening (Spiral Screening) is applied. It produces the quality images in less than a minute. After analyzation, the following of anyone to be resulted.
    5.1 It has to increase false-positive results that means person has lung cancer, but no cancer is present. This leads to no surgeries is required.
    5.2 It may find cases of cancer, but they won't do problem to the patient is called over-diagnosis. This leads to the treatment is not needed.
    5.3 The repeated LDCT screenings result radiations, which may cause healthy person to be affected by cancer.
```

```
Pseudo_Procedure GCPSO (patient_report_images[]):

Step 1: Analyze the scanned images.
Step 2: Compare median, adaptive median, and average filters in which adaptive median are proved to be good.
Step 3: The adaptive histogram equalization technique is used to enhance the image contrast for quality image.
Step 4: This approach proves to have accuracy more than 96% in judging the lung cancer tumor.
```

```
Pseudo_Procedure IOT_as_Virtual_Doctor(GCPSO_Images[]):

Step 1: Trained sensors are used instead of actual doctor, which is called a virtual doctor.
Step 2: The Virtual Doctor is connected automatically to the other analyzation tools like MATLAB and etc.
Step 3: Based on relation and its position, and appearance, the tumor is fixed.
Step 4: Automatically, the screening program assigns the treatment based on the scanning of the type of tumors.
Step 5: Directs the prescription and cost for the treatment.
```

```
Pseudo_Procedure Recommendations_output(report_IOT_Virtual_Doctor):
    Step 1: Based on report from previous module, further LDCT test is suggested based on person immunity.
    Step 2: The follow-up schedule is given in order to increase the false-positives that may extend the survival rate.
    Step 3: Further assessment is followed up after each review, which avoids the death rate.
```

All the above modules when carried out, the first two modules such as COPD_with_LDCT and GCPSO would process the person and outputs the quality image. The remaining two modules such as IOT_as_Virtual_Doctor and Recommendations_Output would analyze the images and fix the tumor spots and suggest the prescriptions and plans to increase survival rate of the person.

The flowchart the shows the flow of activities in order to achieve the goal and produces the prediction with highest accuracy (Fig. 3).

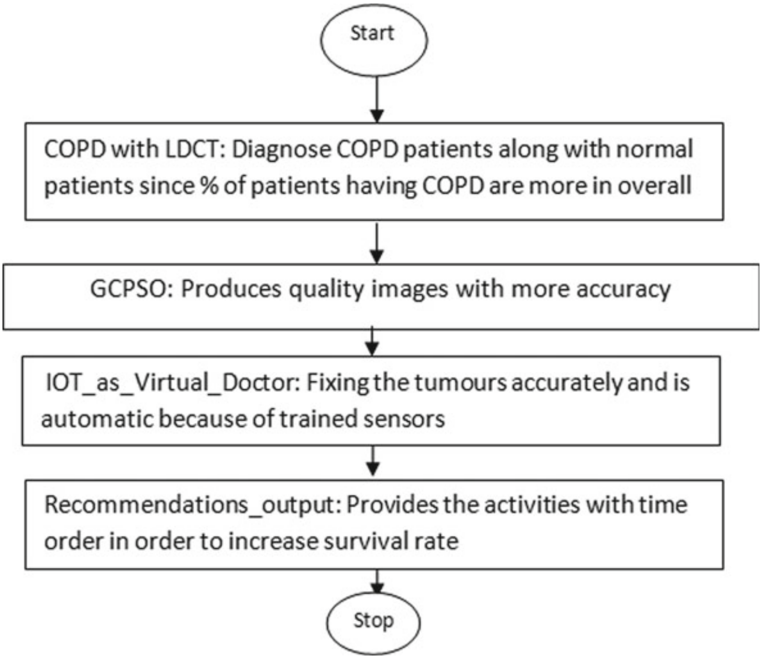

Start
COPD with LDCT: Diagnose COPD patients along with normal patients since % of patients having COPD are more in overall
GCPSO: Produces quality images with more accuracy
IOT_as_Virtual_Doctor: Fixing the tumours accurately and is automatic because of trained sensors
Recommendations_output: Provides the activities with time order in order to increase survival rate
Stop

Fig. 3 Sequence of modules of a hybrid framework for lung cancer detection in a flow graph

4 Results

The output of the defined modules is demonstrated in Fig. 4 for prescribing the actions to do after the analysis.

The process specified in Fig. 4 is integrated with Fig. 2 for efficient extraction of cancer identification. The subsequent recommendations are also guided using virtual doctor as expert and automate the detection in an effective manner.

The retina-based image sensor is used at the IOT based virtual doctor, and there nodules are detected. Based on number of nodules detection, the accuracy is assessed (Fig. 5).

The differences between the traditional LDCT screening and proposed system titled hybrid framework as follows.

The performance of the proposed system is better than only LDCT screening is as follows (Fig. 7).

In Fig. 6, the performance of hybrid LDCT framework is far more additional than the traditional LDCT approach where methods taken on X-axis and % of efficiency on the Y-axis.

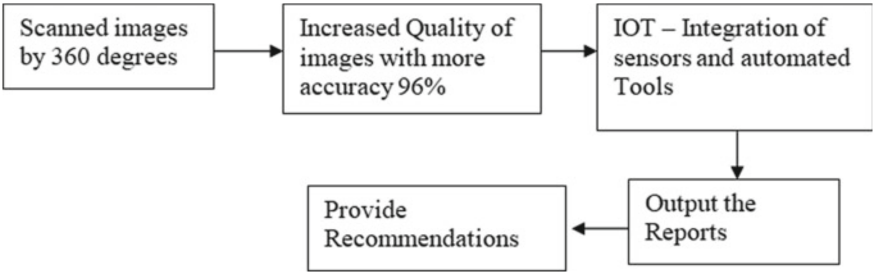

Scanned images by 360 degrees
Increased Quality of images with more accuracy 96%
IOT – Integration of sensors and automated Tools
Output the Reports
Provide Recommendations

Fig. 4 Sequence of outputs produce by the proposed system modules

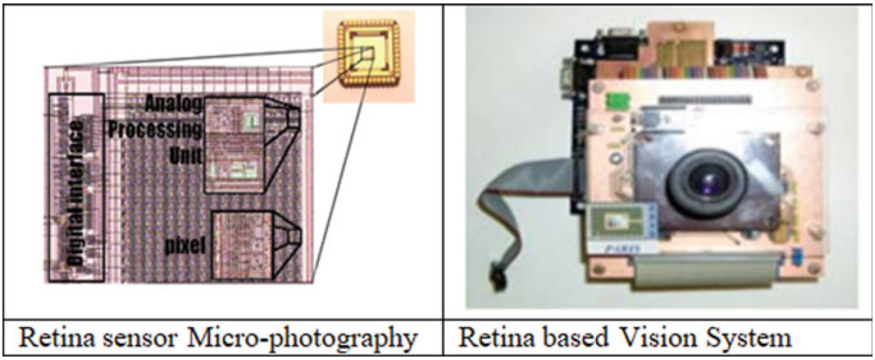
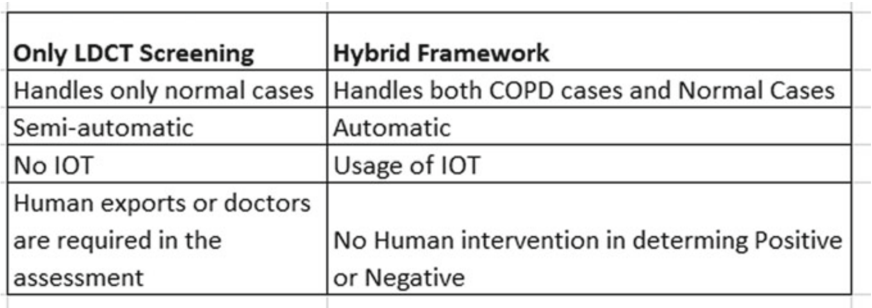


Fig. 5 Retina-based sensor at IOT module



Only LDCT Screening	Hybrid Framework
Handles only normal cases	Handles both COPD cases and Normal Cases
Semi-automatic	Automatic
No IOT	Usage of IOT
Human exports or doctors are required in the assessment	No Human intervention in determing Positive or Negative

Fig. 6 Differences between only LDCT and hybrid framework

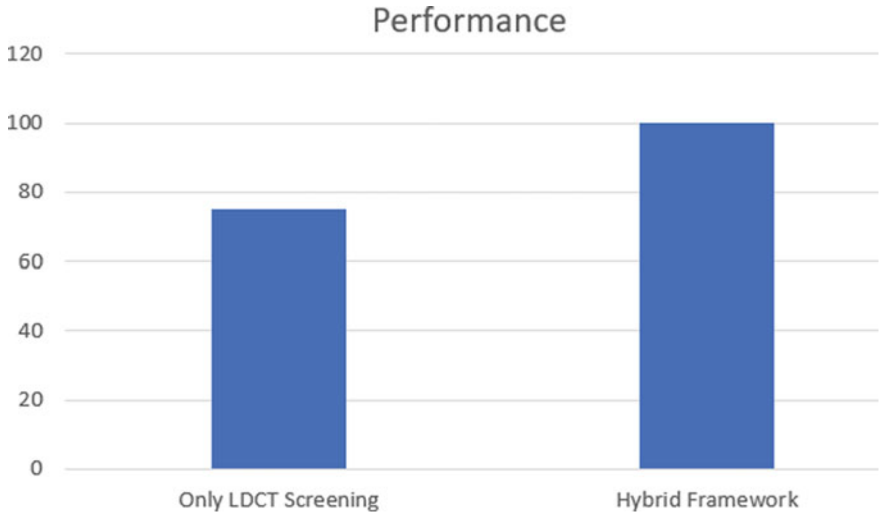


Fig. 7 Performance measured between only LDCT and hybrid LDCT framework

The accuracy of the proposed approach called hybrid LDCT scans versus only LDCT screening is as follows (Fig. 8).

In Fig. 6, the accuracy of hybrid LDCT framework is very good and improved a lot when compared the traditional LDCT approach in terms of NLST and NELSON datasets where methods taken on *X*-axis and % of accuracy on the *Y*-axis.

5 Conclusion

The deaths because of lung cancer are increasing over the globe. To minimize such factor, the traditional LDCT is good and increased the survival rate although the time

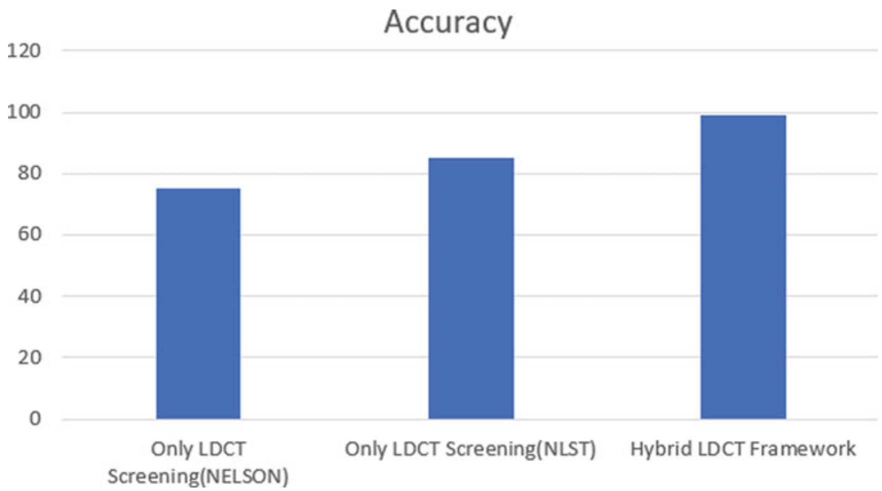


Fig. 8 Accuracy of only LDCT screening (NELSON, NLST) and hybrid LDCT framework

taken to scan is very less compared to other existing approaches. Based on the trend moving toward the automation and accuracy, the LDCT is enhanced by considering COPD as well as GCPSO in producing the quality images that consist of tumors. The sensor with trained feedback is loaded and made as processing unit in analyzing the images. The report is generated with prescription and schedule using inherent automated tools. The suggestions are instructed to the patient and follow-up by time order is done by the recommendation's module. The performance and accuracy of the hybrid LDCT screening framework are observed more when compared to the existing approaches.

References

1. B.-Q. Li, J. You, L. Chen, J. Zhang, N. Zhang, H.-P. Li, T. Huang, X.-Y. Kong, Y.-D. Cai, Identification of lung-cancer-related genes with the shortest path approach in a protein-protein interaction network. **2013**, 8 (2013). Article ID 267375. https://doi.org/10.1155/2013/267375
2. G. Sutedja, New techniques for early detection of lung cancer. Eur. Respir. J. **21**, 57s–66s (2003). https://doi.org/10.1183/09031936.03.00405303
3. A. Tantraworasin, T. Suksomboonchroen, Y. Wannasopha, S. Kongkarnka, S. Saeteng, N. Lertprasertsuke, J. Euathrongchit, B. Chewaskulyong, Lung cancer: diagnosis and treatment approach. Int. Manual Oncol. Prac. 97–144. https://doi.org/10.1007/978-3-319-21683-6_7
4. S.K. Thakur, D.P. Singh, J. Choudhary, Lung cancer identification: a review on detection and classification. **39**(3), 989–998 (2020). https://doi.org/10.1007/s10555-020-09901-x
5. A.A. Judice, Dr. K.P. Geetha, A novel assessment of various bio-imaging methods for lung tumor detection and treatment by using 4-D and 2-D CT images. Int. J. Biomed. Sci. **9**(2), 54–60 (2013). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3708268/
6. Can Lung Cancer Be Found Early? https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/detection.html

7. K.M. Latimer, T.F. Mott, *Lung Cancer: Diagnosis, Treatment Principles, and Screening*. (Naval Hospital Pensacola, Pensacola, Florida). https://www.aafp.org/afp/2015/0215/p250.html
8. K. Senthil Kumar, K. Venkatalakshmi, K. Karthikeyan, Lung cancer detection using image segmentation by means of various evolutionary algorithms (2019). Article ID 4909846, https://doi.org/10.1155/2019/4909846
9. J.L. Mulshine, R.A. Smith, Lung cancer • 2: screening and early diagnosis of lung cancer. https://doi.org/10.1136/thorax.57.12.1071
10. C.I. Henschke, D.I. McCauley, D.F. Yankelevitz, D.P. Naidich, G. McGuinness, O.S. Miettinen, D.M. Libby, M.W. Pasmantier, J. Koizumi, N.K. Altorki, J.P. Smith, Early lung cancer action project: overall design and findings from baseline screening. Lancet **354**(9173), 99–105 (1999). https://doi.org/10.1016/S0140-6736(99)06093-6
11. C. Rampinelli, D. Origgi, M. Bellomi, Low-dose CT: technique, reading methods and image interpretation (2013). https://doi.org/10.1102/1470-7330.2012.0049
12. Low-dose CT scan for lung cancer screening, https://www.swedish.org/services/thoracic-surgery/our-services/lung-cancer-screening-program/low-dose-ct-scan-for-lung-cancer-screening
13. Y. Zhou, X. Xu, L. Song, C. Wang, J. Guo, Z. Yi, W. Li, The application of artificial intelligence and radiomics in lung cancer. Precision Clin. Med. **3**(3), 214–227 (2020).https://doi.org/10.1093/pcmedi/pbaa028
14. Y.W. Kim, B.S. Kwon, S.Y. Lim, Y.J. Lee, J.S. Park, Y.-J. Cho, H.I. Yoon, K.W. Lee, J.H. Lee, J.-H. Chung, E. Ji, C.-T. Lee, Lung cancer probability and clinical outcomes of baseline and new subsolid nodules detected on low-dose CT screening. https://doi.org/10.1136/thoraxjnl-2020-215107
15. H.J. de Koning, C.M. van der Aalst, P.A. de Jong, et al., Reduced lung-cancer mortality with volume CT screening in a randomized trial. N. Eng. J. Med. **382**, 503–13 (2020). https://doi.org/10.1056/NEJMoa1911793, pmid: http://www.ncbi.nlm.nih.gov/pubmed/31995683
16. S. Bharati, P. Podder, M.R.H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images. Inform. Med. Unlocked **20**, 100391 (2020). https://doi.org/10.1016/j.imu.2020.100391
17. T. Patel, V. Nayak, Hybrid approach for feature extraction of lung cancer detection. ICICCT (2018). https://doi.org/10.1109/ICICCT.2018.8473303
18. T. Meraj, H.T. Rauf, S. Zahoor, A. Hassan, M.I.U. Lali, L. Ali, S.A.C. Bukhari, U. Shoaib, Lung nodules detection using semantic segmentation and classification with optimal features. Neural Comput. Appl. https://doi.org/10.1007/s00521-020-04870-2
19. S.A. Khan, S. Hussain, S. Yang, K. Iqbal, Effective and reliable framework for lung nodules detection from CT scan images. Sci. Rep. **9** (2019). Article number: 4989. https://www.nature.com/articles/s41598-019-41510-9
20. W.B. Sentana, N. Jawas, A.E. Wardani, et al., Hybrid CPU and GPU computation to detect lung nodule in computed tomography images, in *2018 Third International Conference on Informatics and Computing (ICIC)*, August 2019. https://doi.org/10.1109/IAC.2018.8780573
21. Dr. P. Tumuluru, Dr. S. Hrushikesava Raju, CH.M.H. Sai Baba, S. Dorababu, B. Venkateswarlu, ECO friendly mask guide for corona prevention. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022047
22. CH.M.H. Sai Baba, Dr. S. Hrushikesava Raju, M.V.B.T. Santhi, S. Dorababu, Er. Saiyed Faiayaz waris, International currency translator using IoT for shopping international currency translator using IoT for shopping. IOP Conf. Ser. Mater. Sci. Eng. **981**, 4. https://doi.org/10.1088/1757-899X/981/4/042014
23. N. Sunanda, S. Hrushikesava Raju, S.F. Waris, A. Koulagaji, Smart instant charging of power banks. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022066
24. R. Mothukuri, Dr. S. Hrushikesava Raju, S. Dorababu, S.F. Waris, Smart catcher of weighted objects. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022002

25. M. Kavitha, Dr. S. Hrushikesava Raju, S.F. Waris, Dr. A. Koulagaji, Smart gas monitoring system for home and industries. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022003
26. Dr. S. Hrushikesava Raju, Dr. L.R. Burra, S.F. Waris, S. Kavitha, IoT as a health guide tool. IOP Conf. Ser. Mater. Sci. Eng. **981**, 4. https://doi.org/10.1088/1757-899X/981/4/042015
27. Dr. S. Hrushikesava Raju, Dr. L.R. Burra, Dr. A. Koujalagi, S.F. Waris, Tourism enhancer app: user-friendliness of a map with relevant features. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022067
28. M. Kavitha, S. Srinivasulu, K. Savitri, P.S. Afroze, P. Akhil, V. Sai, S. Asrith, Garbage bin monitoring and management system using GSM. Int. J. Innov. Technol. Explor. Eng. **8**(7), 2632–2636 (2019)
29. M. Kavitha, K. Anvesh, P. Arun Kumar, P. Sravani, IoT based home intrusion detection system. Int. J. Recent Technol. Eng. **7**(6), 694–698 (2019)
30. M. Kavitha, P.V. Krishna, V. Saritha, *Role of Imaging Modality in Premature Detection of Bosom Irregularity in Internet of Things and Personalized Healthcare Systems* (Springer, Singapore, 2019), pp. 81–92
31. M.K. Gould, Lung cancer screening in individuals with chronic obstructive pulmonary disease. Finding the sweet spot. Am. J. Respiratory Crit. Care Med. **192**(9). https://doi.org/10.1164/rccm.201508-1594ED
32. J.I.Z. Chen, S. Smys, Social multimedia security and suspicious activity detection in SDN using hybrid deep learning technique. J. Inform. Technol. **2**(2), 108–115 (2020). https://doi.org/10.36548/jitdw.2020.2.004
33. S. Smys, A. Basar, H. Wang, Artificial neural network based power management for smart street lighting systems. J. Artif. Intell. **2**(1), 42–52 (2020). https://doi.org/10.36548/jaicn.2020.1.005

An IoT Vision for Dietary Monitoring System and for Health Recommendations

S. Hrushikesava Raju, Sudi Sai Thrilok, Kallam Praneeth Sai Kumar Reddy, Gadde Karthikeya, and Muddamsetty Tanuj Kumar

Abstract Nutritional diet is a common factor for the well-being of all the living creatures and further bodily nourishments. It is important to consider all the statistics of the nutrients that are available in the food in order to deeply understand the facts that are affecting the person's health. Hence, the proposed methodology using IoT as a Dietary monitoring tool is developed to capture the food item and calculates the statistics that help to make a detailed report for the consumer and guide them with proper instructions. The mechanism describes nutrients that are highly recommended, the products that lead to obesity, and by looking at the analysis of the food, and we can accurately determine that the food nutritious and is useful to feed. This application which we proposed supports the sustainability of great health because health depends on feeding nutritious food. This method is more precise because it generates a very high-accurate report of nutrition intake of the person and guides the user to include the necessary nutritional values that are less in his previous dietary habits. Thus, a device is required to sense and provide them with all the statistical values related to their diet in the time period. The chew cycles and the nutrition-deficient food products are identified and can be understood in this approach. Also, this methodology is helpful as a guide that alerts the user with spontaneous analysis and provides future predictions regarding the person's health based on his dietary habits.

Keywords Sensors · Guide · IoT · Recommendations · Prediction · Status · And Analysis

1 Introduction

Food is a basic requirement. The amount of the nutrition required for a person is recommended by taking age, gender, and physical activity levels, and many more factors into consideration. The individual failing to satisfy these standard values for

S. Hrushikesava Raju (✉) · S. S. Thrilok · K. P. S. K. Reddy · G. Karthikeya · M. T. Kumar
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_65

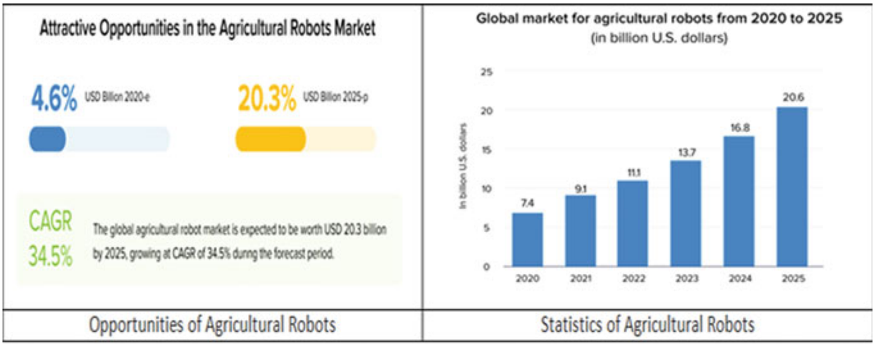


Fig. 1 Significance of agricultural robots and its impact

the nutrients would be considered be under-nutrition, and surplus consumption would end in over-nutrition, which in turn leads to varied sorts of health discrepancies. All in all, malnourishment will affect by rerouting tons of nation's economy toward the medical sectors. Malnutrition is one of the main attributes resulting in neonatal, infant, and maternal mortality diseases and many more. Time trends within the dietary intake pattern in reference to nutritional status over four decades are discussed. The affiliation between dietary insufficiency amid pregnancy and improvement interior the uterine seizing the crease and petite birth mass and their results required upon the dietary adjust in grown-up life are delineated from the data of national-level studies. The chapter has additionally brought among the incidence of fat and over-nutrition, chronic diseases like polygenic disorder, cardiovascular disease, vessel diseases, and cancer, their dietary association, and suggestions for managing these diseases.

IoT is outlined because the network of nodes that gather and convey knowledge via the net. Slowly, the food industry is getting aquatinted with the IoT. With degree of momentous applications of the Internet of Things the food creators, processors, and retailers are experiencing extraordinary openings. Apart from this, the list of points of interest advertised by IoT innovations inside the food industry is sort of long; subsequently, the affect to boot is amazingly noteworthy (Fig. 1).

The steps proposed in this system are:

(1) Extract the energy information from the food
(2) Suggest the health diet to do after consumption of the food each time.
(3) The three categories of result may be under diet, balanced diet, and over diet.

2 Literature Review

There exist certain studies which already performed in order to determine the statistics of the health because of fat, obesity, diabetic, and blood pressure, etc. In which, few works are listed.

With respect to the source specified in [1], the work describes the concept of diet-aware glasses is projected and enforced in a hardware prototype platform, which might offer semi-permanent dietary watching in associate degree retiring and cozy manner. The glasses compute the muscle activity of the temporal region and the micro-controller on-board performs an accurate intake measurement. Once operating beside a mobile phone, it will provide a well analyzed report on intake schedule, the amount of chewing cycles. In regard of description provided in [2], the work demonstrates on dietary monitoring of nutrient intake levels, hydration, food intake rate, and diet selections are all variables known to impact the hazard of unwanted weight gain. This method represents an innovative wearable device taking the shape of a necklace that accumulates information from an inserted piezoelectric sensor capable of capturing skin motion within the lower at the time of food intake. This model was perfectly designed and calibrated to measure every day cycles related to food consumption of an individual accurately.

In the description of [3–17], a precise adjustment of supplement admissions is checked on day-by-day. The module uses a free IoT platform to analyse data and store it with an accuracy of 98.6%. This module helps in dividing the input into frames and removes unwanted frames, then the CNN algorithm captures the food item using smart phone camera and calculates the quantity and nutritional facts of the food.

As a description given in [17], the various web frameworks are taken and are reviewed, and their performances are taken as comparison study. Regarding the demonstration of [18], the details of the video surveillance are monitored and marked colors at the progress bar in order to differentiate the peculiar activities. In the view of mentioned description in [19], the safe driving is guaranteed by alerting road conditions of a journey and security depends on the user activeness. With reference to [20], the BMI and WHR are estimated based on fast food consumption for a sample of 300 students from 2 countries and concluded abnormal obesity is resulted because of fast food habits. In the perspective of [21], the food habits would influence many parts of the body such as respiratory, digestive, cardiovascular system, nervous system, and etc., and concluded to control in eating the outside food or restaurant food. From the point of [22], Harvard and Chan proposed a set of policies and prevention approaches that avoid obesity as well as other health-related issues. As per [23], there were many factors that affect the thermal effect of food, which when reduced leads to insulin resistance, and a specific number of studies are taken into the study. With regard to [24], Harvard's health bulletin states that consuming high-quality food protects the brain and avoids stress while fast food would damage the brain cells. By the discussion from [25], the food that we ate would cause mental health and leads to a link between diet and mental ability. Hence, a proper diet would create good mental health. As per demonstration from [26], the reasons are extracted for diet, depression, and obesity. The consumption of high D-Vitamin food leads to lower depression where other foods lead to more depression. The mind's mood is determined by healthy food. With respect to [27], the foods habits also damage the health in terms of heart attack by high blood pressure and low blood pressure may weaken the heart. From the demonstration of [28], elderly people may suffer from many health issues because of food habits. To overcome and control, the body is equipped with sensors and operates it using a consumer home network and a set of samples is analyzed. With regard to [29], the automation of eyesight detection is done using IoT and equipped with appropriate sensors. An app is designed that reduces human efforts and cost.

In the rest of sources provided in other references describe about the reasons that make the person fatter, the food habits that could make imperfect health, some people may follow health guide tips if follow up is taken for them.

3 Proposed Solution

- Further inputs for peer to peer: Based on the inputs given by the patient, levels of his body's composition and vitamin levels are calculated, and corresponding food is suggested. Along with that we could also suggest the respective exercises that are beneficial to them in getting balanced and on track of health. Like, person with diabetes is suggested to go for a walk every day for minimum distance of 3 km.
- The second scenario: So, in our day to day lives, we see many females facing some dietary issues and all so in order to overcome this issue we need to propose food solution that helps in their tough situations not every time they could come to a clinic and the go for a checkup, so whenever there is a disturbance in the midnight hour, or any odd hours just enter the symptoms that are disturbing them so that they could get a proposed solution of food in order to dismiss the pain that is disturbing the personality.
- So, this will be e a great achievement which will be saving to lives in case of any massive disasters that could cause the fetus to move in the face of early dismissal. Common inputs: Like as to improve the immune rate during this pandemic we could also suggest some foods that boost immunity according to the effectivity of any virus causing diseases. Not everyone knows the right food to eat to raise the immunity and to maintain themselves in a balanced diet and healthy diet some fail and expire and lose hope.
- So, in order to achieve the maximum health and care we need to propose the foods that boost this immunity that is common to everyone based on the studies of medical reports in India. So this helps in fighting against something that is depreciating once health and all. So the second scenario is we find many old-aged people are falling as the victims of muscular pains and bone marrow related issues.
- So, in order to treat them do sudden levels and all we need to propose a strong diet for any supplementary e food that helps in regaining the strength of the bones. So, for achieving this there will be some reports of the patients who are the terms of this bone marrow issues so by examining the reports we could conclude over a particular region that the old-age people are lacking a certain percentage of strength in bones, so according to that the dosage of food will be decided and this will be proposed in a second column (Fig. 2).

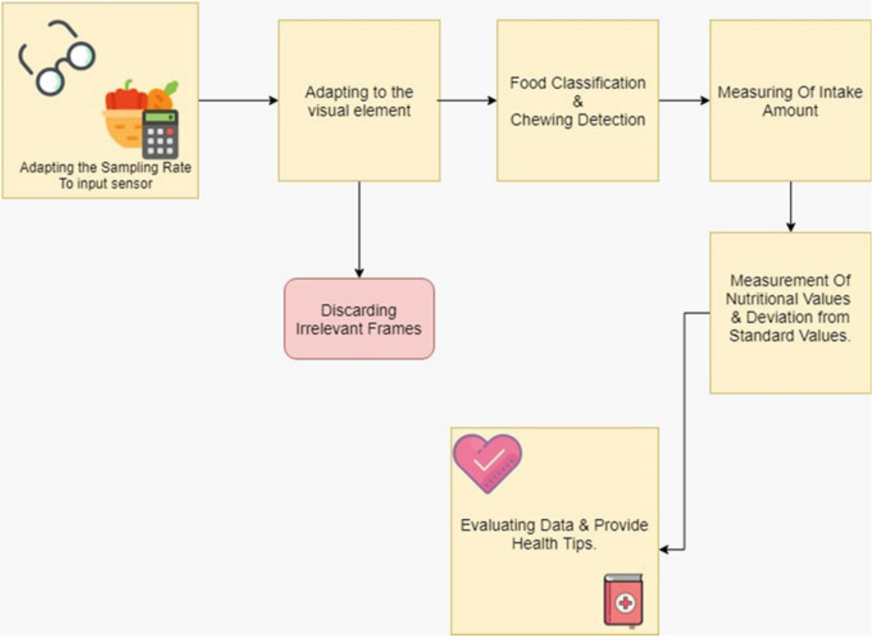
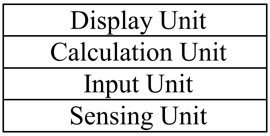

Adapting the Sampling Rate
To input sensor
Adapting to the visual element
Food Classification
&
Chewing Detection
Measuring Of Intake Amount
Discarding Irrelevant Frames
Measurement Of Nutritional Values & Deviation from Standard Values.
Evaluating Data & Provide Health Tips.

Fig. 2 Flow of activities in intended theme

- The scenario is we find many old-aged people are falling as the victims of muscular pains and bone marrow related issues. So, in order to treat to certain levels, we need to propose a strong diet for any supplementary food that helps in regaining the strength of the bones. So, for achieving this, there will be some reports of the patients who are the sufferers of this bone marrow issues. So, by examining the reports, we could conclude over that the old-aged people are lacking a certain percentage of strength in bones, so according to that the dosage of food will be decided and this will be displayed.
- Based on the inputs, levels of his body’s composition and vitamin levels are calculated and corresponding food is suggested. Along with that we could also suggest the respective exercises that are beneficial to them in getting balanced and on track of healthiness. Like, person with diabetes are suggested to go for a walk every day for minimum distance of 3 km (Fig. 3).

The above reinforced architecture is comprised of major layers that are stacked up from the base.

Fig. 3 Components in the proposed architecture



Display Unit
Calculation Unit
Input Unit
Sensing Unit

*Sensing unit allows the happening of sensing the food that is supposed to be examined for nutritional values.

*Input unit allows the user to input the values that are required.

*Calculation unit allows the calculation of deviation from the standard threshold nutritional values of food.

*Display unit shows the output like the deviation, measured values, standard values of a food.

These layers form the complete room for the architecture.

4 Results

Here, the goal is guaranteed using specific sensors that are used to read the food and health bulletin, as well as showing the health condition in a chart. In this, the specific sensors are used in order to carry out the theme. Also, the chart is defined in order to achieve the goal of this proposed system.

4.1 Selected Sensors

The followings are used in the system that was about to take up

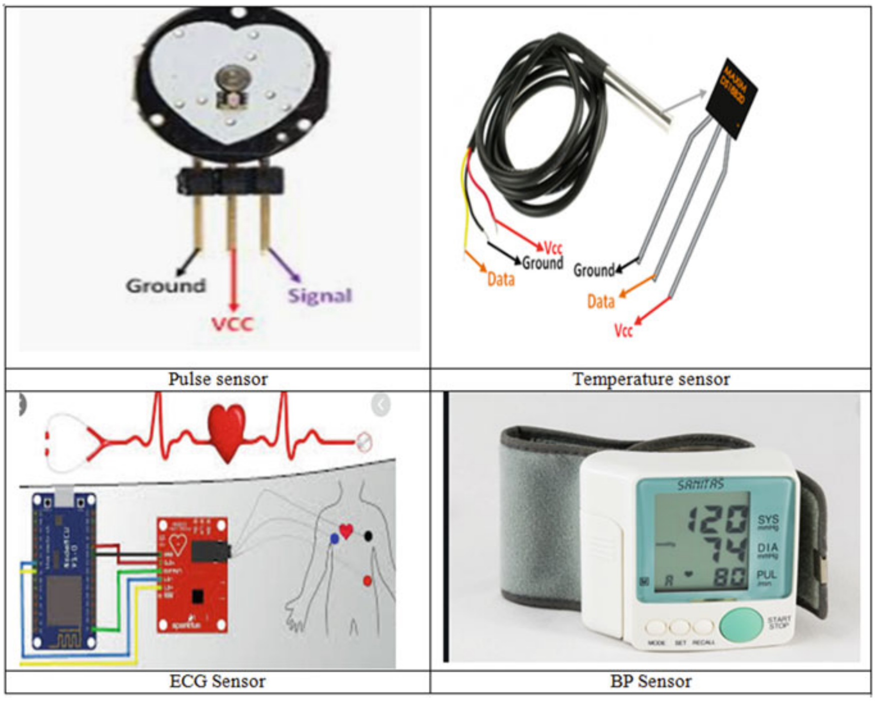
Pulse Sensor: It could be used to check the heart rate of the user. It can be used by any user.

(1) **Pulse Sensor:** It is used to read the heart pulse rate by keeping the chip of this sensor over a finger or earlobe. It is connected to the device though jumper cables and the data of reading information is projected in graphs using specific apps available in the market.

(2) **DS18B20 Temperature Sensor**: It is used to read temperature of an object. Here is to check up the food is fresh or not. The **DS18B20** is a programmable 1-wired temperature sensor from the embedded one. It can be used to detect temperature in environments like chemical solutions, mines. It can measure a wide range of temperature from −55 °C to + 125° with a decent accuracy of ± 5 °C.

(3) **ECG Sensor**: It helps to read electrical impulses from heart on the pathway and provides information according to the heart response during exercises or practice over physical muscle system. An ECG sensor can detect the arrhythmias, coronary heart disease, heart attacks, etc.

(4) **Blood pressure sensor (Sunrom-1437):** The main features of these sensors are easy to operate, switching button to start measuring, it can read all measures such as intelligent device debugging, automatic power to detect, power saving device automatically in 3 min (Fig. 4).



Pulse sensor

Temperature sensor

ECG Sensor

BP Sensor

Fig. 4 Kinds of sensors that are useful for defined system

4.2 Chart

The following is the statistics of composition of the food that is scanned before the user has to take up the food.

From Fig. 5, the various component statistics are extracted and converted those details in to the graph. This can be useful in estimating the energy provided details by the components of the food (Fig. 6).

The system has successfully taken the inputs of food and it returned the standard values of a food, deviation, measured values. The intake of foods in a log has been inputted and successfully calculated the necessity score. It is also displayed type of diet, types of food to be taken, types of exercises to be done.

The existing solution proposes a set of values that are measured when food is inputted like sweets, rice. The values like carbohydrates, fats, fiber, vitamins, etc. There is no additional configuration other than these.

To enhance the functionalities and efficiency, a solution has been proposed and it has some sort of additional configurations like possessing the requirements of diet and the deviation of the values from the standard values and taking the input of several foods that were last eaten by the end user and calculating the threshold and displaying the relevant diet type and the exercises and foods that are to be taken.

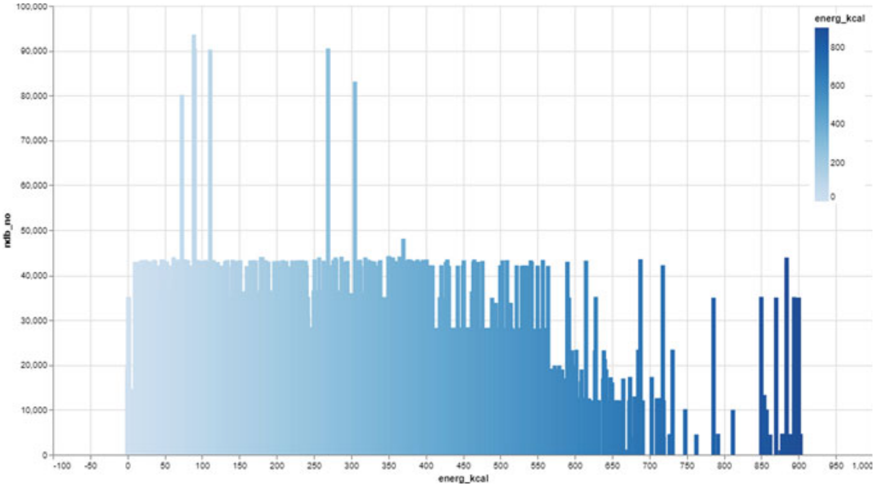


Fig. 5 Showing the statistics of variance of composition of food

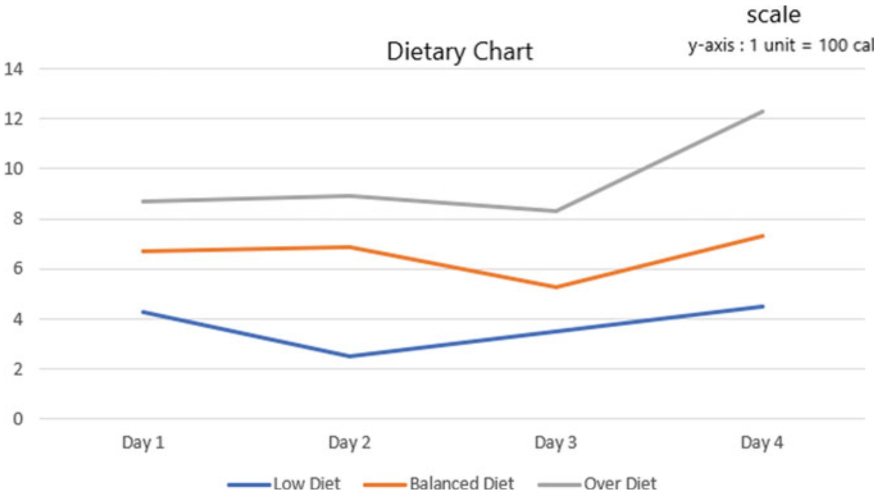


Fig. 6 Showing the dietary modes

5 Conclusion

In this, the person's health is to be predicted after consuming the food. The specific mining and machine learning techniques like classification and prediction approaches are used in order to judge the health bulletin. The Internet of Things plays a key role in determining the various components values using specific sensors. The modes categorized as under, balanced and over diet. These modes are visualized in the experimental setup and are suggested some tips as well as food habits if it is other than balanced diet. Hence, the user who considers this proposed system as a guide for good health maintenance. Further steps to follow to be came to know to avoid dangers to keep track of food statistics that may cause modes. In future, the estimation and accuracy are made better using hybrid approaches or novel techniques in predicting the health status and prescribing the recommendation.

References

1. Q. Huang, W. Wang, Q. Zhang, Your glasses know your diet: dietary monitoring using electromyography sensors. IEEE Internet Things J. **4**(3) (2017)
2. H. Kalantarian, N. Alshurafa, M. Pourhomayoun, S. Sarin, B. Shahbazi, Recognition of nutrition intake. (February, 2015). https://doi.org/10.1109/JSEN.2015.2402652
3. S. Zhang, D.T. Nguyen, G. Zhang, R. Xu, N. Maglaveras, N. Alshurafa,, Habits necklace: a neck worn sensor that captures eating related behavior and more. (2018), pp. 484–487
4. P. Tumuluru, S. Hrushikesava Raju, CH.M.H. Sai Baba, S. Dorababu, B. Venkateswarlu, in *ECO Friendly Mask Guide for Corona Prevention IOP Conference Series Materials Science and Engineering* vol. 981(2) (2020), pp. 10.1. 88/1757–899X/981/2/022047
5. CH.M.H. Sai baba, S. Hrushikesava Raju, M.V.B.T. Santhi, S. Dorababu, Er. Saiyed Faiayaz waris, International currency translator using IoT for shopping international currency translator using IoT for shopping. in *IOP Conference Series Materials Science and Engineering*, vol 981 (2020), pp. 4. https://doi.org/10.1088/1757-899X/981/4/042014
6. N.Sunanda, S. Hrushikesava Raju, S.F. Waris, A. Koulagaji, Smart instant charging of power banks smart instant charging of power banks. in *IOP Conference Series Materials Science and Engineering*, vol 981 (2020). pp. 210. https://doi.org/10.1088/1757-899X/981/2/022066
7. R. Mothukuri, S. Hrushikesava Raju, S. Dorababu, S.F. Waris, Smart catcher of weighted objects smart catcher of weighted objects. in *IOP Conference Series Materials Science and Engineering*, vol 981 (20002), pp. 2. https://doi.org/10.1088/1757-899X/981/2/022002
8. M. Kavitha, S. Hrushikesava Raju, S.F. Waris, A. Koulagaji, Smart gas monitoring system for home and industries smart gas monitoring system for home and industries. in *IOP Conference Series Materials Science and Engineering*. vol 981 (2003), pp. 2. https://doi.org/10.1088/1757-899X/981/2/022003
9. S. Hrushikesava Raju, L.R. Burra, S. Faiayaz Waris, S. Kavitha, IoT as a health guide tool. in *IOP Conference Series, Materials Science and Engineering*. vol 981 (2015), pp. 4. https://doi.org/10.1088/1757-899X/981/4/042015
10. S. Hrushikesava Raju, L.R. Burra, A. Koujalagi, S.F. Waris, Tourism enhancer app: user-friendliness of a map with relevant features. in *IOP Conference Series, Materials Science and Engineering*. Vol 981 (2020), pp. 2. https://doi.org/10.1088/1757-899X/981/2/022067
11. M. Kavitha, S. Srinivasulu, K. Savitri, P.S. Afroze, P. Akhil, V. Sai, S. Asrith, Garbage bin monitoring and management system using gsm, international journal of innovative technology and exploring. Engineering **8**(7), 2632–2636 (2019)
12. M. Kavitha, K. Anvesh, P. Arun Kumar, P. Sravani, IoT based home intrusion detection system. Int. J. Recent Technol. Eng. **7**(6), 694–698 (2019)
13. M. Kavitha, P.V. Krishna, V. Saritha, Role of imaging modality in premature detection of bosom irregularity in internet of things and personalized healthcare systems, (Springer, Singapore, 2019), pp. 81–92
14. S. Anjali Devi, S. Siva Kumar, A hybrid document features extraction with clustering based classification framework on large document sets. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **11**(7) (2020). https://doi.org/10.14569/IJACSA.2020.0110748.

15. S. Anjali Devi, M.B. Harika, I. Harshitha, K. Sarath Sai, Big data set privacy preserving in banking using machine learning techniques. J. Adv. Res. Dynam. Control Syst. **11**(7), 609–620 (2019)
16. V. Rachapudi, S. Venkata Suryanarayana, T. Subha Mastan Rao, Auto-encoder based K-means clustering algorithm. Int. J. Innov. Technol. Explor. Eng. **8**(5), 1223–1226 (2019)
17. V. Rachapudi, N. Nitish, S. Samaikya, U.P. Sathvik, S.A. Devi, Performance comparison of applications with and without web frameworks,. Int. J. Adv. Trends Comput. Sci. Eng. **9**, 1020–1028 (2020). https://doi.org/10.30534/ijatcse/2020/19922020
18. S. Hrushikesava Raju, M. Nagabhushana Rao, N. Sudheer, P. Kavitharani, quick identification of specific activity by processing of large-size videos using advanced spotter. Int. J. Eng. Technol.(UAE) ISSN: 2227–524X. https://doi.org/10.14419/ijet.v7i2.32.15712
19. S. Hrushikesava Raju, M. Nagabhushana Rao, N. Sudheer, P. Kavitharani, Visual safe road travel app over google maps about the traffic and external conditions. Int. J. Eng. Technol.(UAE) (2018). ISSN: 2227–524X. https://doi.org/10.14419/ijet.v7i2.32.15697
20. A. Mohammadbeigi, A. Asgarian, E. Moshir, H. Heidari, S. Afrashteh, S. Khazaei, H. Ansari, Food consumption and overweight prevalence in students and its association with general and abdominal obesity, **59**(3), E236-E240 (2018)
21. N. Butler, RD, LD, The effects of fast food on the body (Sep 2018). https://www.healthline.com/health/fast-food-effects-on-body
22. Harvard TH Chan, Obesity Prevention Source. https://www.hsph.harvard.edu/obesityprevention-source/obesity-causes/diet-andweight/
23. L. de Jonge, G.A. Bray, The thermic effect of the food and obesity: a critical review. **5**(6), 622–631 (1997)
24. M.D. Eva Selhub, Nutritional psychiatry: your brain on food (March 2020). https://www.health.harvard.edu/blog/nutritional-psychiatry-your-brain-onfood-201511168626
25. Diet and Mental Health (October, 2018). https://www.mentalhealth.org.uk/a-to-z/d/diet-and-mentalhealth
26. Food and mood: Is there a connection? (June 2018). https://www.health.harvard.edu/mind-andmood/food-and-mood-is-there-a-connection
27. S. Frothingham, How Does Eating Affect Your Blood Pressure? (August 2019). https://www.healthline.com/health/bloodpressure-after-eating
28. J. Wang, Z. Zhang, B. Li, R. Sherratt, S. Lee, An enhanced fall detection system. Electron. **60**(1), 23–29 (2014). https://doi.org/10.1109/TCE.2014.6780921
29. S. Hrushikesava Raju, L.R. Burra, S.F. Waris, S. Kavitha, S. Dorababu, Smart eye testing, advances in intelligent systems and computing, 2021. in *ISCDA 2020, 1312 AISC*, (2021), pp. 173–181. https://doi.org/10.1007/978-981-33-6176-8_19

Channel Estimation in MIMO-OFDM and PAPR Reduction

A. Gokul, J. N. Sarath, M. Mohit, M. Niranjan, and Aswathy K. Nair

Abstract Multiple Input Multiple Output (MIMO) Orthogonal Frequency Division Multiplexing (OFDM) is a prominent scheme in wireless communication mainly in 5G communications. The signals when transmitted through a wireless channel undergo fading and interference due to multipath induced by reflectors and scatterers present in the surroundings thereby difficult to recover. Proper equalization techniques need to be adopted at the receiver side to recover the signal, prior to that, channel estimation is required to understand the behavior of the channel. We present a Compressed Sensing based pilot assisted channel estimation method over a Rayleigh fading channel and the performance is compared with Least Square (LS) and Minimum Mean Square Error (MMSE). The performance is compared with channel response characteristics with varying pilots and pilot positions and based on the results, number of pilots was optimized. The simulation results show the effectiveness of the method adopted with varying numbers of pilots. The work also addresses PAPR reduction technique. Three PAPR reduction techniques have been applied and compared, viz. Clipping-Filtering, Selective Mapping, and Partial Transmit Sequence (PTS).

Keywords MIMO · OFDM · Channel estimation · Compressed sensing · PAPR · PTS

A. Gokul · J. N. Sarath (✉) · M. Mohit · M. Niranjan · A. K. Nair
Department of Electronics and Communication Engineering, Amrita Vishwa Vidhyapeetham, Amritapuri, India
e-mail: sarathjnair@am.students.amrita.edu

A. Gokul
e-mail: gokulajith@am.students.amrita.edu

M. Mohit
e-mail: mohitmnoja@am.students.amrita.edu

M. Niranjan
e-mail: niranjanm@am.students.amrita.edu

A. K. Nair
e-mail: aswathykn@am.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_66

1 Introduction

OFDM has gained rising popularity in our modern world, in 4G-5G LTE wireless communications. A shortcoming that accompanies the same, is the interference caused due to various multipath components resulting in signal distortion, signal fading, and ultimately limited coverage. This has paved the way for the rise in popularity of MIMO-OFDM systems which utilize multiple antennas having advantages such as reduction in fading and increase in capacity of the system. Without increasing the bandwidth and the transmit power, MIMO systems with multiple antennas help in increasing the average receive SNR and also enhances the range and coverage of the network. Channel estimation is one of the major steps in any wireless transmission technique. The reconstruction of the transmitted signal depends on the accuracy of the channel being estimated. It also compensates for the distortions introduced when the symbols propagate through the channel, taking into account the SNR. The receiver equalizer must unwrap the incoming symbol back to its expected shape to accurately decode it and in two-way communication, channel estimation allows best constellation to be negotiated.

The basic OFDM system is shown in Fig. 1. The data stream is a digital modulated bitstream modulated with 16-QAM. The input bitstream is given to Serial to Parallel Converter (S/P) where the series of bitstreams is divided into small bitstreams. Then the pilot bits (guard bits) are added to the stream via block or comb type pilot insertion method. In comb type pilot insertion technique, the pilot symbols are added in subcarriers periodically and in block type insertion technique, the pilots are inserted into every subcarrier of a block. The IFFT operation of the samples generates Multi-Carrier Modulated (MCM) signals without the use of modulators. The IFFT symbols are added with cyclic prefix, and the symbols are again converted to serial symbols. The addition of cyclic prefix prevents Inter Block Interference (IBI), Inter Carrier Interference (ICI), and Inter Symbol Interference (ISI) [1]. The signal proceeds to be transmitted via the channel where it encounters different fading channel characteristics. Inter Symbol interference can be avoided if the length of channel taps ‘L’ is less

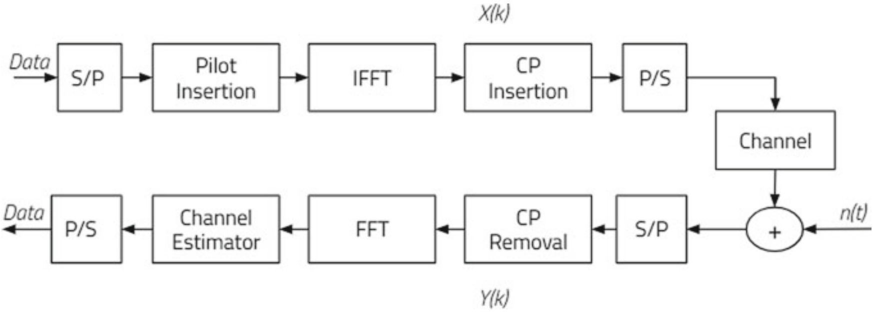


Fig. 1 Basic OFDM system

than the cyclic prefix's length. Signals received at the receiver get added with (Additive White Gaussian Noise) AWGN noise. Cyclic Prefix is removed from the output of Serial to Parallel Converter (S/P), which interferes with the multipath channel. FFT or demodulation of the cyclic prefix removed data block is performed and given to the channel estimation block. Channel estimation is done using different techniques namely Least Square estimation, Minimum Mean Square Error (MMSE), and Compress Sensing, these techniques are discussed in this project. In MIMO systems, more than one receiver and transmitter antenna are used simultaneously. OFDM technique when assisted with MIMO, makes it a promising method to achieve better data rate and capacity. The OFDM modulated data are given to each of the transmitters and the signals received by each receiver antenna are passed through OFDM demodulators. With this kind of implementation, signal loss can be considerably minimized.

In an OFDM system due to the addition of various subcarrier components through an IFFT operation, signals thus transmitted tend to have high peak values. Consequently, Peak-to-Average Power Ratio (PAPR) has become a distinguishing characteristic of OFDM systems with regards to the combination of multiple QAM symbols. The high PAPR will result in the non-linearity of the power amplifiers in the transceivers which in turn results in the non-orthogonality of the system and causes distortion and ICI. PAPR is calculated by obtaining the ratio of peak amplitude square and average value square. It is proportional to the total quantity of subcarriers the system is utilizing. The increasing number of subcarriers could increase the data rate of the system and hence increase the PAPR. The PAPR distribution can be statistically defined using its complementary cumulative distribution function (CCDF). Some methods that accomplish this are clipping, selective mapping, and partially transmit sequence.

2 Related Works

Gong et. al., in [2] introduces a new comb type channel estimation scheme, which considers the property of time variation in different symbols. Doppler spread is high due to the increased speeds of the railway. The results of this simulation elaborate that the system outperforms traditional interpolation methods. This work adopts the iteration and decomposition to detect a signal that results in the target matrix dimensions being diminished Obtained results depict that the proposed detector achieves a superior accuracy with respect to the existing signal detection schemes.

A SISO system [3] in which there is only a single transmitting and receiving antenna. In such a system Linear MMSE gives best performance. Because of matrix inversion and channel correlation. Its complexity is very high. When the delay spread is more than the cyclic prefix ISI and ICI will be introduced in the transmitted signal thus proving that when there is an increase in channel length than cyclic prefix, BER value becomes worse.

The main objective of the paper [4] is a comparison of Least Square and Compressed Sensing. This paper also presents a channel estimation technique that incorporates total Variation and Augmented Lagrangian (TVAL3) method to do channel estimation in Compressed sensing. TVAL3 is a compressed Sensing Reconstruction Algorithm for the reconstruction of $F(m, i)$. In the final section comparison of performance of Least Square Estimation and Compressed Sensing from the plot of (MSE) vs (SNR) for multiple pilots. The TVAL3 based Compressed Sensing showed better performance when a smaller number of pilots than Least Square estimation.

Wang and Luo in [5] proposes a novel filtering technique based on convex optimizations that minimize the error vector magnitude of the symbols in each iteration, significantly reduces the iterations required to reach the desired PAPR value, and have comparatively less distortion along with better out-of-band spreading when compared to the classical ICF technique.

Pilot aided channel estimation techniques Least Square (LS) channel estimation employing frequency domain methods are used to get an initial estimate of the channel in Aghdam and 6. Through a time, domain method, recovered symbols are used to improve the channel estimation. Computer simulations are used to demonstrate the proposed estimator's performance under various channel conditions. In MIMO-OFDM systems, there are two types of pilot placements. Pilots are introduced into a group of subcarriers in an OFDM symbol using Comb type pilot in MIMO-OFDM system, and the channel is predicted via interpolation. It's useful when the channel is fast fading.

Dai et al. in the paper [7] considered a CoSaMP algorithm for fast varying time domain synchronous (TDS) channel and considered three transmission schemes viz. CP, Zero padding and TDS-OFDM. The authors in [8] implemented super imposed pilot based channel estimation in MIMO in USRP experimental setup. In Mowla et al. [9] authors used a composed filtering method after clipping to reduce PAPR. D. Rajeswaran et al. proposed a novel method to reduce the PAPR via. ICA-PCA combination and further clipped to improve the performance [10]. Sam and Nair [11] blind channel and training based channel model in an OFDM system with different types of pilot placements are used to predict the channel via interpolation schemes.

3 Method and Approaches

A random bitstream of size 'n' is generated and is fed to the 16-QAM modulator. The modulated symbols are then given to OFDM modulators. In an OFDM based system, entire bandwidth available is segregated into 'N' number of sub-bands forming an MCM system. The data stream in the OFDM modulator is divided into 'B' blocks. For further calculations, only one block of data is taken into account and the computations will be the same for all other blocks. Using the comb type pilot insertion technique [12], the pilots are added to the data stream. Subcarriers have pilots inserted at certain intervals for each data block. The comb type insertion is written as:

$$X(k) = X(mN + l) = \begin{cases} x_p(m), l = 0 \\ \text{data}, l = 1, \ldots, N - 1 \end{cases} \quad (1)$$

where $x_p(m)$ is the pilot value, N represents the quantity of carriers while $X(mN + l)$ is the position at which pilot symbols are inserted. The N-point IFFT of the pilot inserted data block is taken. The N-point iift of the pilot inserted data provides a $N \times 1$ IFFT block of symbols. Cyclic prefixes (CP) are added to each block after the IFFT operation. The cyclic prefix is always larger than the CIR, 'L' of an OFDM system, to avoid ISI. The prefixed data of size $((N + \text{length(CP)}) \times 1)$ are then reshaped to form a single bitstream for transmission. The signal is transmitted through a wireless channel where it gets affected by different multipath components. The transmitted signals in the multipath channel undergo convolution with multipath channel coefficient, this received signal at receiver is added with AWGN noise of dimension equal to the transmitted signal, $((N + \text{length(CP)}) \times 1)$. The CP added at the transmitter block is removed to neglect the effects of ISI. CP removed data block a $N \times 1$ is fed to the FFT block. A 2×2 MIMO system is considered, i.e., there are 'Tx' $= 2$ transmitter antennas and quantity of receiver antennas is represented by 'Rx' $= 2$. Diversity technique is being used for transmission which helps in reducing the chance of signal fading and increases the capacity. The received signals at the two antennas can be represented as,

$$\begin{aligned} r_1 &= h_{11}x_1 + h_{12}x_2 \\ r_2 &= h_{21}x_1 + h_{22}x_2 \end{aligned} \quad (2)$$

where x_1 and x_2 are the signals transmitted by the transmitter antennas, h_{ij}, $i,j \in$ (1, 2) are the CIR and r_1 and r_2 are the signals received at the two receiver antennas.

3.1 Channel Estimation

This paper compares the LS, MMSE, and Compressed Sensing (CS) method for estimating the channel. The channel may be modeled as a time varying filter, the filtering nature is because of the delays in the multiple signals that are arriving and also the summation of amplitudes at any point in time. Each wireless channel can be characterized by delay (δ) and attenuation (α). The Channel Impulse Response (CIR) [7] of an 'L' tap channel with 's' propagation path is modeled as,

$$h_{i,n} = \sum_{l=0}^{S_i - 1} \alpha_{i,l} \delta\big[n - \tau_{i,l}\big], 0 \le n \le L - 1 \quad (3)$$

The pilot signals present in the transmitted signals are used for estimating the channel. The received pilot OFDM signal is represented as,

$$Y_P = X_P H + Z \tag{4}$$

where the received pilot symbol is Y_P, X_P depicts the transmitted pilot values, H resembles the channel response and Z is the AWGN noise.

3.1.1 LS Channel Estimation

In MIMO-OFDM system, the Least Square (LS) estimated CIR for 'Rx' th receiver can be expressed as:

$$\widehat{H_{\text{LS}}^{(Rx)}} = \left(X_P^{(Tx)}\right)^{-1} Y_P^{(Rx)} \tag{5}$$

where $\widehat{H_{\text{LS}}^{(Rx)}}$ is the linearly interpolated CIR, $Y_P = \left[Y(0), Y(1), ..., Y\left(N_p - 1\right)\right]^T$ is the received pilot sequence and $X_P = \text{diag}\left\{X(0), X(1), ..., X\left(N_p - 1\right)\right\}$ is the transmitted pilot values from each transmitter antenna. The estimated channel response is only for the subcarriers containing pilot carrying symbols, to estimate channel response for given data carriers linear interpolation technique is performed. The Linear Interpolation equation is represented as,

$$I_k(V) = I_k + (I_{k+1} - I_k)(V - V_k)/(V_{k+1} - V_k) \tag{6}$$

where I_k is the estimated channel coefficient at kth subcarrier, V_k is the index of the kth subcarrier and V is the current subcarrier index.

3.1.2 MMSE Channel Estimation

The MMSE channel estimation in MIMO-OFDM for 'Rx' th receiver can be expressed as:

$$\widehat{H_{\text{MMSE}}^{(Rx)}} = R_{\text{HH}} / \left[R_{\text{HH}} + \frac{\beta}{\sigma^2} I_N\right] * \widehat{H_{\text{LS}}^{(Rx)}} \tag{7}$$

where $\widehat{H_{\text{MMSE}}^{(Rx)}}$ is the MMSE estimated channel response, R_{HH} is the frequency domain auto-correlation matrix of the channel, $\beta = E\left(X_k^2\right)/E\left(1/X_k^2\right)$, the transmitted symbol in the frequency domain is X, the variance of AWGN is depicted by σ^2 and I_N is $N \times N$ identity matrix.

3.1.3 Compressed Sensing

The compressed sensing method can be adopted for sparse channel estimation. This technique is prominently used for image reconstruction from its sparse signals which were later adopted for MIMO channel estimation because of the sparsity nature of the multipath channel. The pilot overhead induced in MIMO-OFDM system can be greatly reduced by the inclusion of CS based CE. A comb pilot insertion method with the CS technique-Orthogonal Matching Pursuit (OMP) method is used in the work. OMP is a less complex greedy algorithm used for sparse channel estimation [13]. In this method, the received signal is taken as:

$$Y_P^{\mathrm{CS}} = X_P^{\mathrm{CS}} F_P^{\mathrm{CS}} h + Z_P^{\mathrm{CS}} \tag{8}$$

where Y_P^{CS} is the $N_P \times 1$ received pilot symbols, X_P^{CS} is the $N_P \times N_P$ diagonal matrix of the transmitted pilot symbols, F_P^{CS} is the $N_P \times N_P$ matrix with rows corresponding to the pilot positions are taken from the DFT matrix F and Z_P^{CS} is the $N_P \times 1$ AWGN. h is the channel taps in the time domain. The estimation of the sparse channel h is obtained by convex optimization method represented as follows:

$$\min ||h||_1 \mathrm{s.t} \left|\left| Y_P^{\mathrm{CS}} - \varphi h \right|\right|_2 \le \varepsilon \tag{9}$$

where $||h||_1$ is the 1-norm of the channel, the measurement matrix φ=$X_P^{\mathrm{CS}} F_P^{\mathrm{CS}}$ and ε is the threshold. OMP is a less complex technique used to reconstruct h from the Y_P^{CS} and using the measuring matrix φ. For generating the reconstruction matrix M, the columns of the measurement matrix having largest correlation with Y_P^{CS} are appended to the empty matrix M in the first iteration. After augmenting, the channel is calculated as:

$$\widehat{h_{\mathrm{OMP}_i}} = \left(M_i^H M_i\right)^{-1} M_i^H Y_P^{\mathrm{CS}} \tag{10}$$

For the next iterations, the maximum correlated column of the residue matrix is found and augmented with M. The residue is measured as:

$$R_i = Y_P^{\mathrm{CS}} - M\widehat{h_{\mathrm{OMP}_i}} \tag{11}$$

The iteration takes place until $||R_i - R_{i-1}|| \le \varepsilon$, where ε is the threshold, R_i is the residue of the ith iteration and R_{i-1} is the residue obtained in $i - 1$th iteration. The time domain impulse response is evaluated using spline interpolation method.

4 PAPR Reduction in OFDM System

In an OFDM system due to the addition of various subcarrier components through an Inverse Fast Fourier Transformation (IFFT) operation, transmitted signals tend to have high peak values. Consequently, Peak-to-Average Power Ratio (PAPR) has become a distinguishing characteristic of OFDM systems with regards to the combination of multiple QAM symbols. Owing to the central limit theorem, very high PAPR often accompanies output OFDM symbols. The high PAPR arises because of the IFFT operation done in the system. The major concern raised in an OFDM system is the information loss sustained due to the high PAPR in signals, this occurs when the signal is converted to time domain information at majority of the peaks does not get included and are ultimately lost. Thus, it is imperative to ensure maximum PAPR reduction in the signal. Some signal distorting and signal scattering techniques are used.

4.1 Selective Mapping

It reduces PAPR efficiently while not requiring an increase in power requirement with minimal data rate loss. A full set consisting of candidate signals is generated that carry the same information and signal with the least PAPR as can be seen from Fig. 2. One of the major advantages selective mapping brings to the table is that the peaks will not be eliminated and there is no cap on the number of subcarriers that can be used. However, using this method [14] means it is essential to have an overhead of side information which will be obtained at the receiver to ensure that sent information is recovered.

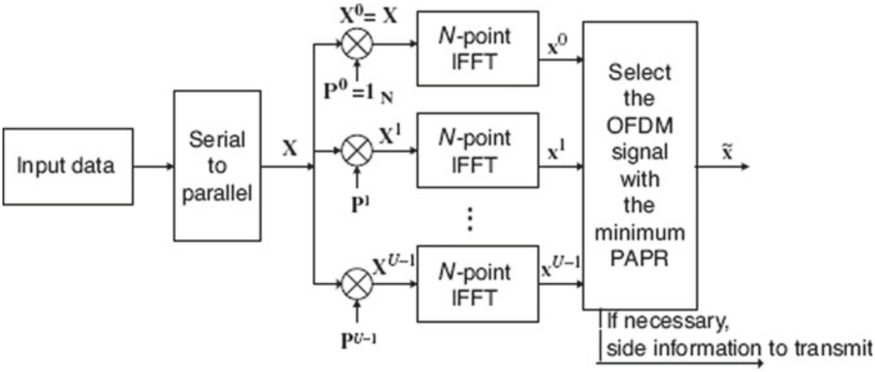


Fig. 2 Selective mapping block diagram

4.2 Partial Transmit Sequence

It is known to provide exceptional PAPR reduction and has a high complexity owing to the nature of its extensive random search through all possible combinations of allowed phase vectors. The technique in [14] is considered to be flexible and effective as it works on the base principle that the entire input data is divided into sub blocks that are non-overlapping, phase shifting each by a constant factor to maintain reduction of PAPR.

$$x = \sum_{m=1}^{M} b_{\mathrm{m}} x_{\mathrm{m}} = \sum_{m=1}^{M} b_{\mathrm{m}} IFFT\{X_{\mathrm{m}}\} = \sum_{m=1}^{M} \text{ƀ}_{\mathrm{m}} x_{\mathrm{m}} \tag{12}$$

$b_{\mathrm{m}} = exp(j\phi_{\mathrm{m}}),\ \phi_{\mathrm{m}}$is the phase factor
X_m is the orthogonal sub blocks

$$[\text{ƀ}_1\ \text{ƀ}_2\ \text{ƀ}_3 \ldots .\text{ƀ}_{\mathrm{m}}] = arg(min(max|\textstyle\sum_{m=1}^{M} b_{\mathrm{m}} x_{\mathrm{m}}|)) \tag{13}$$

Thus, the signal in the time domain with the lowest PAPR will be

$$x = \sum_{m=1}^{M} \text{ƀ}_{\mathrm{m}} x_{\mathrm{m}} \tag{14}$$

4.3 Clipping and Filtering

Clipping and Filtering is a Signal Distortion suppression technique. The root mean square of the signal is taken as the threshold value for clipping. This causes the signal spectrum to spread and thereby increases interference and causes degradation in BER. Passing the signal through a filter helps to reduce out-of-band radiation. Filtering also helps in reducing the PAPR without signal expansion. The clipped signal is convolved with a low pass filter [5].

$$x_k^c = \begin{cases} x_k\ ,\ |x_k| \le A \\ A\ ,\ |x_k| > A \end{cases} \quad 0 \le k \le N-1 \tag{15}$$

where x_k^c is the clipped signal, x_k is the transmitted signal, A is the threshold amplitude. This method is effective in reducing PAPR without spectrum expansion.

5 Results and Analysis

5.1 Channel Estimation

For the simulation of MIMO-OFDM, 16-QAM is used with $N = 1024$ subcarriers in a Rayleigh fading channel. A channel tap of $L = 120$ is used in a rich scattering environment and the number of pilot sets used was 34. For comparing the performance of CS with LS and MMSE varying numbers of pilots were given and the same is decremented up to 9 pilot positions. OMP algorithm is used for CS method and the SNR values varied between 5 and 40 dB, A monte-carlo simulation of 500 runs was performed for each SNR. In this work, LS and MMSE based CE is performed for a 2×2 MIMO spatial modulation scheme using linear and Spline interpolation method. Though both interpolation methods gave satisfactory results in estimating the channel response, spline interpolation gave more accurate channel estimation result compared to linear interpolation method. From the LS estimated channel response of the two receiver antennas Fig. 3, it is evident that the LS- has distortions and does not accurately match with the actual channel response. Envelope of the estimated channel is having prominent distortion when peak values are occurring.

Figures 4 and 5 shows the MMSE estimated channel responses for the two antennas. The MMSE estimated channel for the two receiver antennas shows that the estimated channel envelope traces the same pattern as that of the actual channel response with very minimum distortions, but the values do not match with the actual channel response.

The CS estimated results for the two receiver antennas in all the subcarriers as in Figs. 6 and 7 suggested that the estimated channel is very much comparable with the actual channel response. The distortions in the estimated channel are minimal.

The results of the channel estimation are for 33 pilot carriers. The channel response is estimated using varying pilot numbers and positions. It was observed that when the pilot count is significantly reduced to 9, the estimation with CS technique is expected

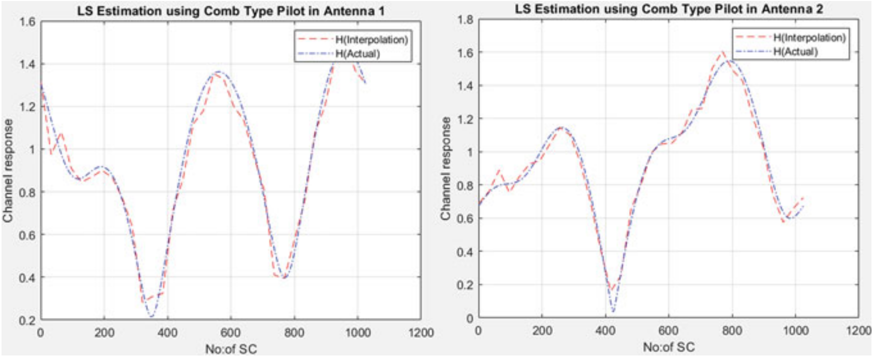


Fig. 3 LS estimated channel response for receiver Antenna 1 and Antenna 2

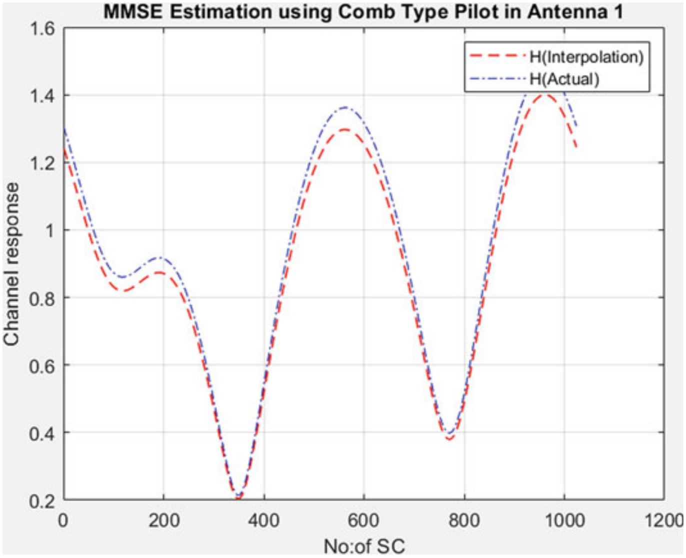


Fig. 4 MMSE estimated channel response for receiver Antenna 1

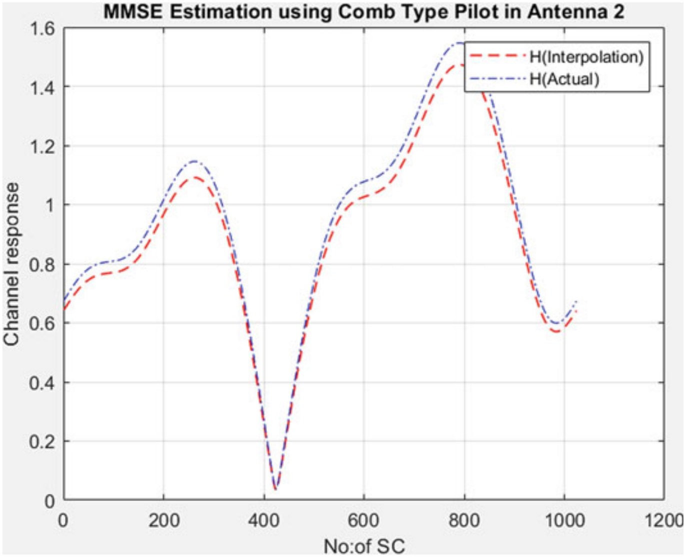


Fig. 5 MMSE estimated channel response for receiver Antenna 2

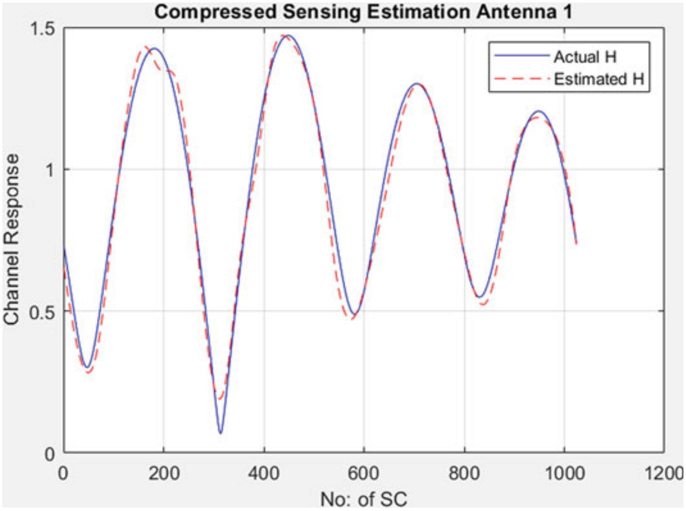


Fig. 6 CS estimated channel response for receiver Antenna 1

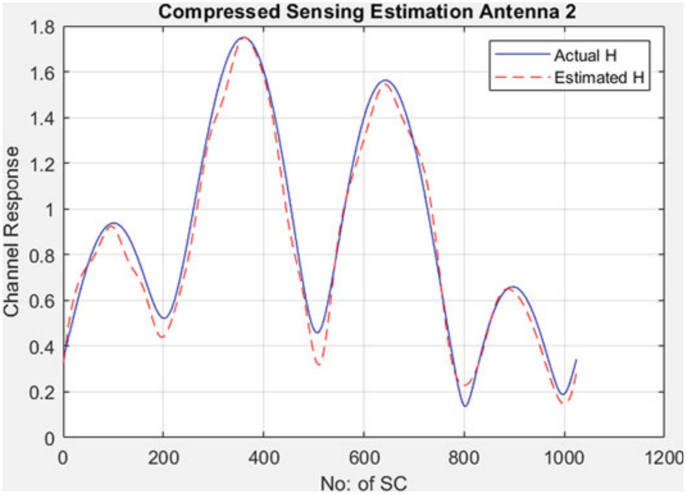


Fig. 7 CS estimated channel response for receiver Antenna 2

to perform better since it follows the concept of sparsity whereas the performance of the LS and MMSE technique depends on increased count of pilots.

Figures 8 and 9 shows the channel estimation with 9 pilot symbols for CS and LS schemes. Hence, we can infer that the performance of LS and MMSE degrades significantly compared to CS. The reduced number of pilots increases the spectral efficiency and reduces the pilot overhead induced by MIMO.

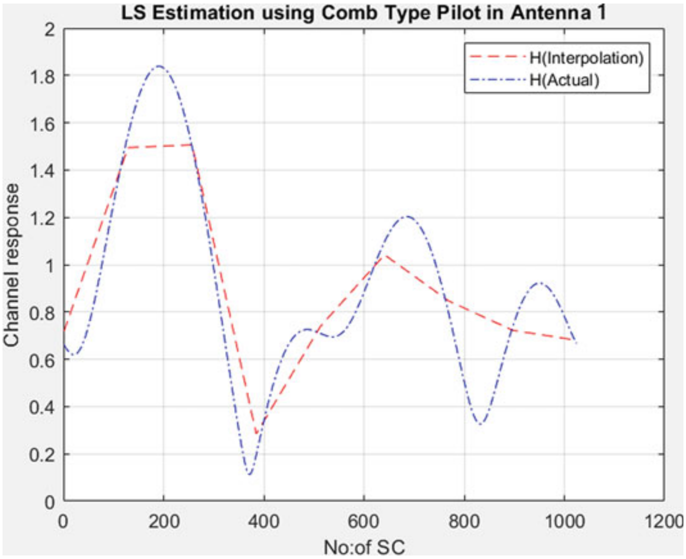


Fig. 8 LS estimation with 9 pilots

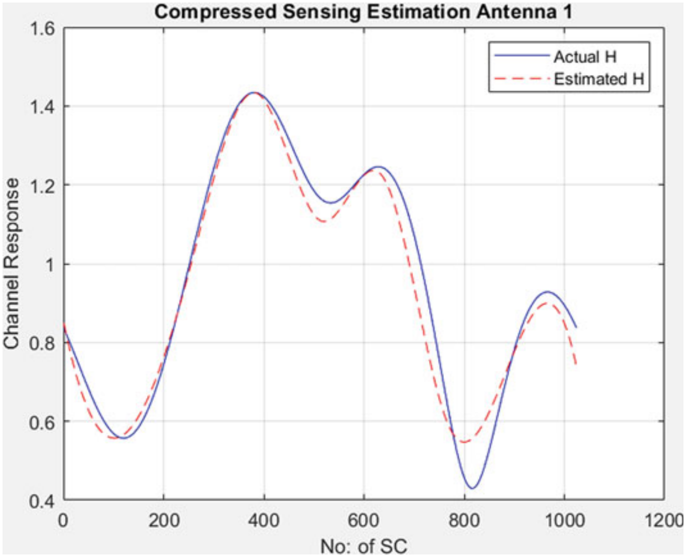


Fig. 9 CS estimation with 9 pilots

5.2 PAPR Reduction

The following section describes the evaluation of PAPR reduction techniques such as clipping and filtering, selective mapping (SM), and PTS. In the simulation with same number of subcarriers and parameters mentioned under CE, a frame size of 10,000 with oversampling factor of 4 were used for 16-QAM. For PTS, a phase factor of $\{\pm 1, \pm j\}$ were chosen from the set. Among PTS and Selective mapping, SM gives much less PAPR value against CCDF with very less complexity as shown in Fig. 10.

PTS is often considered to be a modification of the selective mapping method with better results because it eliminates the need for side information. A much simpler method of clipping followed by filtering was also done to solid the results. The evaluation in Fig. 11 shows that the PAPR for original signal came at 15.3 dB while the PAPR of the clipped signal was determined to be 2.8 dB and that of the clipped and filtered signal was obtained as 10.6 dB.

6 Conclusion

A comb type pilot insertion based channel estimation was presented in this paper. The channel estimation is performed using the received pilot symbols and the estimated values will be only for the pilot carriers. For obtaining the values at data carriers, an interpolation technique is being used. The estimation was carried out with varying pilot counts and significantly reduced to 9 values. From the results, obtained MMSE showed minimum distortions compared with LS estimator. However, with lesser number of pilots CS outperforms LS and MMSE estimators. In a MIMO-OFDM system with larger multipath, CS based channel estimation is appropriate to avoid computational complexity in channel estimation. A joint PAPR reduction technique was also addressed in the work. Clipping has proven to be an effective method to reduce PAPR without spectral expansion but at the cost of introducing out-of-band emission, which is further eliminated using filtering. Selective mapping boasts slightly superior PAPR reduction with lesser redundancy and PTS results depict a similar but lesser PAPR reduction with the added benefit of lower complexity as it eliminates the need for side information.

Sparse channel estimation is becoming prevalent in the channel estimation field, OMP based PN sequence channel estimation in MIMO-OFDM can be implemented as an addition to the work done in this project. Deep learning approach for channel estimation is also an upcoming research field. Particle Swarm Optimisation (PSO) technique can be integrated into PTS to optimize the weighting factor, which can further reduce the complexity of the algorithm.

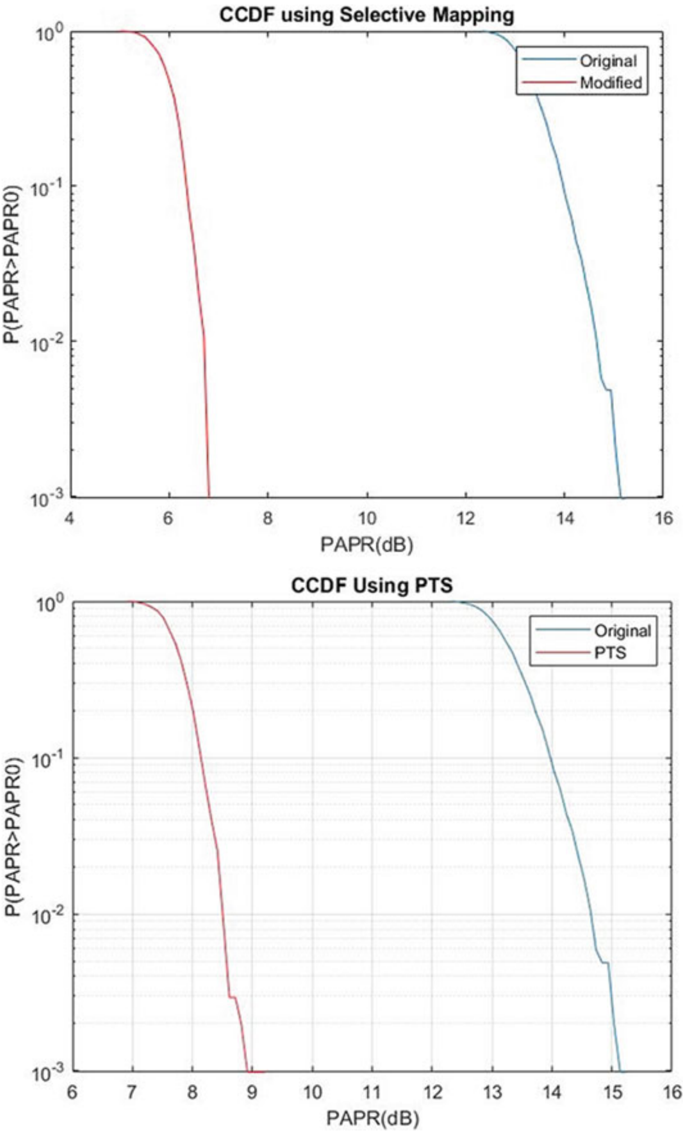


Fig. 10 CCDF plot for selective mapping and PTS

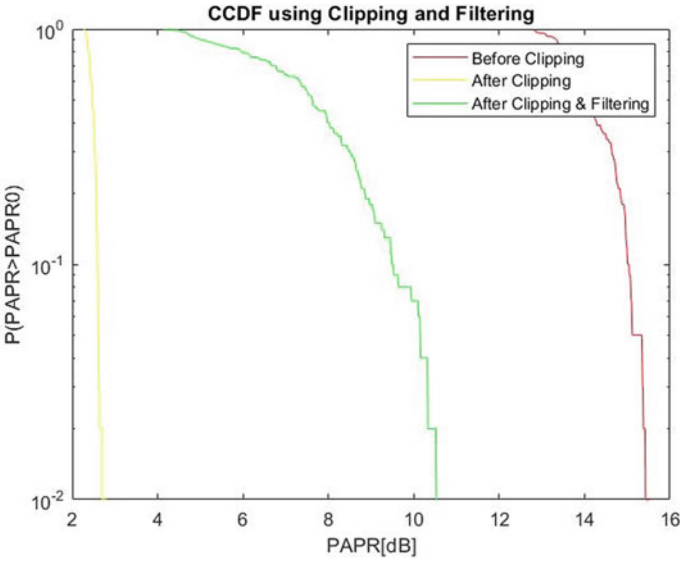


Fig. 11 CCDF plot for clipping and filtering

References

1. T.C. Snehith, K.K. Anil, A.K. Raju, R. Ramanathan, Impact of channel estimation errors on lattice reduction gains in MIMO systems, in *Advance Computing Conference (IACC),2015* (IEEE International, 2015)
2. B. Gong, L. Gui, S. Luo, Y.L. Guan, Z. Liu, P. Fan, Block pilot based channel estimation and high- accuracy signal detection for GSM-OFDM systems on high-speed railways. IEEE Trans. Vehicular Technol. **67**(12), 11525–11536 (2018). https://doi.org/10.1109/TVT.2018.2869679
3. A. Khlifi, R. Bouallegue, Performance analysis of LS and LMMSE channel estimation techniques for LTE downlink systems (2011) arXiv preprint arXiv:1111.1666
4. K.M. Manu, K.J. Nelson, OFDM channel estimation using total variation minimization in compressed sensing, in *International Conference on Contemporary Computing and Informatics (IC3I)* (IEEE, 2014), pp. 1231–1234
5. Y.-C. Wang, Z.-Q. Luo, Optimized iterative clipping and filtering for PAPR reduction of OFDM signals. IEEE Trans. Commun. **59**(1), 33–37 (2010)
6. M.H. Aghdam, A.A. Sharifi, PAPR reduction in OFDM systems: an efficient PTS approach based on particle swarm optimization. ICT Express **5**(3), 178–181 (2019)
7. L. Dai, Z. Wang, Z. Yang, Compressive sensing based time domain synchronous OFDM transmission for vehicular communications. IEEE J. Sel. Areas Commun. **31**(9), 460–469 (2013)
8. S.S, .R, J., S. Kirthiga, Superimposed pilot based channel estimation for MIMO systems, in *Joint International Conference on Artificial Intelligence and Evolutionary Computations in Engineering Systems* (SRM University, Kattankulathor, Chennai, 2016)
9. M. Mowla, M. Ali, R.A. Aoni, Performance comparison of two clipping based filtering methods for PAPR reduction in OFDM signal, arXiv preprint arXiv :1403.3349 (2014)
10. D. Rajeswaran, A.K. Nair, A novel approach for reduction of PAPR in OFDM communication, in *2016 International Conference on Communication and Signal Processing (ICCSP)* (2016), pp. 1978–1981

11. J.A. Sam, A.K. Nair, Analysis and implementation of channel estimation in OFDM system using pilot symbols, in *2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)* (2016), pp. 725–728
12. M. Umesha, S. Swamy, Comb type pilot arrangement based channel estimation for spatial multiplexing MIMO-OFDM systems. Int. J. Eng. Res. Technol. (IRJET) (2018)
13. V. Vahid, E. Saberinia, Performance enhancement of OMP algorithm for compressed sensing based sparse channel estimation in OFDM systems, in *Information Technology-New Generations* (Springer, Cham, 2018), pp. 3–8
14. K.K. Desai, Comparison of SLM and PTS method for PAPR reduction in OFDM. Int. J. Eng. Res. Technol. (IJERT) **03**(06) (2014)

A Wearable Infant HealthCare Monitoring and Alerting System Using IOT

P. Bhuvaneshwari and Rajnish Mahaseth

Abstract In recent years, physical monitoring of the infant is restricted when the infant is kept under critical surveillance. So many neonates/infants are prone to sudden ambiguous death which is technically termed as sudden infant death syndrome (SIDS). To overcome this issue, in this work we have developed an intelligent infant monitoring prototype which is the integration of software as well as the hardware components such as sensors. The proposed model performs the real time monitoring of the baby activities as well as analyzes the basic health parameters using wearable wireless sensor devices. The proposed model is an important low-cost wireless apparatus that uses GSM technology for monitoring and alerting the health condition of the infant to the doctors and parents. Experimental results have shown that the proposed system has the caliber to monitor continuously and notify the life-threatening situations of an infant to the corresponding authorities.

Keywords Sudden Infant Death Syndrome (SIDS) · Intelligent Infant monitoring prototype · Real-time monitoring · Wearable wireless sensor device · GSM technology

1 Introduction

The healthcare sector is going through a huge change with digital technologies, and the way doctors interact with their patients. But still, the number of infant deaths is increasing day by day and exhibits no signs of having suffered [1]. Sudden Infant Death Syndrome (SIDS) is a sudden death that usually occurs during an infant's sleep that is unpredicted even after the forensic autopsy and investigation [2]. It is the major cause of mortality under one year of age of an infant. SIDS is at its peak when the infant is between 2 and 3 months of age. The sudden increase or decrease in physiological parameters like body temperature, heartbeat, breathing rate, and other physiological parameters are the reasons for SIDS [10]. Some of the strategies followed in the hospital to prevent this type of death are, placing an

P. Bhuvaneshwari (✉) · R. Mahaseth
MVJ College of Engineering, Bangalore-67, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_67

infant in a relatively warm and comfortable sleeping environment, utilizing a pacifier and putting it with no loose bedding, and avoidance of exposure to harmful smoke/ tobacco [3]. Generally, in all hospitals during critical situations, the baby is kept under the Intensive Care Unit (ICU) and is separated from mother. The nurses and doctors take care of an infant; however, hospital personnel is also required to do other work at the same time. Sometimes, due to negligence and because of some other factors, it creates an opportunity for SIDS.

This problem motivated us to design an IOT based infant monitoring system to prevent the syndrome and to keep the baby safe by monitoring the baby continuously from remote places. Infant monitoring system is the combination of emerging technology of IOT sensors, cloud technology, wireless communication, and mobile applications which monitor the infant even when he/she is sleeping. In this work, an IoT enabled wireless sensor and electronic devices are used to sense the infant's surrounding parameters like the temperature and humidity, respiratory level as well heartbeat. This prevents any abnormality from happening. Another advantage of this model is the live monitoring of the infant by the users. The data collected from the sensors and camera are stored in the cloud, which enables users to view the details in the system or the mobile phone through a network connection. These data are transmitted to users via wireless telecommunication devices. The specialty of this model is it is enabled with GSM which will automatically trigger an alarm to the doctors and parents if the data readings reach a certain level that can harm the infant.

This paper is organized as follows: In Sect. 2, the problem statement is defined. In Sect. 3, various works done on the infant monitoring system are discussed. The proposed system is briefly described in Sect. 4. The experimental results are discussed in Sect. 5. Finally, the conclusion and future work are presented in Sect. 6.

2 Problem Statement

The increasing mortality rate of newborns or infants is prevalent in developing countries. The major reason for that occurrence is large number of newborns under the control of each physician and the negligence of the serious condition of an infant at the right time. Developing a special health care environment for newborns/infant will effectively decrease the mortality rate. It increases the quality of healthcare and decreases the burden of the existing healthcare system.

3 Literature Review

The IoT (Internet of Things) and cloud storage are the most known technologies of today's Telemonitoring health system. A Telemonitoring health system is a tool that provides an early indication of patient status, allows for immediate diagnosis and intervention, and the details are recorded, evaluated, and controlled [6]. This technology is used to keep track of infants' physiological parameters like body temperature, heartbeat, breathing rate using different sensors like temperature sensor, heartbeat sensor, sound sensor, etc. Some of the proposed systems in monitoring infants are as follows.

Costa et al. [4] designed an incubator that can sense the environment by using a humidity control system. They confirmed that the control of humidity could lead to improving the infant's quality of life by bestowing a thermo-neutral environment. Dive and Kulkarni [5] have developed a similar kind of incubator that notify the doctor and nurse about the baby's condition. The advantage of the work is it alerts the doctor and nurse when the baby cries. For future improvement, they recommended adding parameters such as monitoring of heart pulse and humidity. Lin et al. [8] proposed a model that measures the infant body temperature, sleeping position, and respiratory rate. They experimented with the model by transmitting the result to the server. Fonseca et al. [9] used on-body sensors to detect the heartbeat, temperature, respiratory, and body position of the infant. The proposed model evaluates the risk and sends the alert message to the corresponding person. Ferreira et al. [9] proposed a night watch device that is connected to the infant chest belt to monitor the activities of a baby. The system is dependent on ZigBee to detect the physiological parameters of an infant. This existing system is very expensive.

Kale and Khandelwal [11] developed a telehealth monitoring system with three sensors for measuring temperature, ECG, and pulse rate of the infant. The measurement of these sensors is sent to the physician's computer or phone as SMS. Benharref and Serhani [12] designed a novel healthcare electronic system that relies on the service based cloud architecture to monitor and reduce the impact of infant chronic diseases. The model gathers allows real-time monitoring, extracts data from sensors, identifies risk factors and sends alarm messages to doctors about mishappenings. Patil and Mhetre [13] proposed a baby monitoring system to monitor the baby's body temperature, heart rate, and moisture condition. The model included a sound buzzer that triggers an alarm to initiate the action to control the environment.

Generally, these systems are capable of carefully monitoring infants in the intensive care unit or at home while infants perform their daily activities without affecting their comfort. However, the conventional sensors and clinical instruments can't be utilized for wearable physiological monitoring applications as they are hard to wear for a long time and they additionally create disturbance and stress. Because of the restriction of technologies in sensors, remote systems, and power supply. The traditional wearable sensor system is commonly not satisfactory for the robust long term and easy monitoring of newborn children in real life conditions [7]. Even Though so many works are proposed in the monitoring of the infant, most of the works fail to support portability and compactness in the possible real applications.

In this fast revolutionary world, there is a need to find a special device to take care of the infant even when the parents are busy at work. The main goal of the application of wearable sensor systems in infant health monitoring is to grant centered care every day by providing appropriate alert notifications, reducing the onset of complications, and reducing the cost of hospital-based clinical interventions. Moreover, warning devices that can continuously monitor and alert the caretakers once the threshold reaches will protect the infant from the critical situation. Accordingly, the purpose of the proposed system of infant monitoring system will consist of a warning device and sensors that can measure the infant humidity level, temperature level, respiratory level, and heart rate through a wearable device. Consistent and clear-cut track of physical specification about the child is an enhancing mechanism which is attainable and agreeable. Further, along with a deeper satisfactory review, demands the operation of a unified sensing and automatic alerting model which enhances the accuracy in the health anticipation system.

4 Proposed Methodology

In this work, we have attempted to break the limitation of the existing systems. Figure 1 presents the framework of the proposed model to prevent SIDS. This system mainly contains some sensors like Heartbeat (SN Pulse sensor), temperature and Humidity (DHT11), Respiratory sensor SpO2, Hardware units like Raspberry Pi 3 model b, GSM module, Software units like cloud, website, mobile app for parents.

The infant is tied with a wearable sensor watch that measures the infant's body temperature and humidity through DHT11 temperature and humidity sensor. It is an ultra-low cost sensor that measures the body temperature too. The sleeping position of the infant plays a major role in measuring the heartbeat as well the respiratory level. When the baby sleeps in a prone position, the , as well as the respiratory level, varies from the normal range. A heartbeat SN Pulse sensor is used to measure the heartbeat of an infant and the normal heartbeat rate at active conditions is 100–160 Beats Per Minute (BPM). The Sp02 sensor is used to monitor the respiratory level of an infant. Once the value falls down or rises from the sensors, then the alert message will be sent to the microcontroller. An MCP 3008 ADC is an electronic device that

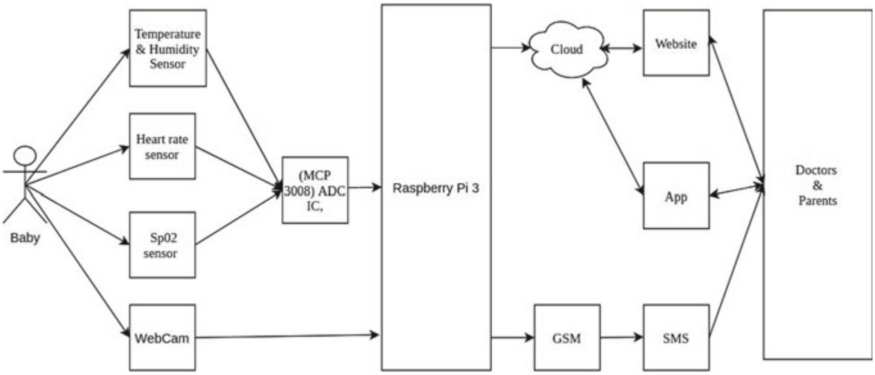


Fig. 1 Proposed architecture for infant monitoring system

is used to convert input from analog to digital. The Raspberry Pi model will take the input from the MCP 3008 ADC and store the output data into the cloud. To run Raspberry Pi we used the Raspberry jersey operating system. The hardware unit is utilized to continuously update the data to the database with the help of the Internet. For all these devices' functionality, Raspberry Pi and Internet connectivity play a major role.

The sensors and hardware unit do various tasks like retrieving the data from infants, update the data in the cloud, and compare these data with threshold values. After receiving the data, this set of information is processed, sent to the Gateway, and it is accessible to the user through the web and app interfaces. If a data rate reaches critical value, then the device will trigger an alarm, and an alert message is sent to the doctor as well as parents through mobile Short Message Service (SMS) with the help of Global System for Mobile communication (GSM).

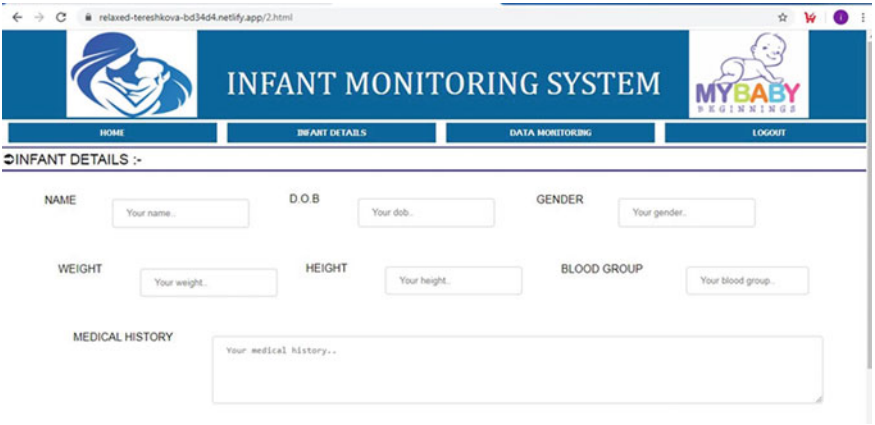
To store data into the cloud we used the Thingspeak cloud. Thigsspeak is an Internet of Things (IoT) based cloud platform. Cloud sends the data to the entity's mobile application as well as the web. We integrated all the sensors, and the proposed prototype is tested in real time under a local network environment. This system helps the hospital representative and parents to take necessary action without delay.

5 Results and Discussions

The proposed model contains various user interfaces and medical service interfaces. The objective of this work is to monitor the infant using wireless and cost-effective sensors from anywhere. The model is successfully deployed and tested in a real-time environment. The result shows that the system is working perfectly. Immediate alerting mechanism using an SMS to mobile is an add on advantage of the model. The real-time sensors data is updated and visualized on the website and in the android app 24×7. In real-time, implementing this safety system would control and lower down the death rate caused by SIDS in Infants.

Once the user login via web/app and search the infant details with the unique infant id, the system displays the infant body temperature, humidity, and heartbeat rate, respiratory rate along with basic information like height, weight, name, blood group, as illustrated in Fig. 2. In Fig. 3, the results of the temperature and humidity sensor are depicted.

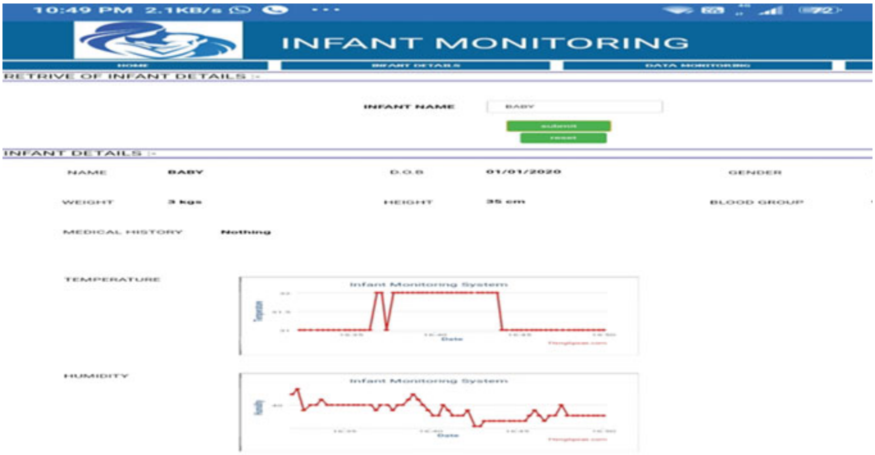
The user can access the infant details on the website as well as using the application. Figure 4 illustrates the view of the infant monitoring system app homepage view. The infant BPM rate is displayed as signals in the app display is shown in Fig. 5. During an emergency ie. Once the sensor values rise or fall the threshold value, then the alert message is sent to patients and doctors as SMS which is shown in Fig. 6. Table 1 shows the heart rate and respiratory values of an infant in the normal condition. In Table 2, the price of the required hardware equipment to develop the proposed model is listed.



INFANT DETAILS :-

NAME Your name..
D.O.B Your dob..
GENDER Your gender..
WEIGHT Your weight..
HEIGHT Your height..
BLOOD GROUP Your blood group..
MEDICAL HISTORY Your medical history..

Fig. 2 Homepage of infant monitoring system Website



RETRIVE OF INFANT DETAILS :-

INFANT NAME BABY

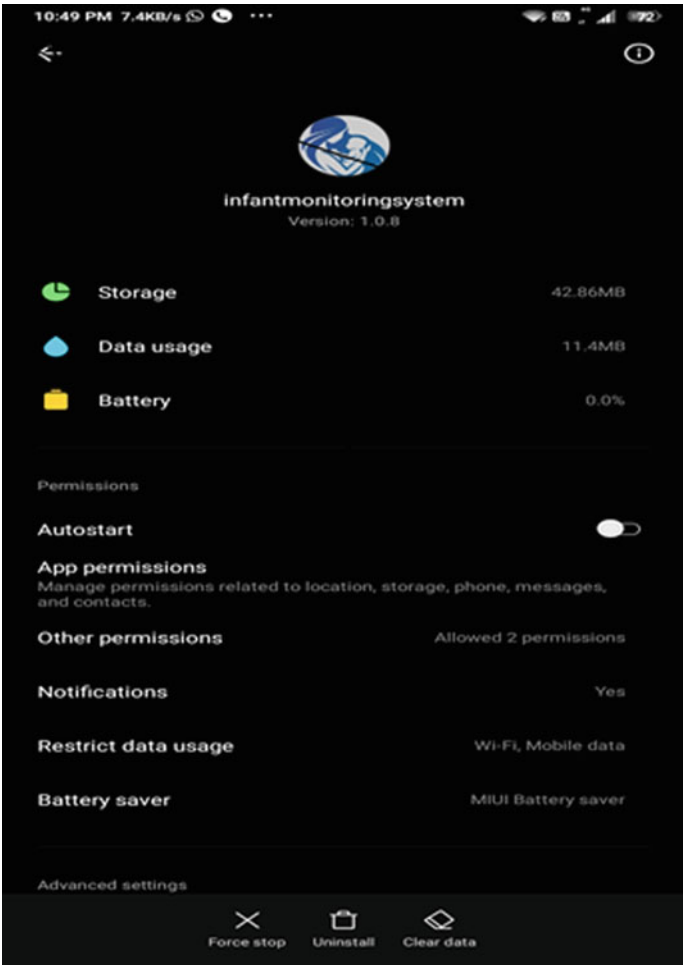
INFANT DETAILS :-

NAME BABY
D.O.B 01/01/2020
GENDER
WEIGHT 3 kgs
HEIGHT 35 cm
BLOOD GROUP
MEDICAL HISTORY Nothing

Fig. 3 User interface with infant details

6 Conclusion and Future Work

One of the critical issues for infants with SIDS problems is monitoring and alerting the user in an emergency. In this paper, we have developed a smart wearable prototype that is responsible for monitoring and reporting the condition of an infant which is kept in a critical unit. The data can be updated and accessed through the web or mobile application all the time. In an emergency case, the alert message is sent to mobile as SMS with the help of GSM. The data rate generated by the wearable device ranges from 29–37 bytes per minute. The most important benefit of this



infantmonitoringsystem
Version: 1.0.8

Storage	42.86MB
Data usage	11.4MB
Battery	0.0%

Permissions

Autostart

App permissions
Manage permissions related to location, storage, phone, messages, and contacts.

Other permissions	Allowed 2 permissions
Notifications	Yes
Restrict data usage	Wi-Fi, Mobile data
Battery saver	MIUI Battery saver

Advanced settings

Fig. 4 Homepage view of Infant monitoring system application

system is the doctors and parents can monitor the physical and medical conditions of the infant from a remote place 24×7. Also, the infant is free from skin irritability, and the inconvenience caused by more wires. Based on the real time evaluation, the developed system works well in the local network environment and provides promising results in terms of reliability, efficiency, portability and compactness. In the future, we have a plan to include a few more parameters like oximeter, barometric pressure sensor, pulse rate sensor, carbon sensor, and motion detector.

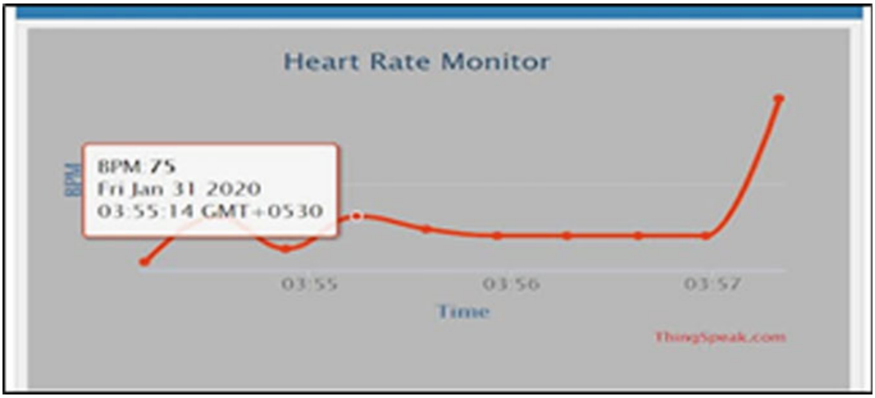
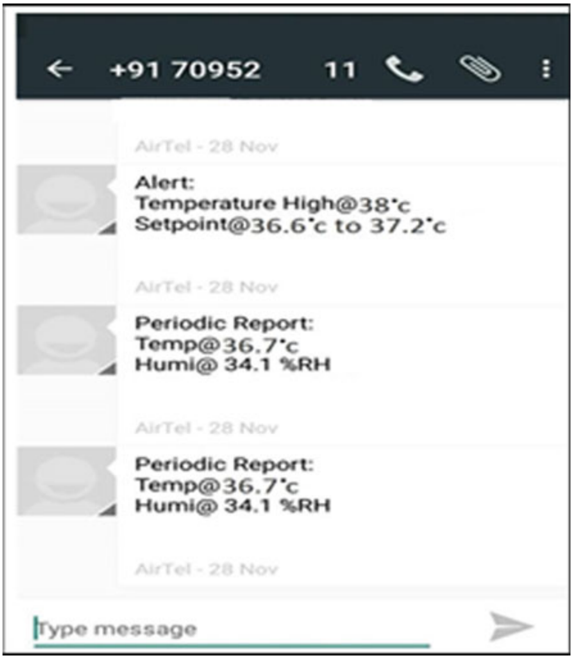


Fig. 5 Real time view of heart rate monitoring

Fig. 6 Sample mobile alert SMS sent to user



AirTel - 28 Nov

Alert:
Temperature High@38°c
Setpoint@36.6°c to 37.2°c

AirTel - 28 Nov

Periodic Report:
Temp@36.7°c
Humi@ 34.1 %RH

AirTel - 28 Nov

Periodic Report:
Temp@36.7°c
Humi@ 34.1 %RH

AirTel - 28 Nov

Table 1 Infant pulse rate

Age category	Heart beats/min	Respiratory Rate/Breath
New Born	100–160	30–60
Infant (1–5 Months)	90–150	30–50
Infant (6–12 Months)	80–140	24–46

Table 2 List of the hardware components with price

S.NO	Types of hardware	Price/Unit (Rs)
1	DHT 11 Temperature and humidity sensor	95.00
2	SN Pulse sensor	300.00
3	Sp02 sensor	1200.00
4	MCP3008N (ADC IC)	242.00
5	SD card	329.00
6	GSM	1099.00
7	Breadboard and jumper wires	172.00
8	Raspberry Pi 3 model B	3090.00
9	Intex webcam	699.00
	Total	7226.00

References

1. P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Vol. 1, (IEEE, Dec, 2001), pp. I-I
2. R. Minetto, N.J. Leite, J. Stolfi, Affleck: Robust tracking of features in variable-zoom videos, in *2009 16th IEEE International Conference on Image Processing (ICIP)*, (IEEE, Nov, 2009), pp. 4285–4288
3. D.L. Hoyert, J. Xu, Deaths; Preliminary data for 2011 (2012)
4. J.L. Costa, C.S. Freire, B.A. Silva, M.P. Cursino, R. Oliveira, A.M. Pereira, F.L. Silva, The humidity control system in the newborn incubator. Fundam. Appl. Metrol. (2009)
5. K. Dive, G. Kulkarni, Design of embedded device for incubator for the monitoring of infants. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **3**(1), 541–546 (2013)
6. I. Murković, M.D. Steinberg, B. Murković, Sensors in neonatal monitoring: Current practice and future trends. Technol. Health Care **11**(6), 399–412 (2003)
7. E. McAdams, A. Krupaviciute, C. Géhin, E. Grenier, B. Massot, A. Dittmar, J. Fayn, Wearable sensor systems: The challenges, in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, Aug, 2011), pp. 3648–3651
8. W. Lin, R. Zhang, J. Brittelli, C. Lehmann, Wireless infant monitoring device for the prevention of sudden infant death syndrome, in *2014 11th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT)* (IEEE, Oct, 2014), pp. 1–4
9. Â.M. Fonseca, E.T. Horta, S. Sendra, J.J. Rodrigues, J.A. Moutinho, A sudden infant death prevention system for babies, in *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)* (IEEE, Oct, 2014), pp. 525–530
10. Research on sudden infant death Syndrome (SIDS), National Institutes of Health, Available online: https//www.nih.gov. 10 Dec 2017
11. S. Kale, C.S. Khandelwal, Design and implementation of real time embedded tele-health monitoring systems, in *2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT)* (IEEE, March, 2013), pp. 771–774
12. A. Benharref, M.A. Serhani, Novel cloud and SOA-based framework for E-Health monitoring using wireless biosensors. IEEE J. Biomed. Health Inform. **18**(1), 46–55 (2013)
13. S.P. Patil, M.M.R. Mhetre, Intelligent baby monitoring system: a review. J. Instrum. Technol. Innov. **4**(1), 16–23 (2014)

Resource Contention and Age of Information in Wireless Sensor Nodes Using Renewal Theory

K. Sureshkumar, V. Rajmohan, Sanju Rajan, and V. R. Prakash

Abstract Age of information (AoI) is an associated metric with the link and is subjected to resource contention of deployed sensors. Resource contention and age of information (RCAoI) is proposed in this work and incorporates renewal theory for calculating and updating the status of nodes. The inter arrival time of packets to the sink is initially considered along with the number of update packets in the deployed scenario for updating the AoI. The protocol subsequently calculates the renewal equation indicating the bounded time for sensors and their reporting frequency. The simulation results thus indicate the insightful information in providing persistence transmissions and scheduling. The scheduling process from sink prioritizes transmission from nodes with values of increasing AoI.

Keywords AoI · Renewal theory

1 Introduction

Age of information is a significant metrics when working in harsh environment and in heterogeneous resources availability amongst sensors in reporting to sink. Influence of channel age of Information has been studied in Huang et al. [1] where the service rate is modelled using "Markovian Modulated service" (MMS). This MMS

K. Sureshkumar (✉)
Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh 522502, India
e-mail: m.k.sureshkumar@kluniversity.in

V. Rajmohan
Department of Electronics and Communication Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, India

S. Rajan
School of Computing Sciences, Hindustan Institute of Technology & Science, Chennai, Tamilnadu, India

V. R. Prakash
Department of Electronics and Communication Engineering, Hindustan Institute of Technology & Science, Chennai, Tamilnadu, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_68

characterizes between good and worst state of a channel, describing the variation of packet arrival and service times according To AoI. The age related metrics has been classified based on two criteria in Yates et al. [2], namely system age and individual age. The former denotes the age used to monitor an event from a group of sensors. The latter chooses an individual sensor amongst the group, and its age using stochastic hybrid system. Tribulation has been in un-slotted sensor which losses its stability on overloading. Age-related information discussions in multi-armed bandit with single-hop network [3]. The optimal scheduling policy to maintain the average cost via decoupling and indexing considering every single arm. However, the work does not consider arrivals to be stochastic in nature. The AoI has been studied to denote the transmission possibilities in Zhou ans Saad [4], AoI at the device and AoI at the receiver. The approach describes collaboration amongst devices to schedule its transmission based on non-uniform sample size of status update. If the status updates are independent from each other, it achieves significant performance. The scenario fails when status updates are not mutually independent or with limited buffer capability. In the proposed work, “resource contention and age of information” describes the bounded time in reporting from sink and status of updates via renewal theory. The number of updates at sink is also monitored so as if a node associated losses connectivity quicker response of renewal update is provided.

This paper is organized as follows, and Sect. 2 deals impact of energy of sensors, traffic profiles for age of information and problem description. Section 3 deals with influence on proposed system via RCAoI. Section 4 deals with simulation of NS2 a discrete event simulator. Section 5 deals with overall age of information and vision for future work.

2 Related Works

Energy harvesting is discussed in Krikidis [5] with two node topology. It transfers an update packet to measure the age of information when the capacitor within a battery is full charged to manage its energy resource. It relies on time divided units with orthogonal channels incorporated in small scale networks without interference and cannot be extended to larger networks. Similarly, the impact of heterogeneous traffic has been studied in Chen et al. [6] to discriminate the difference between grid connected and energy harvesting sensor for AoI. In both sensors deployed, AoI decreases as the energy harvesting increases. Additionally, the observation depicts the AoI increases depending on arrival process prior to saturation. The work assumes multiple packets have been decoded which is confined small scale network.

In general, the modelling system discussion on network control for delay in communication associates it to sampling time and its scheduling policy. This is coarsely divided into three types namely: discrete time, continuous time and hybrid approach [7]. Markov decision process to quantify the nonlinear metric of age is discussed to reduce the staleness of data using continuous and discrete sampling [8]. However, the work of MDP does not consider the service time of queuing to change as per its sampling policy. The impact of sampling policy is determined in the proposed work which has not been discussed earlier.

The age of information in the scenario of multisource and single sink is discussed using M/M/1 and M/G/1. Time sensitive metrics of AoI depend upon inter arrival and service time of packets when single sink is deployed [9]. However, the FIFO queue to be in stationary state which is applicable when packet generation rate is less than service time violation is not discussed. The "Lynapunov optimization" has been discussed in Fountoulakis et al. [10] to balance the trade-off between sampling cost and transmissions error. The algorithm discriminates whether to initiate sampling a new packet or to transmit priory sampled or to remain idle. The process of acknowledging might increase the overhead in networking which is used to attain stable queues in the work of [10]. Incorporating blockchain in AoI has been discussed in Lee et al. [11] with number of transactions and its time out parameter within a specified block size. However, using ledger for channel will increase the communication delay in processing which might but inapplicable in time sensitive information processing. In Azarhava et al. [12] non-orthogonal multiple access (NOMA) has been preferred than the traditional TDMS for calculating AoI. The topology uses some specific centralized sensor based on power for fusion which obtains channel state information. The work of NOMA and TDMA considers data to be received in chronological order with prefect channel state information which is difficult in harsh environment.

Assigning penalty function to characterize age of information in error prone channel is discussed in Chen et al. [13]. It states that main tribulation along bandwidth and power can be refrained by Lagrange multiplier which discriminates individual sensor from a group of sensors. Finally, continuous Markov process is used to schedule as per the optimal policies. The work in Chen et al. [13] states sink which are resource rich and the endogenous factors of channel state incurs same packet loss suitable for scheduling. The quality of service in Emara et al. [14], with IOT networks under time and event triggered traffic is discussed. The macroscopic information of mutual interference modelled via a stochastic geometry influence of microscopic impact of time and event triggered traffic. However, small scale fading and the determining factor in offset of devices which are not considered might increase the interference factor. Unsynchronized transmission also results in increase in AoI. Synchronizing sensor nodes within a delay bound and transmission via autoregressive model provides a suitable solution [15]. However, attaining perseverant synchronization and transmitting is not always feasible in the context of phase offset to transmit may cause an increase in AoI. Maximum AoI threshold based upon network size categorized as small and large. In small network, a cyclic scheduler determines the feasible states since the load is less. Large scale the load factor imposes the scheduler to determine the data rate over a period and perform a polynomial mapping [16]. However, processing a sample within the allocated time slot cannot be incorporated always which might result in accumulating AoI.

Forecasting the AoI has been done in multi-hop network considering the probability mass function. It uses a line network topology where the first stationary links updates the next [17]. The same loss probability cannot be accounted at all slots for networks other than line topology. Effective hydrocast, in Anand and Titus [18], tries to attain energy efficiency updating information for intermittently connected wireless sensors and modelling transmission. However, reducing the number of transmission to maintain efficiency results in stale packets processed at sink. In certain applications, sensors go into sleep state to conserve energy which might influence the sensing time and increase the AoI. This has been resolved as a non-convex problem where in the AoI, and sensing time has been considered in Bedewy et al. [19]. This is achieved where in a group of sensors are consequently share the parameters to an access point. The access point predicts the process of traffic flow and then subsequently sensors goes to sleep state. Assigning transmission times again trade-off of sleep times and battery regime provide better results in Bedewy et al. [19]. However, the work is feasible if source within the range of each other coordinating shared access. The work results in different back log of transmission times when source are out of range to one and another.

2.1 Problem Description

Age of information problem is associated with the following, source of energy of sensor nodes, link queuing strategies for information processing from sensor to sink. The process of finding and associating a fixed policy in calculating AoI failed. The pre-emption policy of calculating the flow results in distribution of heavy tails at the flow [20]. So Pareto solutions for finding the update packet are needed to maintain unbiased traffic flow in a non-pre-emptive manner.

2.2 Event Processing Scenario for Update Packet Interval

The scenario of 5 events and the update packet interval in terms of processing and arrival at sink is shown in Table 1.

The deployed scenario considers 5 events arriving at sink from an individual sensor deployed within the terrain. The service time (ST) is calculated when there is acknowledgement from sensor node. The work assumes processing time of packet can be calculated via acknowledgements. Update packet interval is calculated by Eq. 1 using formula below.

$$\mathrm{UP} = \mathrm{EEP} - \mathrm{SEP} \tag{1}$$

The EEP has been calculated using Eq. 2 which is given below as a sum of SEP and the service time.

$$\mathrm{EEP} = \mathrm{SEP} + \mathrm{ST} \tag{2}$$

Table 1 An example scenario for update packet and its processing time

Events	Inter arrival time (IAT) (s)	Arrival time (AT) (s)	Service time (ST) (s)	Start time of event processing (SEP) (s)	Ending time of Event processing (EEP) (s)	Update packet interval (UP)
1	0.45	0.45	0.37	0.45	0.82	0
2	0.11	0.56	0.40	0.82	1.22	1.22–0.96 = 0.26 s
3	0.16	0.72	0.41	1.22	1.63	1.63–1.23 = 0.40 s
4	0.12	0.84	0.47	1.63	2.10	2.10–1.31 = 0.79 s
5	0.18	1.02	0.62	2.10	2.72	2.72–1.64 = 1.08 s

The above scenario is suitable for single-hop communication. In multi-hop communication, the tribulation in contention and bounded metrics of configuring update is intermittent.

3 Proposed Work

The purpose of resource contention and age of information is to find a deterministic value for sample path which is countable. This is achieved by using renewal theory which calculates the inter arrival times. The flow diagram for resource contention and AoI is shown in Fig. 1.

3.1 Resource Contention and Age of Information

The algorithm initially calculates renewal time for every link within a sample path as in Eq. 3 below.

$$\mu_A = E(X_n) = \int\limits_0^\infty x dF(x) \tag{3}$$

In Eq. 3, the mean is denoted by using $\mu_{A,}$ and the X denotes the inter arrival time of packets in a link. The value of F denotes the function of all inter arrival time within a sample path which is denoted by random variable.

The data transfer takes place where X_1 the first packet from source node reaches the cluster head at time "t" seconds. The work assumes sink is stationary and predetermines which the renewal process of all its sensors. The sensors and path are also

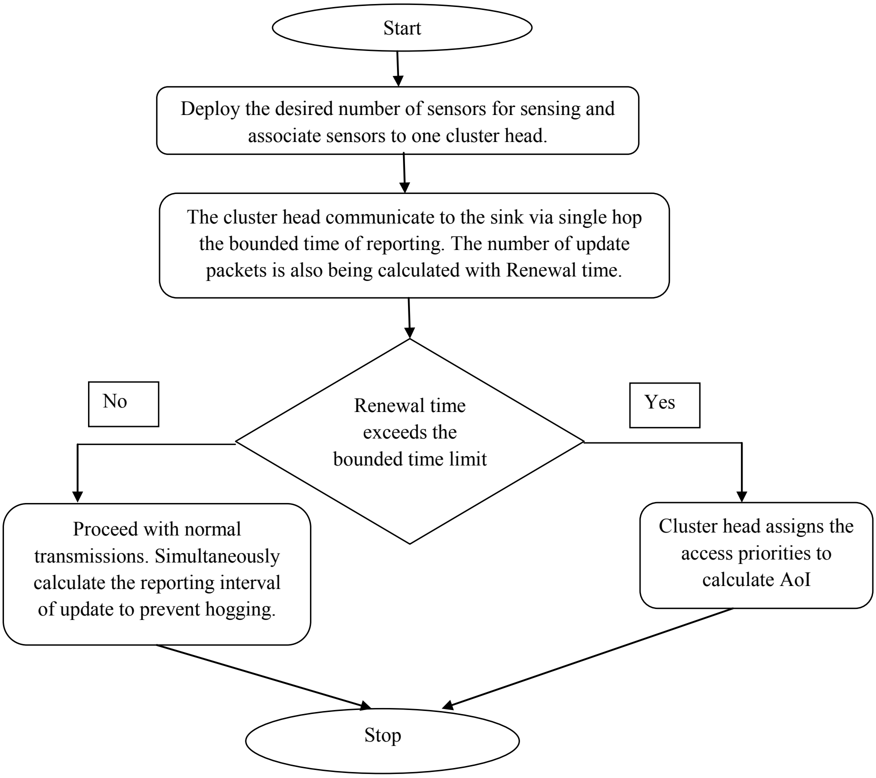

Start
Deploy the desired number of sensors for sensing and associate sensors to one cluster head.
The cluster head communicate to the sink via single hop the bounded time of reporting. The number of update packets is also being calculated with Renewal time.
Renewal time exceeds the bounded time limit
No
Yes
Proceed with normal transmissions. Simultaneously calculate the reporting interval of update to prevent hogging.
Cluster head assigns the access priorities to calculate AoI
Stop

Fig. 1 Flow diagram for resource contention and AoI

assumed to be independent identical distributed. The second, data packet, takes the time interval for $(n-1)$th renewal update subsequently after the nth renewal update.

The time of nth renewal is given by using Eq. 4. The individual time point for each arrival is given by using $I_{1,}$ $I_{2,}$ and summation of all the arrivals is denoted by using I_n.

$$I_n = \sum_{i=1}^{n} X_i \tag{4}$$

The N (t) denotes number of renewals with time "t" and is denoted as in Eq. 5 below.

$$N(t) = \max\{n;\ I_n \leq t\} \tag{5}$$

In most scenarios after instantaneous sensor deployment, the renewal time is known are can be calculated. However, this metrics changes where contention and transmit times alters the updating interval leading to increase in AoI. The renewal equation $R\ (t)$ has been calculated as in Eq. 6.

$$R(t) = E[N(t)] \tag{6}$$

The protocol calculates the first update renewal by conditioning it using the first renewal packet at $X_{1.}$

$$R(t) = \int_0^\infty E[N(t)|X_1 = x]dF(x) \tag{7}$$

If the sensor exceeds, the limit of bounded time in reporting to the sink it has been written as in Eq. (8). This indicates the AoI is increasing and scheduling from cluster head to cluster member has to take place.

$$R(t) = \int_0^\infty E[N(t)|X_1 = x > t] = 0 \tag{8}$$

If the sensor does not exceed the limit of bounded time in reporting to the sink it has been written as in Eq. (9).

$$R(t) = \int_0^\infty E[N(t)|X_1 = x < t] = 1 + R(t - x) \tag{9}$$

Thus, the sink evaluates the AoI via Eq. 8 and schedules the missed nodes in updating the status. Alternatively, it also monitors the number of updating nodes, and its frequency of updates via Eq. 9.

4 Results and Discussion

The results of simulation work using network simulator 2 try to quantify the relation between age of information (AoI) which is tightly coupled with serving time of sink, average number of packets and expected utilization of channel (Table 2).

Table 2 Desired simulation parameters used for implementing RCAoI protocol

Parameter	Value
Terrain	500 × 500 m^2
Maximum number of sensor nodes	100
Total number of sinks	8
Initial energy	5 J
Transmit power	0.60 W
Receive power	0.10 W
Payload size	300 B

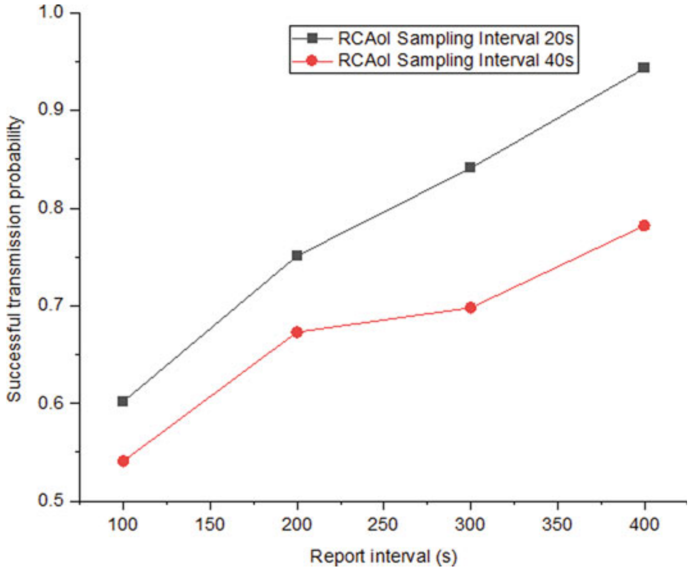


Fig. 2 Successful transmission probability versus report interval

The sampling interval is taken in Fig. 2 states the number of transmission and its success probability. This indicates consecutive data transmission takes place for RCAoI for two different sample intervals.

Influence of channel load and the packet loss rate has been shown in Fig. 3. The loss rate discusses the mismatch between the prior model and the effect of channel load in sending data to sink.

The channel load (CL) is calculated as in Eq. 10 below.

$$\mathrm{CL} = \frac{\mathrm{n}(1 - \text{channel utilization})}{\text{number of slots}} \tag{10}$$

Number of sensors is denoted by "*n*" associated to a selected sink. The channel utilization is calculated as in proposed system with equation. The number of slots is assigned based on the distance of nodes towards the sink. The packet loss rate decreases in both scenarios of proposed work indicating the reduced loss in channel and increase in processing time of sink for updating AoI.

5 Conclusion

The RCAoI protocol balances the successful transmission in terms of terms of different sampling periods indicating it to be superior for different traffic regime. In addition, the protocol also balances the packet loss rate with different channel loads

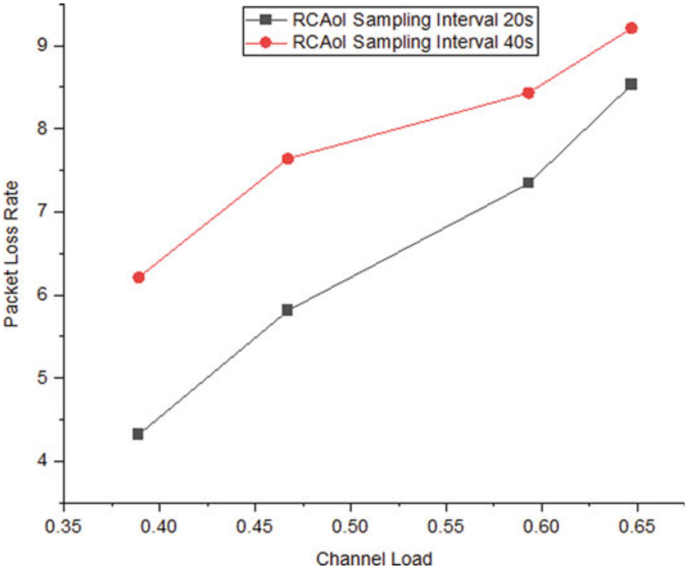


Fig. 3 Packet loss rate versus channel load

by using means of the inter arrival time of packet. The limitation of the proposed work is sensor nodes and sink association consider a fixed threshold using a bounded time in calculating AoI. The influence of multi-hop and difference in holding times and calculation of renewal function is also to be further investigated. Future scope would emphasis on sensor node isolation pertaining to age of information. The isolation process work would focus on unknown functions of bounded delay and renewal process without predetermined hierarchy of nodes in routing.

References

1. L. Huang, L.P. Qian, Age of information for transmissions over Markov channels, in *GLOBECOM 2017–2017 IEEE Global Communications Conference* (IEEE, 2017), pp. 1–6
2. R.D. Yates, S.K. Kaul, Age of information in uncoordinated unslotted updating, in *2020 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2020), pp. 1759–1764
3. V. Tripathi, E. Modiano, A whittle index approach to minimizing functions of age of information, in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (IEEE, 2019), pp. 1160–1167
4. B. Zhou, W. Saad, Minimum age of information in the internet of things with non-uniform status packet sizes. IEEE Trans. Wireless Commun. **19**(3), 1933–1947 (2019)
5. I. Krikidis, Average age of information in wireless powered sensor networks. IEEE Wireless Commun. Lett. **8**(2), 628–631 (2019)

6. Z. Chen, N. Pappas, E. Björnson, E.G. Larsson, Age of information in a multiple access channel with heterogeneous traffic and an energy harvesting node, in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (IEEE, 2019), pp. 662–667
7. K. Liu, A. Selivanov, E. Fridman, Survey on time-delay approach to networked control. Annu. Rev. Control. **48**, 57–79 (2019)
8. Y. Sun, B. Cyr, Sampling for data freshness optimization: Non-linear age functions. J. Commun. Netw. **21**(3), 204–219 (2019)
9. M. Moltafet, M. Leinonen, M. Codreanu, On the age of information in multi-source queueing models. IEEE Trans. Commun. **68**(8), 5003–5017 (2020)
10. E. Fountoulakis, N. Pappas, M. Codreanu, A. Ephremides, Optimal sampling cost in wireless networks with age of information constraints, in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (IEEE, 2020), pp. 918–923
11. S. Lee, M. Kim, J. Lee, R.H. Hsu, T.Q. Quek, Is blockchain suitable for data freshness? An age-of-information perspective. IEEE Netw **35**(2), 96–103 (2021)
12. H. Azarhava, M.P. Abdollahi, J.M. Niya, Age of information in wireless powered IoT networks: NOMA vs. TDMA. Ad Hoc Netw. **104**, 102179 (2020)
13. Y. Chen, H. Tang, J. Wang, J. Song, Optimizing age penalty in time-varying networks with markovian and error-prone channel state. Entropy **23**(1), 91 (2021)
14. M. Emara, H. ElSawy, G. Bauch, A spatiotemporal model for peak AoI in uplink IoT networks: time versus event-triggered traffic. IEEE Internet Things J. **7**(8), 6762–6777 (2020)
15. J.V. Anand, K. Sundeep, N.D.P. Rao, Underwater sensor protocol for time synchronization and data transmissions using the prediction model, in *2020 International Conference on Inventive Computation Technologies (ICICT)* (IEEE, 2020), pp. 762–766
16. C. Li, S. Li, Y. Chen, Y.T. Hou, W. Lou, AoI scheduling with maximum thresholds, in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications* (IEEE, 2020), pp. 436–445
17. O. Ayan, H.M. Gürsu, A. Papa, W. Kellerer, Probability analysis of age of information in multi-hop networks. IEEE Netw. Lett. **2**(2), 76–80 (2020)
18. J.V. Anand, S. Titus, Energy efficiency analysis of effective hydrocast for underwater communication. Int.J. Acoust. Vib. **22**(1), 44–50 (2017)
19. A.M. Bedewy, Y. Sun, R. Singh, N.B. Shroff, Optimizing information freshness using low-power status updates via sleep-wake scheduling, in *Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing* (2020), pp. 51–60
20. J.P. Champati, R.R. Avula, T.J. Oechtering, J. Gross, On the minimum achievable age of information for general service-time distributions, in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications* (IEEE, 2020), pp. 456–465

Investigations on Power-Aware Solutions in Low Power Sensor Networks

S. S. Vidhya and Senthilkumar Mathi

Abstract Power management is a very vast topic and the solution spans around hardware and software approaches. The power efficiency of IoT low-power devices becomes an important component of modern communication environments since it is very costly or impossible to replace or change device batteries in deployed environments. Energy management in sensor networks is an open challenge to researchers Hence, this article is investigated the power management solutions introduced in various literature. A detailed investigation of energy harvesting-based techniques and network-based solutions for efficient utilization of available energy is explored. The paper also highlights the recent advancement in technologies to improve battery life by adding low power components, circuitry, and low power communication protocols such as ZigBee, RPL, wirelessHART, Bluetooth low energy, and LoRAWAN. The analysis drawn from the investigation is the combination of the power provisioning approach with power control-based solutions are the best suited for designing power-efficient schemes.

Keywords Sensor network · Low power networks · ZigBee · RPL · WirelessHART · Bluetooth low energy · LoRAWAN

1 Introduction

A wireless low-power network is composed of tiny sensor nodes powered by a battery. The key component of sensor networks [1] is processing module, sensing module, transceiver module, and power unit as shown in Fig. 1. Energy management in low power nodes is achieved through various protocols to manage power supply units and efficient utilization of available energy in a sensor node. To tackle the

S. S. Vidhya · S. Mathi (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: m_senthil@cb.amrita.edu

S. S. Vidhya
e-mail: ss_vidhya@cb.students.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_69

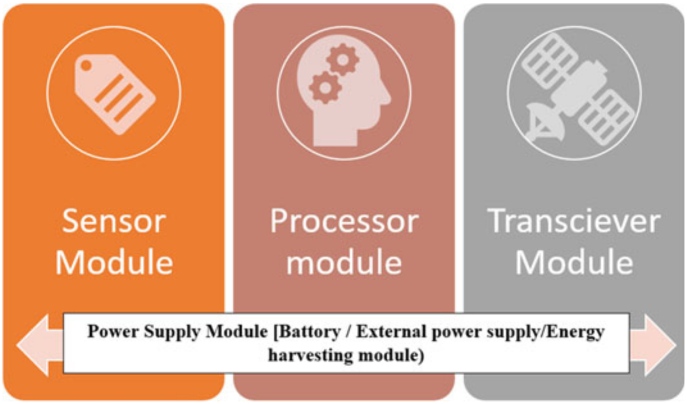
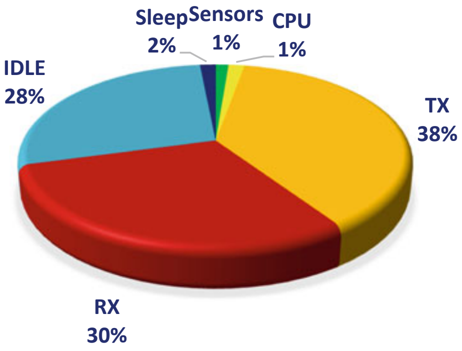

Sensor Module
Processor module
Transciever Module
Power Supply Module [Battory / External power supply/Energy harvesting module)

Fig. 1 Sensor node—architecture

energy scarcity in sensor nodes balanced management between use and supply is required. The power consumption modules in a sensor node and its rate of energy usage are shown in Fig. 2. The transceiver unit is the most energy-hungry module in a sensor node. So that the software-based solutions are mainly concentrated on the communication part.

The common power source for a sensor node is a battery. Consumer battery lithium-ion is suitable for common applications such as smart homes, parking lots, etc. But extreme environments, like a cold chain which is used to monitor frozen foods, pharmaceuticals, etc. demand bobbin-type batteries. Common batteries are prone to capacity losses (e.g., 30% loss after 1000 cycles). The degradation is directly proportional to environmental conditions such as temperature, humidity, etc. Another problem that reduces the life of a battery is self-discharge. The loss rate depends on the temperature and chemical reactions inside the cell. The constraints of battery supply led the researchers to propose alternate power provisioning techniques by using ambient energy [2]. These energy harvesting techniques have some limitations

Fig. 2 Energy consumption modules in the sensor node



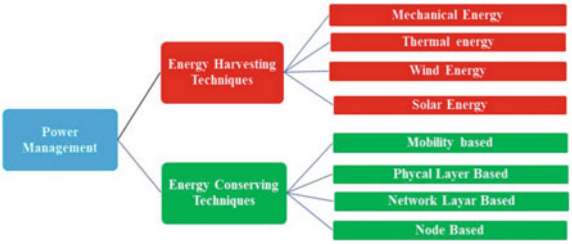


Fig. 3 Top tier division

since there are some situations where the harvesting chance is less than the required power [3].

In the present work, the energy management solutions are divided into two namely, energy harvesting and energy conserving. The top-tier taxonomy described in this literature is shown in Fig. 3. Most power management schemes considered that data collection consumes less power than data transmission. Hence, many research works happened in the area of transmit power control and routing-based solutions [4]. The challenges in network power management are as follows.

(1) Limited power source such as battery powered
(2) Sensor network deployment areas such as dense forest, mining area, and smart furnace.
(3) Dynamic network topology due to adhoc nature
(4) Node mobility in applications like cattle monitoring.

2 Energy Harvesting Techniques

One of the solutions to overcome the energy constraint issue of low power networks is energy harvesting techniques. Different sort of energy-producing techniques such as solar, wind, thermal and mechanical energy converts different sources of energy to electrical power. The general modules of a harvesting system are shown in Fig. 4. The photoelectric cells are used to convert light energy to electric energy. It will not work efficiently during cloudy and night time. The wind-based system uses turbines to produce electric power. The thermal system uses mechanical resources with an electrostatic generator. The majority of the systems use rechargeable batteries to store generated power.

The most commonly used resource is light energy. It can be artificial or natural light. This resource is a cheap, pollution less, inexhaustible and clean source of energy applicable for outdoor IoT applications. The amount of power harvesting will depend on the light intensity, atmospheric conditions and the cell area. Another parameter is the incidence of light. For full efficiency, the light source and the cell array should be perpendicular. The disadvantage of this system is (i) not suitable for indoor (ii) Depends on light and incident angle. Kansal et al. [5] conducted a study

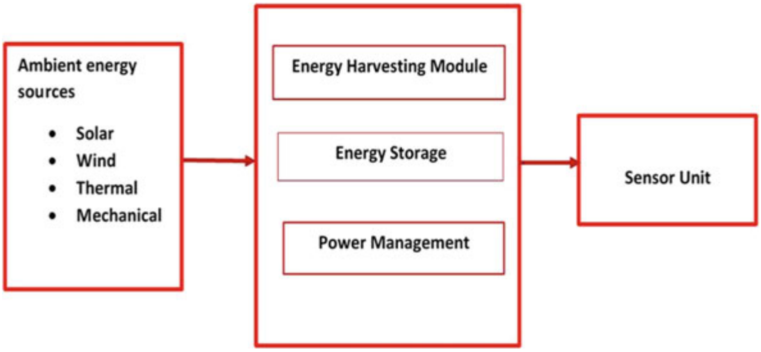


Fig. 4 Components of energy harvesting system

of voltage properties of different solar cell and associated storage devices in different environments. Similar works based on solar energy is shown in Table 1.

In most of the solar power-based works hybrid schemes are used, that is to manage the harvested energy efficiently with hardware or software-based modules. The energy generating module in cooperation with the energy management module is presented in [8–11]. A better choice of power management is based on prediction-based approaches. This can be achieved in two different ways namely, predicts the future energy needs of the communication nodes and predicts the future energy production from the harvesting sources. Table 2 summarises light energy-based hybrid schemes. Wind energy is another popular approach in the field of low power

Table 1 Energy harvesting solutions: Light energy-based

Source of energy /Node	Storage	Merit	Shortcoming
Heliomote [5]	NiMH battery	• Harvesting aware performance scaling algorithms • Select paths that have more solar availability	• Harvesting efficiency is low, single energy resource based
Ambimax [6]	Sup-capacitor	• Computes maximum power point tracking, the semiconductor is charged based on this • Harvesting is done from the combination of solar and wind	• Efficiency depends on the deployment area
Prometheus [7]	Sup-capacitor	• Intelligently manages energy transfer • Lifetime is maximized by managing mote power levels	• No MPPT circuit

Table 2 Energy harvesting solutions: Light energy-based hybrid schemes

Research work	Strategy used	Merit	Shortcoming
Abbas et al. [11]	Analytical model	Used Maximal Power Transferring Tracking (MPTT) algorithm, predicts upcoming energy, and dynamic selection of transmit power	Not discussed network-level energy management
Zhang et al. [10]	Duty cycle based	Energy efficiency is improved through opportunistic duty cycling. Adjusts duty cycle with local history	Harvesting module and management modules are independent
Dehwah et al. [9]	Routing	Dynamic programming-based routing policy optimization	High computation power
Wang, J et al. [8]	Node distribution	Distributed approximation algorithm-based solar head placement strategy and wireless charging of other nodes	Efficiency depends on the deployment area
Kosunalp [18]	Energy prediction	Q-learning based solar energy prediction Improved prediction accuracy	High computation power
Yang et al. [19]	Energy prediction	Each node energy consumption-based distributed power-aware data collection scheme	Prediction based on weathercast

networks. It requires bulky hardware, which lowers the feasibility of sensor network implementations. The proposals made in [12–17] is based on wind power. Most of the proposals are based on a prediction approach since the availability of wind will depend on the weather and the previous wind history.

Vibration-based works are widely used in WSNs, especially in body area networks [20] and aquatic sensor networks [21]. The power source used in these works is piezoelectric materials. Mechanical stress is converted into electric energy from motion or vibration. The paper proposes the conversion of vehicle vibration into electric energy in smart roadways [22]. Electrostatic based energy generation module uses the principle of distance change between two capacitive plates kept at a constant charge. During vibration, the distance between the plates changes and the energy is produced from the model.

The research works in [22–26] proposed mechanical energy based harvesting system for low power networks. Knight et al. proposed a thermal energy-based scheme, the thermostatic device is used as a power module. It is also applicable in the sensor networks deployed on water surfaces. The works based on power resources other than light is listed in Table 3.

3 Energy Conserving Techniques

The innovative sensor applications in the area of technology-assisted sensor networks are found to be the most demanding innovation in the area of technology-assisted living.

The power harvesting assisted sensor nodes are always not a good solution for all the applications because of extra hardware requirements along with the actual nodes. In this section, the various approaches proposed by researchers to manage battery dependant node's efficiency for keeping the network live are discussed. Here, the paper presents a broad idea of different approaches done in network protocols. Such a scheme can be categorized into three categories based on its characteristics as given in Fig. 5.

Data-driven approaches—energy management through data-driven concentrates on collected samples. This is achieved through two basic schemes (i) sensor data transmission based and (ii) sample data acquisition based. Data compression is one of the techniques that can be adopted to reduce the size of sending data.

Many compression algorithms are proposed for minimal centering power sensor nodes [27–29]. An alternative approach in the data reduction is, the part of sample data can be predicted so that the transmission rate towards the sink node can be lowered. Constant prediction [30], Exponential smoothing [31] and ARIMA [32] are the proposed prominent works based on this scheme.

Routing based approaches—one of the promising network-level energy management technique is energy-aware routing schemes. Variant transmitting modes and cluster management are the auspicious approaches in this area [33–35]. The sensor nodes are divided into different clusters with single or multiple CH to reduce the communication distance. This cluster level communication management can reduce energy depletion between sink node and data collection points. Most of the schemes try to select cluster head intelligently for the efficient management of energy [36–39]. One such approach is LEACH [40]. Many extensions to this protocol are proposed in [41–43]. Recently some variants of hierarchical routing such as tree-based, location-based and chain based are proposed [44, 45] for managing energy efficiency. The routing schemes such as energy-efficient routing [46], Ad-hoc on-demand distance vector routing (AODV) [47], routing protocol for low power network (RPL) [48] etc. ensures low energy consumption at communication nodes. Geographical random forwarding is a location-aware approach for selecting relay nodes [49].

Duty cycling based approaches—the network life can be increased by switching the node state between sleep and active mode. Most of the research works in this area focused on variant sleep node selection criteria. Adaptive self-configuring follows such an approach [50]. One of the efficient approaches proposed in [51, 52] is making some number of redundant mobile nodes to sleep and keeping others in

Table 3 Energy harvesting solutions: Non-solar energy resources

Scheme	Strategy used	Merit	Shortcoming
Wind energy [14]	Prediction based: weather-conditioned moving average	Used MPTT algorithm, predicts upcoming energy availability, Dynamic selection of transmit power	Single resource dependency
Wind energy [15]	Prediction based: wind energy predictor (WEP)	Wind energy availability is predicted based latest energy generation information Improved wind energy conservation rate	Single resource dependency
Wind energy [16]	Prediction based: weather forecast based duty cycle power management	Weather forecast based wind energy availability prediction and storage policies based on the predicted value	Weather forecast-based prediction
Wind energy [17]	Non-predictive MPTT	MPTT circuit-based energy optimization with energy storage circuit	Depends on wind availability
Wind energy [20]	Energy prediction	A proper rectifier bridge is selected using predicted power and follows an adaptive strategy	Energy management and harvesting modules are independent
Wind energy [21]	MPTT based	Low power management hardware circuit using MOSFETs to efficiently manage low wind speed condition MPTT is done through resistor emulation	Depends on wind availability
Mechanical [23]	Combined energy-aware interface	Vibration-based harvesting module and software-based energy flow management	Huge size
Mechanical [22]	Road vibration	Vehicle speed-based road vibrational frequency piezoelectric cantilever beams	Hardware size is high
Mechanical [24]	Human motion-based	Statistical characteristics of human motion stochastic model	Used for body area networks
Mechanical [25]	Water pressure based	Water pressure is applied to piezoelectric material and used for underwater communication	Applicable for specific applications

(continued)

Table 3 (continued)

Scheme	Strategy used	Merit	Shortcoming
Thermal [26]	Thermostatic device-based	Solar energy is converted to electric power	Applicable for specific applications

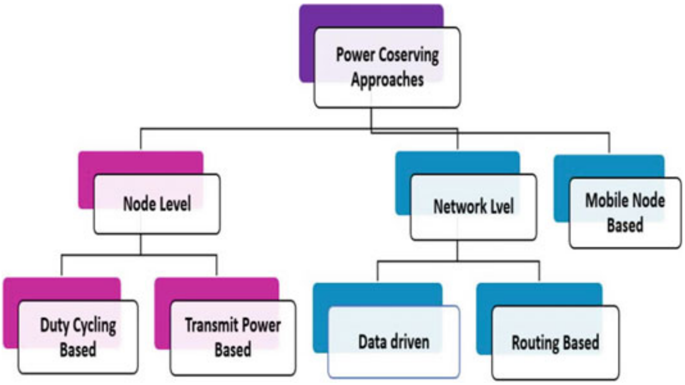


Fig. 5 Energy conserving approaches

wake-up mode. In research work, [53] proposes a sleep scheduling scheme-based linear distance approach. The sleep decision is taken by a node based on the probability which is proportional to the distance from the sink node. In the basic energy-conserving scheme [54] three states are defined namely active, sleep and idle for each node. Based on the routing or application layer information, nodes switch among these states. Dynamic sensor MAC [55] proposed a dynamic sleeping cycle based on power availability at nodes and network latency. The major problem with these schemes is latency unless the scheduling scheme chosen is not effective.

Mobile node-based approaches—the mobility in a low-power network is termed micro-mobility since the network contains few mobile nodes and the mobility environment is a limited area. In the present work, Greedy maximum residual energy [56], proposed a mobile sink node to collect sensor information from nodes. Another approach is made in [57], which is based on a mobile relay, and data collection is done through message ferries. These nodes move around the fields, collect data and send it to the destination node. In the literature [58] and [59] has done an experiment based study to show the effect of transmit power and energy consumption in adhoc networks. The results are shown in the Figs. 6 and 7. Various power conservation-based schemes discussed are summarized in Table 4 (Fig. 6).

Low power network protocols—as the popularity and applications of low power networks are increasing day by day, different standard protocols are exclusively defined for low power networks. The characteristics of the network protocols are

Fig. 6 Remaining energy for different transmit power in mobile network

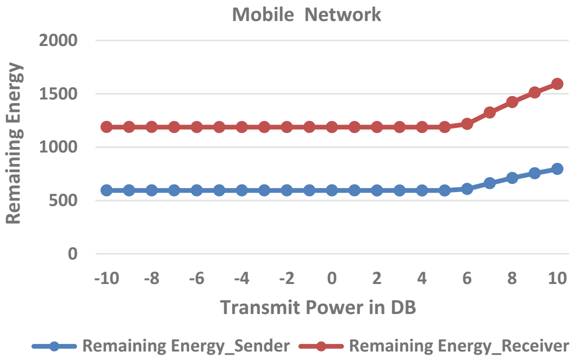
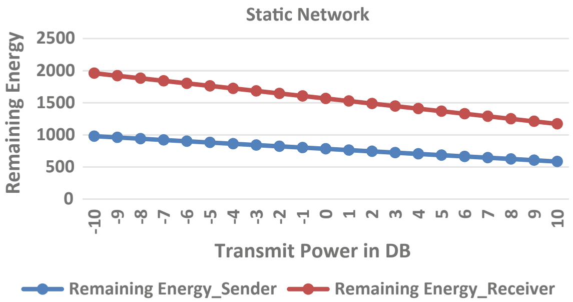


Fig. 7 Remaining energy for different transmit power in static network



given in Table 5. Conventional communication protocols are defined for sending a large amount of data (Fig. 7).

Traditionally the sensor nodes are dealing with small scalar values like pressure, temperature, humidity, etc. It becomes wastage of energy and bandwidth if traditional protocols are used to communicate with these small-sized data. So, the protocols like low power Wi-Fi, Bluetooth Low Energy, Zigbee, *Z*-Wave, and LoRaWAN [65–69] are introduced for low power communication networks.

The inferences made from the investigations are as follows. (1) The power management schemes in low power networks can be categorized into two: energy harvesting and energy conserving. (2) The energy harvesting schemes are efficient in terms of long-life networks. (3) The efficiency of energy harvesting schemes depends on (i) deployment environments and (ii) the proper management and storage of harvested energy, since the availability of common resources like light, wind, etc. depends on environmental factors. (4) The network-based energy conserving protocol demands prior knowledge of network power distribution and its consumption. (5) Duty cycling-based schemes are prone to delay because the sleep node selection and synchronization is a hurdle. (5) The power level management needs proper power level selection for efficient communication.

Table 4 Summary of energy-conserving solutions

Protocol	Scheme	Merit	Shortcoming
AP-routing [30]	Data-driven approach	Predicts next probe time based on the environmental changes	Low data accuracy
SIP [31]	Data-driven approach	Data filtering of sensed data	Applicable for engine monitoring applications
ARIMA [32]	Data-driven approach	Sensor node readings are predicted using the time series model	Data accuracy depends on the prediction
EECRP [33]	Routing based	Cluster heads are rotated based on the energy load	Node density is not considered
LEACH-AP [34]	Routing based	Nearest neighbor selection from the received power and distance formula	Needs information about node position
TEAR [36]	Routing based	Cluster head (CH) selection is based on node initial energy, residual energy, and traffic load	Node density is not considered
PSO-UFC [37]	Routing based	Swarm optimization-based CH selection	Node density is not considered
EF—LEACH [38]	Routing based	Remaining energy-based dynamic CH change	Node density is not considered
RARZ [41]	Routing based	Remaining energy-based CH selection and location-based routing	Node density is not considered
SEECH [42]	Routing based	Separates CH and relays based on node eligibility, a distance-based scheme is employed	CH energy is not considered
LA-MHR [42]	Routing based	Learning automata-based CH selection	High computation
EQR [46]	Routing based	Data prioritization and mobile sink-based	Low data accuracy
SPEED [47]	Routing based	Feedback control and geographic forwarding	Needs the knowledge of node location

(continued)

Table 4 (continued)

Protocol	Scheme	Merit	Shortcoming
GeRaF [49]	Routing based	Location and node contention-based data forwarding	Needs the knowledge of node location
ASCENT [50]	Duty cycling	Adaptively elects active nodes based on node density	Chance of prolonged active state of a single node
LDS [53]	Duty cycling	Linear distance-based sleeping node selection	Chance of prolonged sleep state of a single node
AODV and DSR [54]	Duty cycling	Application-level information and node density-based sleep node selection	intra-layer communication is needed
DSMAC [55]	Duty cycling	Analytical based duty cycle adjustment	Computation overhead
Variable-range transmission power control [60]	Node level	A model based on the routing protocol and signaling overhead	Service message overhead
DDPC [61]	Node level	Dynamic varying of transmission power	Power level is varied based on a feedback mechanism
Power Control Protocol [62]	Node level	Different power levels for service messages and data messages	Power reduction is based on service message count
EPARN [63]	Node level	Residual energy and expected future energy needs based routing	Computational overhead
CLPC [64]	Node level	Average RSSI-based routing	RSSI is environment dependant
Controlled sink mobility [56]	Mobile node based	Route selection using mixed-integer linear programming and the sink movement based on Greedy algorithm	Computational overhead

4 Conclusions

This investigation work presented different energy-aware schemes in IoT network communication. The review is done in two directions: energy harvesting approaches and energy-conserving techniques. A tabular-based summary with performance metric comparison is presented for all energy management kinds of literature discussed in this work. It is recommended that a hybrid scheme that combines energy harvesting and energy-conserving techniques, which will be the best choice

Table 5 Power consumption and properties of low power network protocols

	Power consumption			Properties
	Sleep (μ W)	Transmit (mW)	Receive (mW)	
Low power Wi-Fi [65]	300	350	270	Dynamic energy consumption, high throughput, penetration through walls, and other obstacles, range up to 1 km
Bluetooth Low Energy [66]	8	60	53	Low latency, high security, high speed, Range up to 100 m
ZigBee [67]	4	72	84	Self-organization, low cost, range up to 100 m, and IPv6 support
Z-wave [68]	3	70	65	RF-based, less radiation, highly reliable, range up to 100 m
LoRaWAN [69]	1	77	18	Wide coverage up to kilometers, a single node can handle thousands of devices, low cost, point-point communication with end devices

for increasing network lifetime. However, energy management is still an open challenge to researchers, and lots of studies are required towards the efficient functioning of low power networks.

References

1. R. Zagrouba, A. Kardi, Comparative study of energy efficient routing techniques in wireless sensor networks. Information **12**(1), 42 (2021)
2. A. Sharma, A. Kakkar, A review on solar forecasting and power management approaches for energy-harvesting wireless sensor networks. Int. J. Commun. Syst. **33**(8) (25 May 2020)
3. K.A.M. Zeinab, S.A.A. Elmustafa, Internet of things applications, challenges and related future technologies. World Sci. News **2**(67), 126–148 (2017)
4. R. Chiwariro, Quality of service aware routing protocols in wireless multimedia sensor networks: survey. Int. J. Inf. Technol. **29**, 1–2 (2020)
5. K. Lin, J. Yu, J. Hsu, S. Zahedi, D. Lee, J. Friedman, A. Kansal, V. Raghunathan, M. Srivastava, Heliomote: Enabling long-lived sensor networks through solar energy harvesting, in *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems*, (2005), pp. 309–309
6. C. Park, P. H. Chou, Ambimax: Autonomous energy harvesting platform for multi-supply wireless sensor nodes, in *2006 3rd Annual IEEE Communications Society on Sensor and ad hoc Communications and Networks*, vol. 1. (IEEE, 2006), pp. 168–177
7. X. Jiang, J. Polastre, D. Culler, Perpetual environmentally powered sensor networks, in *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005* (IEEE, 2005), pp. 463–468
8. C. Wang, J. Li, Y. Yang, F. Ye, Combining solar energy harvesting with wireless charging for hybrid wireless sensor networks. IEEE Trans. Mob. Comput. **17**(3), 560–576 (2017)
9. A.H. Dehwah, J.S. Shamma, C.G. Claudel, A distributed routing scheme for energy management in solar powered sensor networks. Ad Hoc Netw. **67**, 11–23 (2017)
10. J. Zhang, Z. Li, S. Tang, Value of information aware opportunistic duty cycling in solar harvesting sensor networks. IEEE Trans. Industr. Inf. **12**(1), 348–360 (2015)

11. M. M. Abbas, M. A. Tawhid, K. Saleem, Z. Muhammad, N. A. Saqib, H. Malik, and H. Mahmood, Solar energy harvesting and management in wireless sensor networks. Int. J. Distrib. Sensor Netw. **10**(7), 436107 (2014)
12. Y. Wu, B. Li, F. Zhang, Predictive power management for wind powered wireless sensor node. Future Internet **10**(9), 85 (2018)
13. S. Kosunalp, An energy prediction algorithm for wind-powered wireless sensor networks with energy harvesting. Energy **139**, 1275–1280 (2017)
14. A. Jushi, A. Pegatoquet, T.N. Le, Wind energy harvesting for autonomous wireless sensor networks, in *2016 Euromicro Conference on Digital System Design (DSD)* (IEEE, 2016), pp. 301–308
15. Y. Wu, W. Liu, Y. Zhu, Design of a wind energy harvesting wireless sensor node, in *2013 IEEE Third International Conference on Information Science and Technology (ICIST)* (IEEE, 2013), pp. 1494–1497
16. D. Porcarelli, D. Spenza, D. Brunelli, A. C Energy-aware approaches ammarano, C. Petrioli, and L. Benini, Adaptive rectifier driven by power intake predictors for wind energy harvesting sensor networks. IEEE J. Emerg. Sel. Top. Power Electron. **3**(2), pp. 471–482 (2014)
17. Y.K. Tan, S.K. Panda, Optimized wind energy harvesting system using resistance emulator and active rectifier for wireless sensor nodes. IEEE Trans. Power Electron. **26**(1), 38–50 (2010)
18. S. Kosunalp, A new energy prediction algorithm for energy-harvesting wireless sensor networks with q-learning. IEEE Access **4**, 5755–5763 (2016)
19. S. Yang, X. Yang, J.A. McCann, T. Zhang, G. Liu, Z. Liu, Distributed networking in autonomic solar powered wireless sensor networks. IEEE J. Sel. Areas Commun. **31**(12), 750–761 (2013)
20. F. Akhtar, M.H. Rehmani, Energy harvesting for self-sustainable wireless body area networks. IT Prof. **19**(2), 32–40 (2017)
21. R. Kumar, D. Bhardwaj, M. K. Mishra, Enhance the lifespan of underwater sensor network through energy efficient hybrid data communication scheme, in *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)* (IEEE, 2020), pp. 355–359
22. Y. Song, C.H. Yang, S.K. Hong, S.J. Hwang, J.H. Kim, J.Y. Choi, S.K. Ryu, T.H. Sung, Road energy harvester designed as a macro-power source using the piezoelectric effect. Int. J. Hydrogen Energy, **41**(29), pp. 12 563–12 568, (2016)
23. T. Ruan, Z.J. Chew, M. Zhu, Energy-aware approaches for energy harvesting powered wireless sensor nodes. IEEE Sens. J. **17**(7), 2165–2173 (2017)
24. S. Zhang, A. Seyedi, Statistical models for harvested power from human motion. IEEE J. Sel. Areas Commun. **33**(8), 1667–1679 (2015)
25. A. Bereketli, S. Bilgen, Remotely powered underwater acoustic sensor networks. IEEE Sens. J. **12**(12), 3467–3472 (2012)
26. C. Knight, J. Davidson, Thermoelectric energy harvesting as a wireless sensor node power source, in *Active and Passive Smart Structures and Integrated Systems 2010*, vol. 7643. (International Society for Optics and Photonics, 2010), pp. 76431E
27. R. Middya, N. Chakravarty, M.K. Naskar, Compressive sensing in wireless sensor networks–a survey. IETE Tech. Rev. **34**(6), 642–654 (2017)
28. R. Soua, P. Minet, A survey on energy efficient techniques in wireless sensor networks, in *2011 4th Joint IFIP Wireless and Mobile Networking Conference* (IEEE, 2011), pp. 1–9
29. T. Aneeth, R. Jayabarathi, Energy-efficient communication in wireless sensor network for precision farming, in *Artificial Intelligence and Evolutionary Computations in Engineering Systems* (Springer, 2016), pp. 417–427
30. I. Ragoler, Y. Matias, N. Aviram, Adaptive probing and communication in sensor networks, in *International Conference on Ad-Hoc Networks and Wireless* (Springer, 2004), pp. 280–293
31. D.J. McCorrie, E. Gaura, K. Burnham, N. Poole, R. Hazelden, Predictive data reduction in wireless sensor networks using selective filtering for engine monitoring, in *Wireless Sensor and Mobile Ad-Hoc Networks.* (Springer, 2015), pp. 129–148
32. A. Abd Manaf, S. Sahibuddin, R. Ahmad, S. M. Daud, E. El-Qawasmeh, Informatics engineering and information science, in *Conference proceedings ICIEIS* (Springer, 2011), pp. 42

33. J. Shen, A. Wang, C. Wang, P. C. Hung, C.-F. Lai, An efficient centroid-based routing protocol for energy management in WSN-assisted IoT. IEEE Access **5**, 18 469–18 479 (2017)
34. I. Sohn, J.-H. Lee, S.H. Lee, Low-energy adaptive clustering hierarchy using affinity propagation for wireless sensor networks. IEEE Commun. Lett. **20**(3), 558–561 (2016)
35. S.S. Aswanth, A. Gokulakannan, C.S. Sibi, R. Ramanathan, Routing in wireless sensor network based on swarm intelligence, in *2019 3rd International Conference on Trends in Electronics and Informatics* (IEEE, 2019), pp. 502–508
36. D. Sharma, A.P. Bhondekar, Traffic and energy aware routing for heterogeneous wireless sensor networks. IEEE Commun. Lett. **22**(8), 1608–1611 (2018)
37. T. Kaur, D. Kumar, Particle swarm optimization-based unequal and fault tolerant clustering protocol for wireless sensor networks. IEEE Sens. J. **18**(11), 4614–4622 (2018)
38. T.M. Behera, U.C. Samal, S.K. Mohapatra, Energy-efficient modified leach protocol for IoT application. IET Wireless Sens. Syst. **8**(5), 223–228 (2018)
39. D. Sharma, A. Ojha, A.P. Bhondekar, Heterogeneity consideration in wireless sensor networks routing algorithms: a review. J. Supercomput. **75**(5), 2341–2394 (2019)
40. S. Pal, D. Bhattacharyya, G.S. Tomar, T.-h. Kim, Wireless sensor networks and its routing protocols: a comparative study, in *2010 International Conference on Computational Intelligence and Communication Networks* (2010), pp. 314–319
41. R.N. Jadoon, W. Zhou, W. Jadoon, I. Ahmed Khan, Rarz: ring-zone based routing protocol for wireless sensor networks. Appl. Sci. **8**(7), 1023, (2018)
42. S. Tanwar, S. Tyagi, N. Kumar, M.S. Obaidat, La-mhr: learning automata based multilevel heterogeneous routing for opportunistic shared spectrum access to enhance lifetime of WSN. IEEE Syst. J. **13**(1), 313–323 (2018)
43. M. Tarhani, Y.S. Kavian, S. Siavoshi, Seech: Scalable energy efficient clustering hierarchy protocol in wireless sensor networks. IEEE Sens. J. **14**(11), 3944–3954 (2014)
44. X. Liu, Atypical hierarchical routing protocols for wireless sensor networks: A review. IEEE Sens. J. **15**(10), 5372–5383 (2015)
45. A.C.J. Malar, M. Kowsigan, N. Krishnamoorthy, S. Karthick, E. Prabhu, K. Venkatachalam, Multi constraints applied energy efficient routing technique based on ant colony optimization used for disaster resilient location detection in mobile ad-hoc network. J. Ambient Intell. Humaniz. Comput. 1–11 (2020)
46. B. Nazir, H. Hasbullah, Energy efficient and QoS aware routing protocol for clustered wireless sensor network. Comput. Electr. Eng. **39**(8), 2425–2441 (2013)
47. T. He, J.A. Stankovic, C. Lu, T. Abdelzaher, Speed: A stateless protocol for real-time communication in sensor networks, in *23rd International Conference on Distributed Computing Systems, 2003. Proceedings* (IEEE, 2003), pp. 46–55
48. S. Vidhya, S. Mathi, Investigation of next generation internet protocol mobility-assisted solutions for low power and lossy networks. Procedia comput. Sci. **143**, 349–359 (2018)
49. L. Cheng, J. Niu, J. Cao, S.K. Das, Y. Gu, Qos aware geographic opportunistic routing in wireless sensor networks. IEEE Trans. Parallel Distrib. Syst. **25**(7), 1864–1875 (2013)
50. A. Cerpa, D. Estrin, Ascent: Adaptive self-configuring sensor networks topologies. IEEE Trans. Mob. Comput. **3**(3), 272–285 (2004)
51. F. Cuomo, A. Abbagnale, E. Cipollone, Cross-layer network formation for energy-efficient IEEE 802.15. 4/zigbee wireless sensor networks. Ad Hoc Netw. **11**(2), 672–686 (2013)
52. M. Li, B. Yang, A survey on topology issues in wireless sensor network. in *ICWN. Citeseer* (2006), pp. 503
53. J. Deng, Y.S. Han, W.B. Heinzelman, P.K. Varshney, Scheduling sleeping nodes in high density cluster-based sensor networks. Mob. Net. Appl. **10**(6), 825–835 (2005)
54. R.S. Bhadoria, G.S. Tomar, S. Kang, Proficient energy consumption aware model in wireless sensor network. Int. J. Multimedia Ubiquit. Eng. **9**(5), 27–36 (2014)
55. P. Lin, C. Qiao, X. Wang, Medium access control with a dynamic duty cycle for sensor networks, in *2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No. 04TH8733)*, vol. 3, (IEEE, 2004), pp. 1534–1539

56. S. Basagni, A. Carosi, E. Melachrinoudis, C. Petrioli, Z.M. Wang, Controlled sink mobility for prolonging wireless sensor networks lifetime. Wireless Netw. **14**(6), 831–858 (2008)
57. W. Zaho, A message ferrying approach for data delivery in sparse mobile ad hoc networks-macro, in *proceedings of MobiHoc'04* (2004)
58. M. Krunz, A. Muqattash, S.J. Lee, Transmission power control in wireless ad hoc networks: challenges, solutions and open issues. IEEE Netw. **18**(5), 8–14 (2004)
59. S. Blakeway, A. Kirpichnikova, M. Schaeffer, E.L. Secco, Transmission power and effects on energy consumption and performance in manet. EAI Endorsed Trans. Mob. Commun. Appl. **19**(16) 2019
60. J. Gomez, A.T. Campbell, Variable-range transmission power control in wireless ad hoc networks. IEEE Trans. Mob. Comput. **6**(1), 87–99 (2006)
61. A. Spyropoulos C.S. Raghavendra, Energy efficient communications in ad hoc networks using directional antennas, in *Proceedings. 21 Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 1, (2002), pp. 220–228.
62. D. Seth, S. Patnaik, S. Pal, EPCM–an efficient power controlled mac protocol for mobile ad hoc network. Int. J. Electron. **101**(10), 1443–1457 (2014)
63. L. Femila V. Vijayarangan, Transmission power control in mobile ad hoc network using network coding and co-operative communication, in *2014 international conference on communication and network technologies.* (IEEE, 2014), pp. 129–133
64. A.S. Ahmed, T.S. Kumaran, S.S.A. Syed, S. Subburam, Cross-layer design approach for power control in mobile ad hoc networks. Egypt. Inform. J. **16**(1), 1–7 (2015)
65. D.M. Dobkin, B. Aboussouan Low power wi-fi™ (IEEE802. 11) for IP Smart Objects. GainSpan Corporation (2009)
66. J. Tosi, M. Taffoni, R. Santacatterina, D. Sannino, Formica, performance evaluation of bluetooth low energy: A systematic review. Sensors **17**(12), 2898 (2017)
67. Alliance Z. Zigbee alliance. WPAN industry group, http://www.zigbee.org/. The industry group responsible for the ZigBee standard and certification (2010)
68. C.W. Badenhop, S.R. Graham, B.W. Ramsey, B.E. Mullins, L.O. Mailloux, The z-wave routing protocol and its security implications. Comput. Secur. **68**, 112–129 (2017)
69. J. Haxhibeqiri, E. De Poorter, I. Moerman, J. Hoebeke, A survey of LoRaWAN for IoT: From technology to application. Sensors **18**(11), 3995 (2018)

Efficient Certificate-less Signcryption Scheme for Vehicular Ad Hoc Networks

Aminul Islam, Fahiem Altaf, and Soumyadev Maity

Abstract The number of accidents on roads has brought challenges in the rapid development of intelligent transportation systems (ITS). Gradually people have started using private vehicles in place of the public transport system which has resulted in an increased load on existing intelligent transportation infrastructure. In the near future, secure communication and better mobility will be the primary goal of ITS. A vehicular ad hoc network (VANETs) that utilizes the functionality of MANETs will be a suitable candidate for its rapid development. In this paper, we have presented a new certificate-less signcryption scheme that encrypts and signs the message parallelly in a single step due to which protocol becomes efficient and has a low computational cost.

Keywords Certificateless cryptography · Signcryption · Security · VANETs · Pairing-based cryptography · Vehicle · Intelligent transportation system

1 Introduction

The rapid growth and development of intelligent transportation system (ITS) have amplified our interest in vehicular ad hoc networks (VANETs). There are three main entities in VANET, namely on-board unit (OBU), roadside unit (RSU) and a trusted authority (TA). Each vehicle is equipped with OBU to provide communication between vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). Both V2V and V2I communication are based on IEEE802.11p standard such as dedicated short-range communication (DSRC). These technologies help to make wireless communication efficient and easier. In a vehicular network, each vehicle broadcasts

A. Islam (✉) · F. Altaf · S. Maity
Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, India
e-mail: pcl2015006@iiita.ac.in

S. Maity
e-mail: soumyadev@iiita.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_70

a safety message containing vehicle-specific information that needs to be rectified before transmitting to the other vehicles, and if the personal information (e.g. vehicle identity, location, etc.) is accessible to all the other entities in the network, then the malicious vehicle may use that information for its own benefits. So, it implies the need of security services.

The basic security requirement of VANETs includes data authentication, confidentiality, integrity as well as non-repudiation. Various security schemes have been introduced for this purpose. But those schemes have various drawbacks [18]. For example, certificate management is a common problem in traditional PKI-based schemes. To overcome it, identity-based public-key cryptosystem was introduced. But later, it was observed that this cryptosystem is facing the key escrow problem. That gave the idea to introduce a new public-key system called certificate-less public-key cryptosystem. But some of those schemes have high computation and communication cost and also are vulnerable to various attacks that are explained in later sections.

1.1 Motivation and Contribution

Normally, in vehicular communication, all the traffic-related messages are required to be signed but not encrypted. But in some cases, broadcasting can result in leak of private data to a malicious user. Therefore, we have to ensure that such messages are not accessible to any unauthorized user in the network. To overcome this problem, we propose a signcryption scheme which fulfils all the above-mentioned security requirements for VANETs. Signcryption is a combination of digital signature and encryption which ensures authenticity as well as confidentiality of all vehicle-related statistics. The specific contribution of this proposed work is as follows:

- First, we have designed an efficient generic signcryption scheme based on the bilinear pairing. This scheme is made efficient and lightweight by excluding the use of map-to-point operation.
- Secondly, we have adapted this generic scheme for VANETs and proposed an efficient privacy-preserving signcryption scheme. The proposed scheme satisfies all the basic security requirements in VANET including authentication, confidentiality, integrity and non-repudiation.
- Finally, we have compared our protocol with other existing protocols in terms of computational cost, communication cost as well as signcryption and unsigncryption cost.

The rest structure of the paper is as follows: Sect. 2 presents related work, and Sect. 3 presents preliminaries. In Sect. 4, we introduce our generic signcryption scheme, and in Sect. 5, we present the signcryption scheme for VANETs. Section 6 presents the performance analysis, and finally, in Sect. 7, we present the concluding remarks.

2 Related Work

Various researchers tried to improve the security and privacy of VANETs with the use of various cryptographic techniques. One of the traditional schemes is a public-key infrastructure (PKI)-based scheme which is widely used while performing authentication during vehicular communication. A scheme designed by Raya et al. [15] proposes a PKI-based scheme in which TA generates certificates along with a large number of key pairs which results in complex certificate management. Also, a scheme by Lu et al. [1] discusses a method to overcome delay in the anonymous authentication of safety messages. In this protocol, TA generates system parameters like a public key, private keys and certificates and utilizes them with bilinear pairing in order to sign and verify the traffic-related statistics. For any movement of vehicles in the RSU, it will have an anonymous certificate, which results in the prohibition of various attacks since the actual identification of vehicles is unrevealed. However, these protocols suffer from high computational, communication as well as storage overhead due to the use of certificates along with a large number of key pairs.

With an aim to overcome the complexity of certificate management, an identity-based cryptographic technique is proposed by Adi Shamir et al. [19] which uses unique ID like username, mobile number or email as a public key. A scheme by Zhang et al. [2] suggests an ID-based batch verification scheme for both V2V as well as V2I communication. However, this scheme requires high computation cost and suffers from a key escrow attack. Also, in [3] by Jianhong et al., an improved secure batch verification is proposed for the purpose of vehicular communication. But it requires three bilinear pairing operations and has a high computational cost. Another ID-based scheme by Zhang et al. [4] magnifies pseudonymous authentication. This scheme preserves privacy against traceability attacks. However, the protocol does not support batch verification.

A local ID-based anonymous message authentication protocol was proposed by Wang et al. [5]. In this protocol, both bilinear pairing and batch verification are used to secure the vehicular ad hoc network. Liu et al. [6] propose a V2I communication protocol that utilizes both bilinear pairing and batch verification in order to secure the ad hoc network. However, the verification of messages is not performed efficiently in this protocol. In ID-based cryptography, the generation of private keys is performed by a private key generator (PKG). If somehow PKG is compromised, then whole VANETs will become vulnerable which leads to the key escrow problem. To resolve this issue, Al-Riyami and Paterson [7] first introduced certificate-less cryptography in the year 2003. Later, Ming et al. [8] came up with a certificate-less authentication scheme for secure authentication which neither uses a map-to-point nor bilinear pairing operations due to which the average message delay as well as average message loss ratio gets reduced significantly leading to improved efficiency. Also, in vehicular communication, for the purpose to fulfil the security requirements of all the traffic-related statistics, a message is either first signed and then encrypted or vice versa. At the receiver end, while decrypting, the contrary steps are followed when compared to the sender's end. But due to the use of more machine cycles, these schemes suffer from high computational cost.

A hybrid signcryption authentication protocol is proposed by Han et al. [9] in vehicular ad hoc network (VANET), but the scheme does not ensure privacy preservation as real identities of vehicles can be revealed in a network. Hong et al. [10] overcome this problem and presented a signcryption scheme based on ID-based cryptography. This scheme neither uses a map-to-point operation nor bilinear pairing and hence, is efficient when compared to the previous one. Various researchers also proposed identity-based signcryption schemes in [11–14], but all of these schemes suffer from key escrow attacks. To further enhance privacy and security, a certificate-less signcryption scheme was proposed by Li et al. [16] which does not use point multiplication operation. Later, an efficient generalized certificate-less scheme is proposed by Zhang et al. [13] without the use of bilinear pairing. Also, Karati et al. [17] proposed a certificate-less signcryption that supports public verifiability.

3 Preliminaries

3.1 System Model

Typically, the network layer of VANETs is divided into two layers—upper layer and a lower layer. The upper layer includes trusted authority (TA) and application server, whereas lower layer includes the vehicles (with pre-installed OBUs) and roadside unit (RSU) as shown in Fig. 1.

- On-Board Unit(OBU): The OBU is basically mounted on a vehicle. It consists of GPS, micro-sensors, etc. The OBU sends all the traffic-related statistics to the RSU

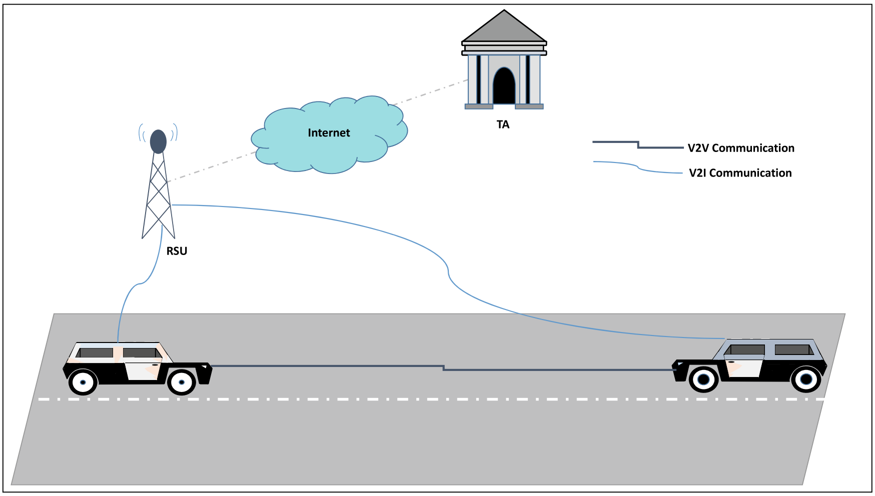


Fig. 1 VANET architecture

and communicates with other vehicle's OBU using IEEE802.11p communication standard. DSRC is an open-source wireless communication protocol that supports in-range vehicle communication. OBU includes a tamper-proof device attached with it with the responsibility of storing secret information of vehicles.

- Roadside Unit (RSU): RSU allows vehicles to connect to the Internet using DSRC protocol. It basically verifies all the traffic-related statistics and re-broadcast the traffic notifications to in-range vehicles.
- Trusted Authority (TA): It plays a very important role in VANET and is responsible for registering all OBUs and RSUs. It generates all the system parameters and maintains the database of all the registered vehicles. In addition, TA also generates pseudo-identity of all the registered vehicles so that the real-identity of a vehicle remains anonymous to all the other entities in the network.

3.2 Design Goals

In order to design a reliable and efficient vehicular network, the proposed protocol must provide efficient security services. The proposed scheme should considers all basic security requirement which are mentioned below:

- Message Authentication: It guarantees that the source entity of a traffic-related message is authentic.
- Message Confidentiality: It promises the avoidance of illegal access by any illegitimate user in the network.
- Integrity: Integrity ensures that the data is not modified or altered by any malicious entity during the communication.
- Non-Repudiation: It guarantees that the vehicles cannot deny the origin of messages later which have already been sent.
- Privacy: It ensures that the real-identity of a vehicle is anonymous to all the other entities in the network. The true identity of the vehicle can only be traced by TA.

3.3 Bilinear Pairing

The bilinear pairing can be defined as $\hat{e}$: $G_1 X G_1 \longrightarrow G_2$ where G_1 denotes an additive group, and G_2 denotes the multiplicative group both of order q. It has three main properties as given below

- Bilinearity: $\hat{e}(aP, bQ) = \hat{e}(bP, aQ) = e(P, Q)^{ab}$ where P,$Q \in G_1$ and a,$b \in Z_q^*$
- Non-degeneracy: $\hat{e}(P, Q) \neq 1$ where 1 is element G_1.
- Computability: To compute $\hat{e}(P, Q)$, there exist an efficient algorithm.

Table 1 Notations

Symbol	Description
G_1, G_2	q ordered two cyclic groups
MSK	Master secret key
P, Q	Two generator $\in G_1$
m	Plaintext message
$Params$	Global parameters
S_k'	Partial private key
ID_A, ID_B	Real-identity of vehicle A and B, respectively
$\text{PID}_A, \text{PID}_B$	Pseudo-identity of vehicle A and B, respectively
P_{kA}, P_{kB}	Public key of vehicle A and B, respectively
s_{kA}, s_{kB}	Private key of vehicle A and B, respectively
x_{kA}, x_{kB}	Secret key of vehicle A and B, respectively

3.4 Complexity Assumptions

In this section, we define Diffie–Hellman-based complexity assumptions supporting the security of proposed protocol. Here, $P \in G_1$ and $a, b, c \in Z_q^*$

Computational Diffie–Hellman (CDH): Given a tuple (P, aP, bP), it is computationally hard to compute abP.

Bilinear Diffie–Hellman Inversion (BDHI): Given a tuple $(P, aP, a^2P, a^3 P, \ldots, a^q P)$, it is compuationally hard to compute $\hat{e}(P, P)^{a^{-1}}$.

Strong Diffie–Hellman (SDH): Given a tuple $(c, P, aP, a^2P, a^3P, \ldots, a^q P)$, it is computationally hard to compute $\hat{e}(P, P)^{(c+a)^{-1}}$.

4 Proposed Generic Signcryption Scheme

In this section, we introduce a generic signcryption scheme with the aim to achieve the confidentiality as well as message authentication simultaneously. List of notations employed in this article is given in Table 1.

Set-up: The responsibility of executing this phase is with trusted authority (TA). The security parameter λ is chosen by TA that generates q ordered cyclic groups, namely G_1 and G_2. Here, G_1 depicts additive cyclic group, and G_2 depicts a multiplicative cyclic group. In addition to that P, $Q \in G_1$ is an arbitrary generator for bilinear pairing $\hat{e}(., .)$. Furthermore, TA chooses hash function H: $\{0, 1\}^l \longrightarrow Z_q^*$, α_1 and $\alpha_2 \in Z_q^*$. Thereafter, it computes the following parameters:

$$U = \alpha_1.P$$

$$V = \frac{\alpha_2}{1+\alpha_1}.Q$$

$$W = \hat{e}(P, Q)^{\alpha_2}$$

Finally, it declares the global parameters as $Param=\{q, G_1, G_2, P, U, V, W, \hat{e}, H\}$ and stores the master secret key $(MSK) = (\alpha_1, \alpha_2, Q)$.

Extract Partial private key: All users have their unique identity (ID). TA generates the partial private key (S'_k) of a user by using its MSK, $Params$ and a user's ID.

$$S'_k = \frac{\alpha_2}{\alpha_1 + ID}.Q$$

After calculating partial private key, the trusted authority returns S'_k to the user via a secured channel.

Generate user secret value: After collection of partial private key, a user first validates it as per following equation $W \doteq \hat{e}(S'_k, ID.P + U)$. Thereafter, that device chooses a secret key $x \in Z_s^*$ to generate public key $P_k = x.S'_k$.

Generate Full Private key: After calculating the secret key(x) & partial private key, a user sets its full private key $s_k = x$.

Signcrypt: In this phase, if one of participating user A desires to send message to another user B, then both users need to compute some parameters. User A computes the following parameters using $Params$, Message $(m) \in \{0, 1\}^l$, its secret key (x_A), public key of user B (P_{kB}), its own public key (P_{kA}) as well as its private key (s_{kA}). This algorithm is summarized in Fig. 2. User A randomly picks $t_1, t_2 \in Z_q^*$ and computes

$$\Psi = W^{t_2}.$$

$$\beta_A = H(ID_A||ID_B||P_{kA}||P_{kB})$$

$$s_1 = m.\Psi$$

$$S_2 = P_{kB}(t_2 + \beta_A t_1)$$

$$S_3 = \frac{t_1}{x_A}(U + P.ID_A)$$

$$s_4 = H(H(m)||ID_A||ID_B||\Psi)$$

Finally, user A transmits $\{s_1, S_2, S_3, s_4\}$ to user B via a public channel as signcryption of m.

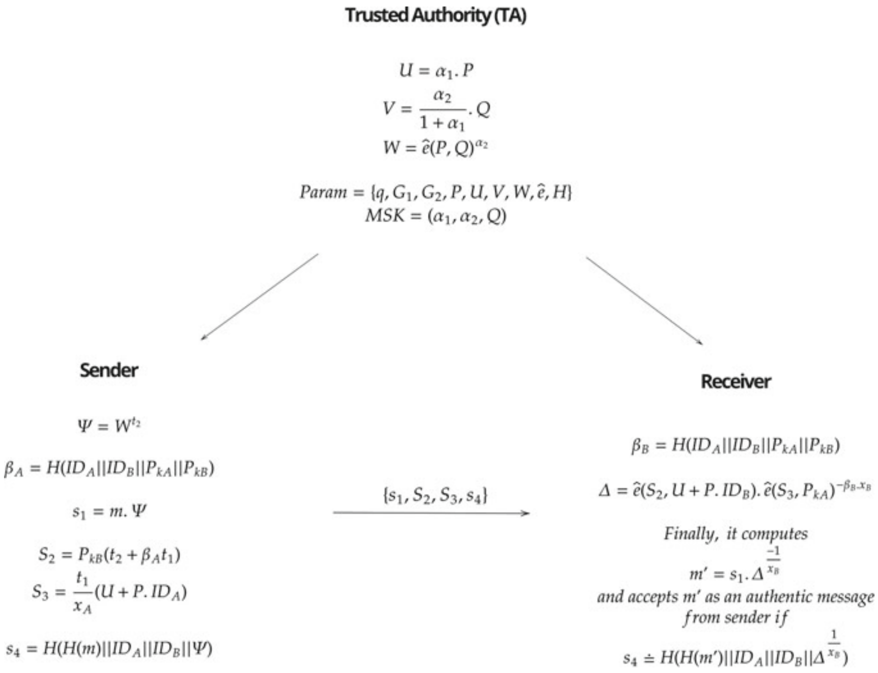

Trusted Authority (TA)
$U = \alpha_1.P$
$V = \frac{\alpha_2}{1+\alpha_1}.Q$
$W = \hat{e}(P,Q)^{\alpha_2}$
$Param = \{q, G_1, G_2, P, U, V, W, \hat{e}, H\}$
$MSK = (\alpha_1, \alpha_2, Q)$
Sender
$\Psi = W^{t_2}$
$\beta_A = H(ID_A||ID_B||P_{kA}||P_{kB})$
$s_1 = m.\Psi$
$S_2 = P_{kB}(t_2 + \beta_A t_1)$
$S_3 = \frac{t_1}{x_A}(U + P.ID_A)$
$s_4 = H(H(m)||ID_A||ID_B||\Psi)$
$\{s_1, S_2, S_3, s_4\}$
Receiver
$\beta_B = H(ID_A||ID_B||P_{kA}||P_{kB})$
$\Delta = \hat{e}(S_2, U + P.ID_B).\hat{e}(S_3, P_{kA})^{-\beta_B.x_B}$
Finally, it computes
$m' = s_1.\Delta^{\frac{-1}{x_B}}$
and accepts m' as an authentic message from sender if
$s_4 \doteq H(H(m')||ID_A||ID_B||\Delta^{\frac{1}{x_B}})$

Fig. 2 Generic signcryption scheme

Unsigncrypt: On receiving the tuple $\{s_1, S_2, S_3, s_4\}$ from user A, user B runs the unsigncrypt algorithm in order to read the message m from user A. This algorithm is summarized in Fig. 2. User B computes

$$\beta_B = H(ID_A||ID_B||P_{kA}||P_{kB})$$

$$\Delta = \hat{e}(U + P.ID_B, S_2).\hat{e}(S_3, P_{kA})^{-\beta_B.x_B}$$

Finally, it computes

$$m' = s_1.\Delta^{\frac{-1}{x_B}}$$

and accepts m' as an authentic message from user A if

$$s_4 \doteq H(H(m')||ID_A||ID_B||\Delta^{\frac{1}{x_B}})$$

Correctness Proof:

We know

$$\begin{aligned}\Delta &= \hat{e}(U + P.ID_B, S_2).\hat{e}(S_3, P_{kA})^{-\beta_B.x_B}\\ &= \hat{e}(\alpha_1.P + P.ID_B, P_{kB}(t_2 + \beta_A.t_1)).\hat{e}(\frac{t_1}{x_A}(U + P.ID_A), x_A.S_k^{'})^{-\beta_B.x_B}\\ &= \hat{e}(P, Q)^{x_B\alpha_2 t_2}.\hat{e}(P, Q)^{x_B\alpha_2\beta_A t_1}.\hat{e}(P, Q)^{-x_B\alpha_2\beta_A t_1}\\ &= \hat{e}(P, Q)^{x_B\alpha_2 t_2}\end{aligned}$$

Therefore,

$$\begin{aligned}m^{'} &= s_1.\Delta^{\frac{-1}{x_B}}\\ &= s_1.\hat{e}(P, Q)^{x_B\alpha_2 t_2.\frac{-1}{x_B}}\\ &= s_1.\hat{e}(P, Q)^{-\alpha_2 t_2}\\ &= s_1.W^{-t_2}\\ &= m.\Psi.W^{-t_2}\\ &= m.W^{t_2}.W^{-t_2}\\ &= m\end{aligned}$$

5 Proposed Signcryption Scheme for VANETs

As we know that privacy and security have always been a major concern for VANETs. So, to provide better security, now we need to adopt the above generic scheme. In this section, we will propose a privacy-preserving signcryption scheme suitable for VANETs. Our protocol consists of the following phases:

Set-up: The responsibility of executing this phase lies with a trusted authority (TA). On input of a security parameter λ, TA generates q ordered cyclic groups, namely G_1 and G_2 and bilinear map $\hat{e}$. Here, G_1 denotes additive cyclic group, and G_2 denotes a multiplicative cyclic group. In addition, TA randomly picks two generators $P, Q \in G_1$. TA also chooses collision-resistant hash functions H: $\{0, 1\}^l \longrightarrow Z_q^*$ and $H_1 : G_1 \longrightarrow \{0, 1\}^l$ as well as randomly picks α_1 and $\alpha_2 \in Z_q^*$. Thereafter, it computes the following parameters:

$$U = \alpha_1.P$$

$$V = \frac{\alpha_2}{1 + \alpha_1}.Q$$

$$W = \hat{e}(P, Q)^{\alpha_2}$$

Finally, it declares the global parameters as $Param = \{q, G_1, G_2, P, U, V, W, \hat{e}, H, H_1\}$ and stores the master secret key $(MSK) = (\alpha_1, \alpha_2, Q)$.

Pseudo-identity Generation: In this phase, a vehicle generates its pseudo-identities with the help of TA. First, a vehicle picks a random number $r \in Z_q^*$, computes $\mathrm{PID}_1 = rP$ and sends the tuple $\{\mathrm{PID}_1, ID\}$ to TA, where ID is the real-identity of the vehicle. On receiving the above tuple, TA first validates the identity of the vehicle. If the real identity of the vehicle is not valid, then rejection of that vehicle is considered; otherwise TA computes

$$\mathrm{pid}_2 = ID \oplus H_1(\alpha_1.\mathrm{PID}_1 || U).$$

The tracing of real identities of the vehicle can only be performed by TA as follows:

$$ID = \mathrm{pid}_2 \oplus H_1(\alpha_1.\mathrm{PID}_1 || U)$$

Extract Partial Private Key: After the generation of pseudo-identity, generation of partial private key (S'_k) for a vehicle by using MSK, $Params$ and ID is carried out by TA as follows:

$$S'_k = \frac{\alpha_2}{\alpha_1 + \mathrm{PID}}.Q$$

After calculating partial private key, TA returns S'_k and $\mathrm{PID} = (\mathrm{PID}_1, \mathrm{pid}_2)$ to the vehicle through a secure channel.

Generate Secret Value: After collection of partial private key, a vehicle first validates and accepts the partial private key by checking if $W \doteq \hat{e}(S'_k, \mathrm{PID}.P + U)$. Thereafter, it randomly selects a secret key $x \in Z_s^*$ and sets its public key $P_k = x.S'_k$.

Generate Full Private Key: The full private key of a vehicle is constituted by partial private key S'_k and secret key $s_k = x$. Vehicle then stores these keys in its tamper-proof OBU.

Signcrypt: In this phase, if one of the participating vehicles desires to send a message to another vehicle or RSU, then both entities need to compute some parameters. The sender vehicle will be refereed as vehicle A and receiving vehicle as vehicle B. Vehicle A computes the following parameters using the global parameters $Params$, message $m \in \{0, 1\}^n$, its pseudo-identity (PID_A), pseudo-identity of vehicle B (PID_B), secret key of vehicle A (x_A), public key of vehicle B (P_{kB}), public key of vehicle A (P_{kA}) and private key of vehicle A (s_{kA}). This algorithm is summarized in Fig. 3. Vehicle A randomly selects $t_1, t_2 \in Z_q^*$ and computes

$$\Psi = W^{t_2}.$$

$$\beta_A = H(\mathrm{PID}_A || \mathrm{PID}_B || P_{kA} || P_{kB})$$

$$s_1 = m.\Psi$$

$$S_2 = P_{kB}(t_2 + \beta_A t_1)$$

$$S_3 = \frac{t_1}{x_A}(U + P.\mathrm{PID}_A)$$

$$s_4 = H(H(m)||\mathrm{PID}_A||\mathrm{PID}_B||\Psi)$$

Finally, vehicle A sends $\{s_1, S_2, S_3, s_4, T\}$ to vehicle as signcryption of message (m), where T is the timestamp.

Unsigncrypt: On recieving the signcryptext $\{s_1, S_2, S_3, s_4, T\}$ from vehicle A, vehicle B first validates the timestamp as $T_a - T_d \geq \Delta T$, where T_a and T_b are the arrival time and departure time, respectively. ΔT is the fixed value. If timestamp is not valid, then rejection of that signcryptext is considered; otherwise, vehicle B runs the unsigncrypt algorithm to extract the message m from vehicle A. This algorithm is summarized in Fig. 3. For that, vehicle B computes

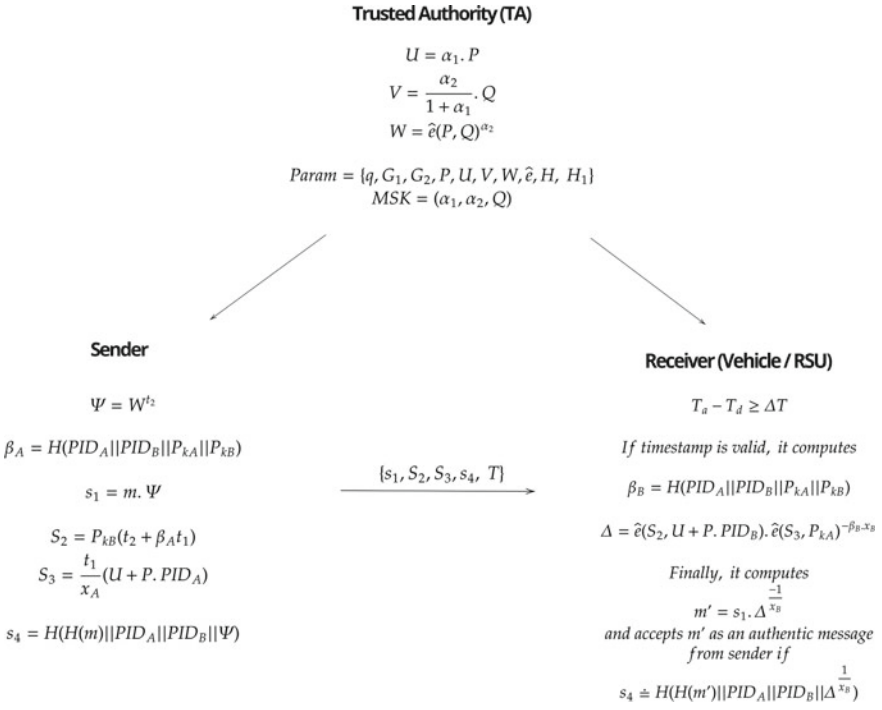

Trusted Authority (TA)
$U = \alpha_1.P$
$V = \frac{\alpha_2}{1+\alpha_1}.Q$
$W = \hat{e}(P,Q)^{\alpha_2}$
$Param = \{q, G_1, G_2, P, U, V, W, \hat{e}, H, H_1\}$
$MSK = (\alpha_1, \alpha_2, Q)$
Sender
$\Psi = W^{t_2}$
$\beta_A = H(PID_A||PID_B||P_{kA}||P_{kB})$
$s_1 = m.\Psi$
$S_2 = P_{kB}(t_2 + \beta_A t_1)$
$S_3 = \frac{t_1}{x_A}(U + P.PID_A)$
$s_4 = H(H(m)||PID_A||PID_B||\Psi)$
$\{s_1, S_2, S_3, s_4, T\}$
Receiver (Vehicle / RSU)
$T_a - T_d \geq \Delta T$
If timestamp is valid, it computes
$\beta_B = H(PID_A||PID_B||P_{kA}||P_{kB})$
$\Delta = \hat{e}(S_2, U + P.PID_B).\hat{e}(S_3, P_{kA})^{-\beta_B \cdot x_B}$
Finally, it computes
$m' = s_1.\Delta^{\frac{-1}{x_B}}$
and accepts m′ as an authentic message from sender if
$s_4 \doteq H(H(m')||PID_A||PID_B||\Delta^{\frac{1}{x_B}})$

Fig. 3 Signcryption scheme for VANETs

$$\beta_B = H(\mathrm{PID}_A||\mathrm{PID}_B||P_{kA}||P_{kB})$$

$$\Delta = \hat{e}(U + P.\mathrm{PID}_B, S_2).\hat{e}(S_3, P_{kA})^{-\beta_B.x_B}$$

Finally, it computes the

$$m' = s_1.\Delta^{\frac{-1}{x_B}}$$

and accepts m' as a valid message if

$$s_4 \doteq H(H(m')||\mathrm{PID}_A||\mathrm{PID}_B||\Delta^{\frac{1}{x_B}})$$

Correctness Proof:

We know

$$\begin{aligned}\Delta &= \hat{e}(U + P.\mathrm{PID}_B, S_2).\hat{e}(S_3, P_{kA})^{-\beta_B.x_B}\\ &= \hat{e}(\alpha_1.P + P.\mathrm{PID}_B, P_{kB}(t_2 + \beta_A.t_1)).\hat{e}(\frac{t_1}{x_A}(U + P.\mathrm{PID}_A), x_A.S_k^{'})^{-\beta_B.x_B}\\ &= \hat{e}(P, Q)^{x_B\alpha_2 t_2}.\hat{e}(P, Q)^{x_B\alpha_2\beta_A t_1}.\hat{e}(P, Q)^{-x_B\alpha_2\beta_A t_1}\\ &= \hat{e}(P, Q)^{x_B\alpha_2 t_2}\end{aligned}$$

Therefore,

$$\begin{aligned}m^{'} &= s_1.\Delta^{\frac{-1}{x_B}}\\ &= s_1.\hat{e}(P, Q)^{x_B\alpha_2 t_2.\frac{-1}{x_B}}\\ &= s_1.\hat{e}(P, Q)^{-\alpha_2 t_2}\\ &= s_1.W^{-t_2}\\ &= m.\Psi.W^{-t_2}\\ &= m.W^{t_2}.W^{-t_2}\\ &= m\end{aligned}$$

Batch Unsigncrypt: When vehicle B receives multiple signcryptions $\{s_{1_i}, S_{2_i}, S_{3_i}, s_{4_i}, T_i\}$ of n messages from n different vehicles, it first validates the timestamp T_i as discussed earlier, where $i = 1, 2, 3, \ldots, n$. Vehicle B then computes

$$\beta_{B_i} = H(\mathrm{PID}_i||\mathrm{PID}_B||P_{k_i}||P_{kB})$$

$$\prod_{i=1}^{n}\Delta_i = \hat{e}(\sum_{i=1}^{n} S_{2_i}, U + P.\mathrm{PID}_B).\hat{e}(\sum_{i=1}^{n} S_{3_i}, \sum_{i=1}^{n}\beta_{B_i}.P_{k_i})^{-x_B}$$

Finally, it computes the

$$m_i' = s_{1_i}.\prod_{i=1}^{n}\Delta_i^{\frac{-1}{x_B}}$$

and accepts all messages as valid if

$$s_{4_i} \doteq H(H(m_i')||\mathrm{PID}_i||\mathrm{PID}_B||\Delta_i^{\frac{1}{x_B}})$$

Correctness Proof: We know

$$\begin{aligned}\prod_{i=1}^{n} \Delta_i &= \hat{e}(\sum_{i=1}^{n} S_{2_i}, U + P.\mathrm{PID}_B).\hat{e}(\sum_{i=1}^{n} S_{3_i}, \sum_{i=1}^{n} \beta_{B_i}.P_{k_i})^{-x_B}\\ &= \hat{e}(P, \sum_{i=1}^{n} t_{2_i}.Q)^{x_B\alpha_2}\end{aligned}$$

Therefore,

$$\begin{aligned}m_i' &= s_{1_i}.\prod_{i=1}^{n} \Delta_i^{\frac{-1}{x_B}}\\ &= s_{1_i}.W_i^{-t_{2_i}}\\ &= m_i.\Psi_i.W_i^{-t_{2_i}}\\ &= m_i.W_i^{t_{2_i}}.W_i^{-t_{2_i}}\\ &= m_i\end{aligned}$$

6 Performance Analysis

In this section, the performance evaluation of the proposed scheme on the basis of computation cost and communication cost is measured. It may be noted that while measuring performance evaluation, implementation of the protocol is done using ecpy library in Python programming language. The comparison of our protocol with other protocols [13, 16, 20, 21] is shown in Table 2.

6.1 Computational Cost

The proposed protocol requires only three bilinear pairing out of which one bilinear pairing is used in the setup phase, whereas the other two pairings are used in unsigncryption phase. The protocol requires only two exponentiation operation out of which one is used in signcryption and the other is used in unsigncryption phase. It may be noted that our protocol requires 45.8 ms and 67.36 ms during the signcryption and unsigncryption phase, respectively. It is lesser than the protocol proposed

Table 2 Comparisons of the computational cost and communication cost

Schemes	Computational cost			Communication cost
	Signcrypt (in ms)	Unsigncrypt (in ms)	Total Cost (in ms)	
Li et al. [16]	84.9	106.4	191.3	$\|G_2\| + 2 * \|Z_q^*\| + 2 * \|G_1\|.$
Zhang et al. [13]	102.7	82.2	184.9	$2 * \|G_q\| + \|m\|$
Zhou et al. [20]	91.5	131	222.5	$2 * \|G_1\| + \|G_1\| + 2 * \|G_1\|$
Jin et al. [21]	131.5	180.5	312	$2 * \|G_1\| + \|G_1\| + 2 * \|G_1\|$
Our scheme	45.8	67.36	113.16	$\|G_2\| + \|Z_q^*\| + 2 * \|G_1\|.$

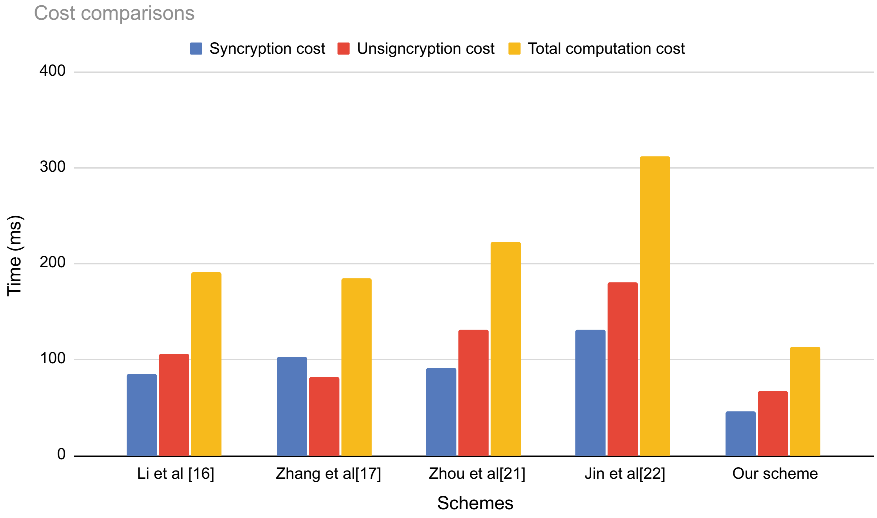


Fig. 4 Computational cost comparisons

by Li et al. [16] which requires 191.3 ms, Zhang et al. [13] which requires 184.9 ms, Zhou et al. [20] which requires 222.5 ms and Jin et al. [21] which requires 312 ms as shown in the comparison chart in Fig. 4.

6.2 Communication Cost

The communication cost is derived from the cost of the total number of variables transmitted during the communication. In the proposed scheme, four signcrypt $\{S_1, s_2, S_3, s_4\}$ variables are basically transmitted during V2V communication.

The cost of transmitting these four signcrypt $\{S_1, s_2, S_3, s_4\}$ variables is measured as $|G_2| + |Z_q^*| + 2 * |G_1|$. Table 2 also depicts the communication cost comparison of proposed protocol with other schemes.

7 Conclusion

In this paper, we have introduced an efficient generic signcryption scheme based on the bilinear pairing that can be applied in any network like IoT, etc. After that, we have applied our generic signcryption scheme in the vehicular ad hoc network and proposed an efficient signcryption scheme. It ensures all the security requirements which signcrypt a message in a single logical step. Due to this, the performance of the signcryption method has been increased. At the end, we have shown a performance analysis of our proposed protocol based on the computational and communicational cost which indicates that our protocol is more efficient as compared to an existing protocol. In future, it will open more ways for researchers to enhance the security and privacy of the network by taking advantage of certificate-less signcryption schemes.

References

1. R. Lu, X. Lin, H. Zhu, P. Ho, X. Shen, ECPP: efficient conditional privacy preservation protocol for secure vehicular communications, in *IEEE INFOCOM 2008—The 27th Conference on Computer Communications* (Phoenix, AZ, 2008), pp. 1229–1237
2. C. Zhang, P.-H. Ho, J. Tapolcai, On batch verification with group testing for vehicular communications. Wirel. Netw. **17**(8), 1851 (2011)
3. Z. Jianhong, X. Min, L. Liying, On the security of a secure batch verification with group testing for VANET. Int. J. Netw. Secur. **16**, 355–362 (2014)
4. Y. Zhang, L. Yang, S. Wang, An efficient identity-based signature scheme for vehicular communications, in *2015 11th International Conference on Computational Intelligence and Security (CIS)* (Shenzhen, 2015), pp. 326–330
5. S. Wang, N. Yao, LIAP: a local identity-based anonymous message authentication protocol in VANETs. Comput. Commun. **112**, 154–164 (2017)
6. J. Liu, et al., An efficient privacy preserving batch authentication scheme with deterable function for VANETs, in *International Conference on Network and System Security* (Springer, Cham, 2018)
7. S.S. Al-Riyami, G.P. Kenneth, Certificateless public key cryptography, in *International conference on the theory and application of cryptology and information security* (Springer, Berlin, Heidelberg, 2003)
8. M. Yang, H. Cheng, Efficient certificateless conditional privacy-preserving authentication scheme in VANETs. Mob. Inf. Syst. **2019** (2019)
9. Y. Han, et al. SCHAP: the aggregate signcryption based hybrid authentication protocol for VANET, in *International Conference on Internet of Vehicles* (Springer, Cham, 2014)
10. Z. Hong, F. Tang, W. Luo. Privacy-preserving aggregate signcryption for vehicular Ad Hoc networks, in *Proceedings of the 2nd International Conference on Cryptography, Security and Privacy* (2018)
11. A. Yin, H. Liang, Certificateless hybrid signcryption scheme for secure communication of wireless sensor networks. Wirel. Person. Commun. **80**(3), 1049–1062 (2015)

12. Z.H.O.U. Caixue, Certificateless signcryption scheme without random oracles. Chin. J. Electr. **27**(5), 1002–1008 (2018)
13. Z. Bo, Z. Jia, C. Zhao. An efficient certificateless generalized signcryption scheme. Secur. Commun. Netw. **2018** (2018)
14. I. Ali et al., *An Efficient Hybrid Signcryption Scheme with Conditional Privacy-Preservation for Heterogeneous Vehicular Communication in VANETs* (IEEE Trans. Veh, Technol, 2020)
15. R. Maxim, J.-P. Hubaux. The security of vehicular ad hoc networks, in *Proceedings of the 3rd ACM workshop on Security of Ad Hoc and Sensor Networks* (2005)
16. F. Li, Y. Han, C. Jin, Certificateless online/offline signcryption for the Internet of Things. Wirel. Netw. **23**(1), 145–158 (2017)
17. K. Arijit, C.-I. Fan, R.-H. Hsu, Provably secure and generalized signcryption with public verifiability for secure data transmission between resource-constrained IoT devices. IEEE IoT. J. **6**(6), 10431–10440 (2019)
18. A. Islam , S. Ranjan, A.P. Rawat, S. Maity, A comprehensive survey on attacks and security protocols for VANETs, in *Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems*, vol. 171, ed. by H.S. Saini, R. Sayal, A. Govardhan, R. Buyya (Springer, Singapore, 2021). https://doi.org/10.1007/978-981-33-4543-0_62
19. A. Shamir, Identity-based cryptosystems and signature schemes, in *Advances in Cryptology. CRYPTO 1984. Lecture Notes in Computer Science*, vol. 196, ed. by G.R. Blakley, D. Chaum (Springer, Berlin, Heidelberg, 1985). https://doi.org/10.1007/3-540-39568-7_5
20. F. Zhou, Y. Li, and Y. Ding, Practical V2I secure communication schemes for heterogeneous VANETs. Appl. Sci. **9**(15) (2019)
21. C. Jin, G. Chen, C. Yu, J. Shan, J. Zhao, Y. Jin, An efficient heterogeneous signcryption for smart grid. PLOS ONE **13**(12), 1–16 (2018)

Detection of Attacks Using Multilayer Perceptron Algorithm

S. Dilipkumar and M. Durairaj

Abstract Intrusion detection system (IDS) refers to a software system that alerts the network or computer activities and identifies the occurrence of any mischievous operations. New issues such as malware and worms are added as the internet is bursting into civilization. Henceforth, the users will utilize various techniques such as password cracking, where unencrypted text detection is used to cause system vulnerabilities. Therefore, the users require some protection mechanism to protect their device against the intruders. The main purpose of this research work is to include a comparative analysis on intrusion detection by using different machine learning and deep learning techniques. Various machine learning techniques have been used to develop IDS, and they are Back Propagation Neural Network (BPN), Feed Forward Neural Network (FNN), Recurrent Neural Network (RNN) and Multilayer Perceptron (MLP) based on real time neural network datasets such as IDS datasets and UNSW datasets. The proposed system can be analyzed in terms of error rate and accuracy values.

Keywords Intrusion detection · Deep learning · Network traffic · Neural networks · Multilayer perceptron

1 Introduction

Machine learning algorithms are often classified as supervised or unsupervised. Supervised algorithms rely on a software scientist or data analyst with machine learning expertise to improve each input and expected output, further providing assessment on prediction accuracy along with the training algorithm time. Data scientists regulate which variables or characteristics of the model should be analyzed and used for predictions development. Once the training has been done, the algorithm must adapt what has been taught to new data. There is no need to practice non-supervised algorithms with optimal performance. Rather by using an insistent

S. Dilipkumar (✉) · M. Durairaj
Department of CSE, Bharathidasan University, Trichy, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_71

technique named deep learning to analyse the details to get the final outcomes [1–3]. Uncontrolled learning algorithms are reversed for more complicated processing function than supervised learning process, along with recognition of images, speech-to-text and generation of languages. Those are the neural networks operate by integrating coaching data across multiple samples and automatically recognizing often subtle correlations amongst several variables. The algorithm was once trained using its confederation bank to illustrate new data. Only such an algorithm became viable in the age of big data, because they depend upon large quantities of data for coaching.

Algorithms for machine learning are identified as supervised or unsupervised results. Semi-supervised machine learning algorithms can be decreased over the supervised and unsupervised learning process by the use of both the labelled and unlabelled data for training process. Usually, there will be a small amount of labelled data and an oversized amount of unlabelled data. The systems that can be used in this method are able to appreciably improve learning accuracy [4, 5].

The feedback is required for the agent to look out which action is best; this is often remarked because the reinforcement signals. Machine learning attempts to work on the vast quantities of information. Although it typically provides quick and reliable results to allow spot lucrative possibilities or dangerous threats, where it often needs overtime and money to properly coach it. The combination of machine learning with AI and intellectual technologies will analyze the large amounts of data in a simpler way.

2 Methodology

The IDS is often distinguished on the premise of where the detection will be performed and also the way or by which technique is being detected. The IDS is classified into two segments, one being network intrusion detection system, and thus, another is host intrusion detection system. The first system mentioned helps within the analysis the arriving networking traffic and although the HIDS functioning is dependent on operating system operation. The key conditions of information mining on IDS, which were primitively discussed, were called clustering and classification. As there is no initial label on clustering problem data collection, the item generated for the clustering algorithm has been allocated with identical data records to the same class.

The packet's action was called a traditional class or peculiar class to keep up with existing data's features and characteristics. This works on burrowing from data previously clustered in classification. This means the content is labelled. Classification can well be a technique for processing knowledge that is used to analyze a collection of information. Classification plays an important role in classifying information within this field of continuous streaming data [6, 7]. Many algorithms like decision tree, rule-based induction, Bayesian network, genetic algorithm, etc., are accustomed to classify the data. In existing framework implement, machine learning techniques like Random forest, Naive Bayes, Support Vector machine algorithms are implemented to detect the intrusion from network datasets. Existing framework could also be provide high warning and low accuracy [8–10].

3 Novel Intelligent Based Ids

Deep learning has become a popular topic in the world of machine learning. It is subfield of machine learning in artificial neural networks. Using deep learning approach within the applying area, we are able to process on great amount of things required to be trained. Process is placed on numerous data points. Deep learning learns different features from the information. If the pile of knowledge is on the market, it can reduce the system performance. For achieving better accuracy in terms of performance, deep learning is considered as compatible learning mechanism. Learning varies in three major categories, i.e. supervised, semi-supervised and unsupervised. Here, the intrusion detection is implemented with relevance to the deep learning approach. Intrusion is the term, which might offend the security of automatic processing system or network. Another technique is intrusion detection, which remains tactic to investigate intrusion. Intrusion detection technique is assessed based on two methods, i.e. anomaly detection or misuse detection. Security has become a very important issue for computer systems with the rapid expansion of the computer networks over the past decade [11–14].

Specific machine learning based approaches for intrusion detection systems are being introduced in recent years. This research provides an introduction to intrusion detection through networks. A Multilayer Perceptron (MLP) is used to track interference assisted by an off-line approach to analytics. The classifying records are of two general classes—normal and assault—this analysis requires unraveling a multi-class problem because the neural network is still detecting the threat. MLP is often usually a stacked feed forward network equipped with static back propagation (BP). Such networks carried out positive analysis of static patterns through countless deployments.

3.1 Pre-Processing

Pre-processing data is a key step in the [data extraction] process. The expression “garbage in, garbage out” especially applies to machine learning and data processing projects. Methods of data collection are usually poorly regulated, dominant to out-of-range values, impossible combinations of data, missing values, etc. Resolve data for which these problems have not been carefully tested, it may yield unclear results for the process. Therefore, first and foremost, the representation and consistency of information are before an experiment is performed. When there is much irrelevant information present, then the discovery of knowledge is focused during the training process. Preparation and filtering of data steps can take considerable time interval. During this module, eliminate the irrelevant and missing values in uploaded datasets.

3.2 Classification

As the proliferation of network activity growth and confidential information on network infrastructure increases, more and more companies become vulnerable to a wider kind of attack. It is essential to protect network systems from interference, interruption and other suspicious behaviours from undesirable attackers. The network should be protected from intruders, disruption and other suspicious behaviours is important. A Multilayer perceptron (MLP) can be a type of feed forward artificial neural network. An MLP subsists on a network of at least three layers of nodes. In addition to the input nodes, any node may be a neuron that uses a nonlinear activation function. MLP's method of studying used for training data sets, which is called as the back propagation method. The multiple layers and the nonlinear activation differentiate between Multilayer Perceptron and linear perceptron. It can discern data, which cannot be separated linearly. Multilayer perceptron is consistently referred to as neural networks called "vanilla," particularly once they always had a secret layer. A perceptron may be a linear classifier; that this is an input classification algorithm by splitting a line from two groups. In python, select the option classify and select the feature options to execute the class attribute provided by Multilayer perceptron. Data usually is a property of vector x, multiply by a wand of weights added to a bias (Fig. 1).

$$B : y = w * x + b.$$

4 Experimental Work

The proposed research work uses the KDD Cup Dataset, which is used to test intrusion detection problems. The dataset may be a series of assumed crude TCP dump data on a LAN over a span of 9 weeks training data was collected from seven weeks of network traffic to around 5 million connections records and about 2 million connection records were given fortnight of testing data. And also upload the UNSW datasets. During this phase, we will upload the network datasets within the sort of CSV file. The accuracy, false positive ratio and training time of samples are compared with traditional algorithms. [http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html] [15] (Fig. 2; Table 1).

Proposed algorithm has better accuracy rates compared to existing ones due to the use of Multilayer perceptron technique that will be helpful in training of input of attack signature which will be fed as input to the architecture (Fig. 3).

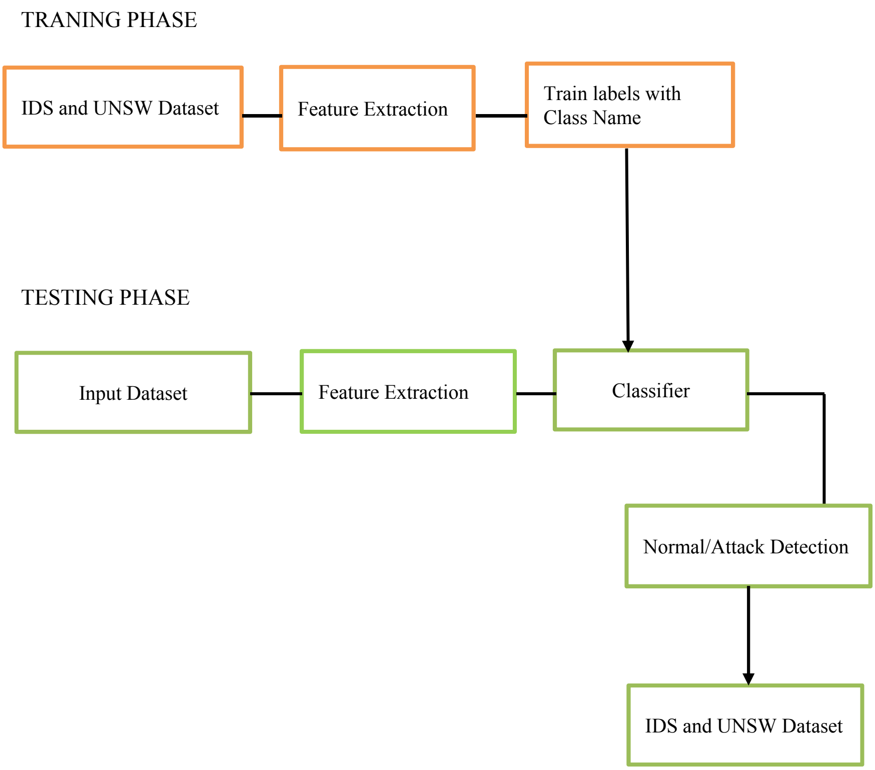


Fig. 1 Proposed intelligent based IDs engine

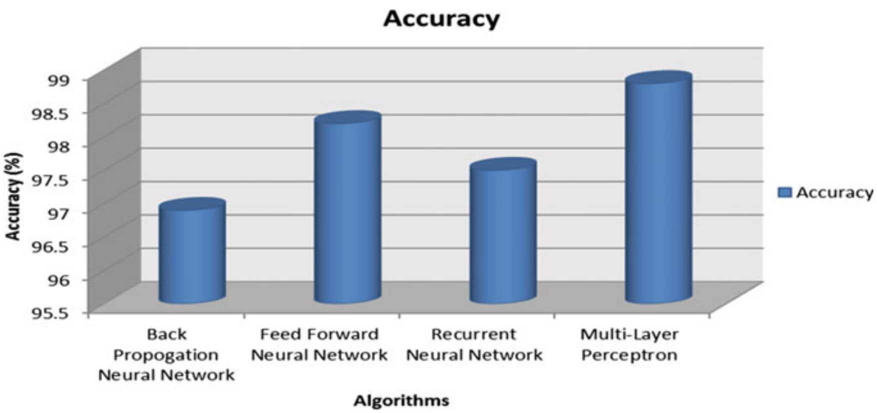


Fig. 2 Accuracy of various algorithms

Table 1 Accuracy of various algorithms

Algorithm	Accuracy
Back propagation neural network (BPN)	96.9
Feed forward neural network (FF)	98.2
Recurrent neural network (RNN)	97.5
Multilayer perceptron (MLP)	98.8

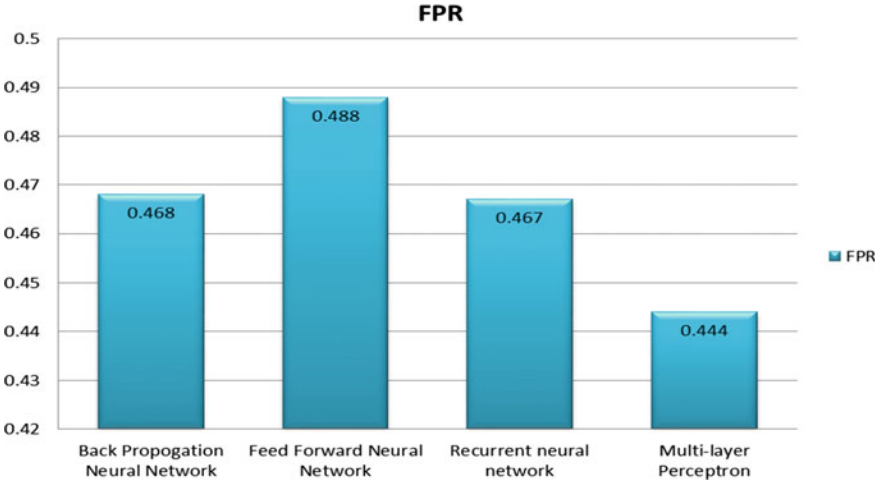


Fig. 3 False positive ratio

The proposed method achieves less false positive ratio by properly identifying the correct attack based signatures based on the proper classification algorithms deployed to analyse the input collected form the network. The proposed method is able to achieve better FPR even within the presence of malicious nodes in the network (Fig. 4).

The percentage of identifying correct attacks based on the training and testing samples based on the algorithm used is a key factor. Here in our proposed technique, the use of MLP enhances key parameter TPR to decide which one are malicious or benign.

Our tests use the KDD Cup Dataset that is used to test intrusion detection problems. The dataset may be a series of assumed crude TCP dump data on a LAN over a span of 9 weeks training data was collected from seven weeks of network traffic to around 5 million connections records and about 2 million connection records were given fortnight of testing data. And also, upload the UNSW datasets. During this module, we will upload the network datasets within the sort of CSV file (Fig. 5).

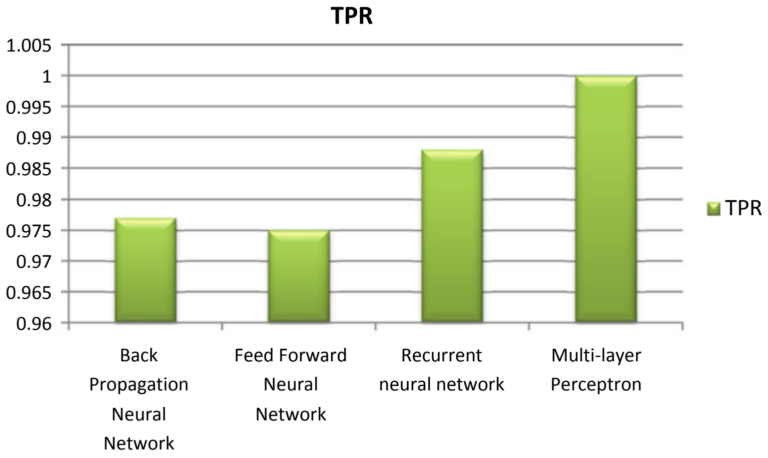


Fig. 4 True positive rate

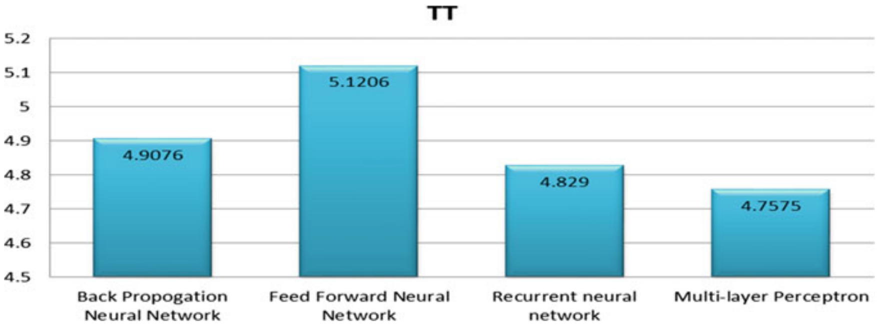


Fig. 5 Training time of samples

5 Conclusion

Detection of intrusion plays a very important role within network security, since the applications and their behaviour change every day. In recent years, network intrusion detection has been thoroughly researched, and a number of techniques are introduced including machine learning and deep learning techniques. As a result, there increased the requirement for accurate classification of the network flows. Here, we’ve got proposed deep learning model using Multilayer perceptron with feature selection for the accurate classification of intrusion detection. During this project, we demonstrated the development of a lightweight neural network capable of detecting intrusion from the network in real time. We also provided more insight into the methodologies used by various classification schemes in the process. We addressed possible analysis and optimization techniques that can be extended to other supervised methods of machine learning. We also outlined a quick method of identifying key attributes that supported the connection weights within the neural network and compared the deep learning algorithm (MLP) with BPN, FNN and RNN algorithm. Comparison done based error metrics (False positive rate, True Positive Rate, Training Time) and Accuracy metrics. From the above comparison, MLP is often provided less error metrics and highest accuracy 98.9% than the prevailing machine learning algorithms.

References

1. R.C. Staudemeyer, Applying long short-term memory recurrent neural networks to intrusion detection. S. Afr. Comput. J. **56**(1), 136–154 (2015)
2. Y. Xin, L. Kong, Z. Liu, Y. Chen,Y. Li, H. Zhu, C. Wang, Machine learning and deep learning methods for cybersecurity. IEEE Access (2018)
3. N. Hubballi, Pairgram: modeling frequency information of look ahead pairs for system call based anomaly detection, in *Communication Systems and Networks(COMSNETS), 2012 Fourth International Conference* (IEEE, 2012), pp. 1–10
4. S. Venkatraman, M. Alazab, Use of data visualisation for zero-day malware detection. Secur. Commun. Netw. **1728303**, 13 (2018). https://doi.org/10.1155/2018/1728303
5. H. Kayacik, A.N. Zincir-Heywood, M.I. Heywood, Selecting features for intrusion detection: a feature relevance analysis on KDD 99 intrusion detection datasets, in *Proceedings of the Third Annual Conference On Privacy, Security and Trust 2005, PST 2005* (DBLP, 2005)
6. Z. Jiong, M. Zulkernine, A. Haque, Random forests-based network intrusion detection systems. IEEE Trans. Syst. Man Cyber. Part C (Appl. Rev.) **38**(5), 649–659 (2008)
7. M. Alazab, S. Huda, J. Abawajy, R. Islam, J. Yearwood, S. Venkatraman, R. Broadhurst, A hybrid wrapper-filter approach for malware detection. J. Netw. **9**(11), 2878–2891 (2014)
8. T. Kim, B. Kang, M. Rho, S. Sezer, E.G. Im, A multimodal deep learning method for android malware detection using various features. IEEE Trans. Inf. Forensics Secur. **14**(3), 773–788 (2019)
9. R. Thanuja, A. Umamakeswari, Black hole detection using evolutionary algorithm for IDS/IPS in MANETs. Clust. Comput. **22**(2), 3131–3143 (2019)
10. R. Thanuja, A. Umamakeswari, Unethical network attack detection and prevention using fuzzy based decision system in mobile Ad-hoc networks. J. Electr. Eng. Technol. **13**(5), 2086–2098 (2018)
11. A. Saracino, D. Sgandurra, G. Dini, F. Martinelli, Madam: effective and efficient behavior-based android malware detection and prevention. IEEE Trans. Dependable Secure Comput. **15**(1), 83–97 (2018)
12. S. Naseer, Y. Saleem, S. Khalid, M.K. Bashir, J. Han, M.M. Iqbal, K. Han, Enhanced network anomaly detection based on deep neural networks. IEEE Access **6**, 48231–48246 (2018)
13. S. Smys, B. Abul, W. Haoxiang, Hybrid intrusion detection system for internet of things (IoT). J. ISMAC **2**(04), 190–199 (2020)
14. V. Suma, W. Haoxiang, Optimal key handover management for enhancing security in mobile network. J. Trends Comput. Sci. Smart Technol. (TCSST) **2**(4), 181–187 (2020)
15. [http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html]

IoT-Based Smart Monitoring of online Transformer

G. Gajenthiran, Chindamani Meyyappan, J. Vishnuprakash, R. Arjun, T. Viswak Sena, Yegappan Sethu, and R. Sharan Prasanna

Abstract The IOT based online transformer monitoring records the electrical quantities such as load current, transformer oil, earth resistance, voltage, power factor, ambient temperatures. This online monitoring system integrates a Wi-Fi module (ESP8266), with stand alone Arduino microcontroller and sensor packages placed at site, those electrical quantities are stored in ADC and received parameters are processed and recorded in system memory, it sends the saved data to cloud, the cloud processing system gets data from microcontroller, and it integrates the details and represents the details in graphical manner. If there is any abnormality with transformer, graphical representation shows it in monitoring device.

Keywords Online transformer · Internet of Things (IOT) Wi-Fi module · Short circuit current · Open circuit voltage · Earth leakage current · Arduino microcontroller

G. Gajenthiran · C. Meyyappan (✉) · J. Vishnuprakash · R. Arjun · T. Viswak Sena · Y. Sethu · R. Sharan Prasanna
Sri Ramakrishna Engineering College, Coimbatore, India
e-mail: Chindamani.meyyappan@srec.ac.in

G. Gajenthiran
e-mail: gajenthiran.1703158@srec.ac.in

J. Vishnuprakash
e-mail: Vishnuprakash.1703164@srec.ac.in

R. Arjun
e-mail: arjun.1703066@srec.ac.in

T. Viswak Sena
e-mail: viswaksena.1703154@srec.ac.in

Y. Sethu
e-mail: yegappan.1603153@srec.ac.in

R. Sharan Prasanna
e-mail: sharanprasannaa.1903121@srec.ac.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_72

1 Introduction

To monitor and make statistical analysis of the distribution transformer and record the parameters like voltage, current, vibration, temperature, earth resistance and float level in cloud computing technique, it has the Arduino microcontroller, Wi-Fi module and sensor to monitor the parameters. Hardware consists of power supply, Arduino microcontroller, Wi-Fi module, sensors, potential transformer and current transformer. Various faults like over-voltage, over-current, increase in temperature, oil-level, humidity, etc., are steadily monitored by microcontroller and passes regular health information through Wi-Fi component, and the data can be accessed remotely by a web application.

For online monitoring of distribution transformer, most power companies utilize SCADA (supervisory control and data acquisition) system, but the SCADA system is an expensive proposition. At present, it is being done physically by hand, impersonate and the disadvantage is, it fails to update the transformer inappropriate status such as overloads voltage and current and overheating of transformer oil and windings. So the lifespan and performance get reduced. Vadirajacharya et al. [1] explain how the system and the operator are communicated. Kumar et al. [2] explain power distribution transformer stations located in outskirts are supervised by using wireless GPRS. GR47—as the date communication module, LPC2132—as main processor are used at control station and how they are communicated is also investigated [3]. This paper demonstrates the need for a modern load scheme and introduces the new technology of intelligent load shedding [4]. This paper describe the design and implementation of an automatic method of protecting transformer as an alternative to the fuse protection technique [5]. This study describe the design and implementation of microcontroller based system for protecting transformer [6]. This study provides a practical guidance to help maintenance personnel for the best utilization of the power transformer in electric utilities current maintenance practice is reviewed and an effective maintenance planning has been proposed in order to prevent the failures [7]. In this case study the fault and defects that occurred in 400kV/220kV/132kV/66kV sub station can be found by DGA. Dissolved Gas Analysis is a technique used to assess incipient faults of the transformer [8]. The objective of this paper is mainly to analyze available data from DGA to arrive at conclusion of type of fault to further investigate or to save the transformer and correlate the fault found with DGA data [9] introduced as a comprehensive protection technique for power transformer against the traditional differential protection schemes, current transformer error and saturation, tap changer operation, energising inrush current, over-fluxing etc. [10] explains the fault detection and rectifying techniques in multilevel inverter using different methodologies.

.

2 Methodology

IOT is used to acquire real-time data of transformer. To monitor the transformer temperature sensor, potential transformer, vibration sensor, one float and current transformer are used. The analog values are delivered to Arduino controller of ATmega328 family which has inbuilt analog to digital converter. And the methodology is well explained in block diagram in Figs. 1 and 2 show the simulation done in Proteus circuit design. The measured sensors values are sent sequentially to a Wi-Fi module under TCP IP protocol to a dedicated IP, and the data's are displayed in the web interfaced personal computer system.

2.1 Transformer

The voltage is stepped down from (0–230 V) to (0–9 V) using potential transformer in which the primary turns should be greater than secondary turns. Depending upon the wire gauge, the voltage in the secondary decreases meanwhile the current or ampere rating on the secondary side increases and vice versa. The output from the secondary coil is fed to the next part rectifier of the circuit.

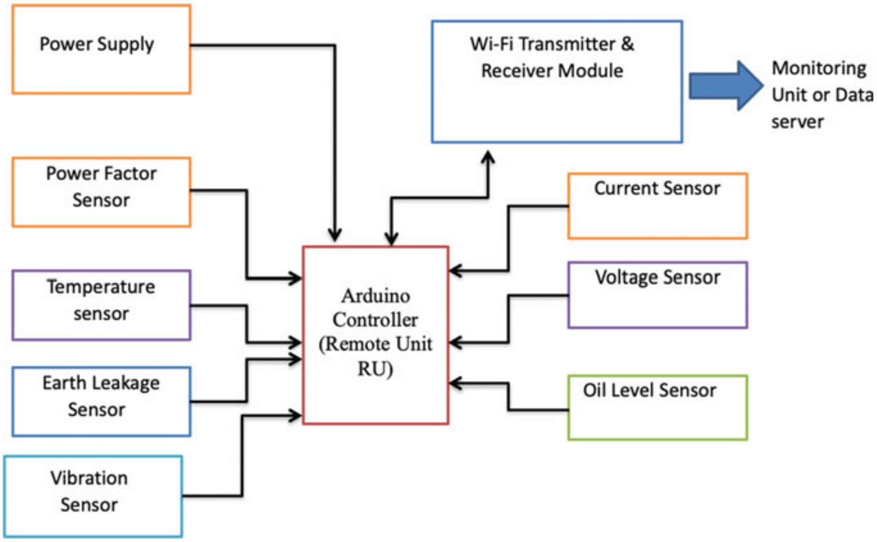


Fig. 1 Block diagram

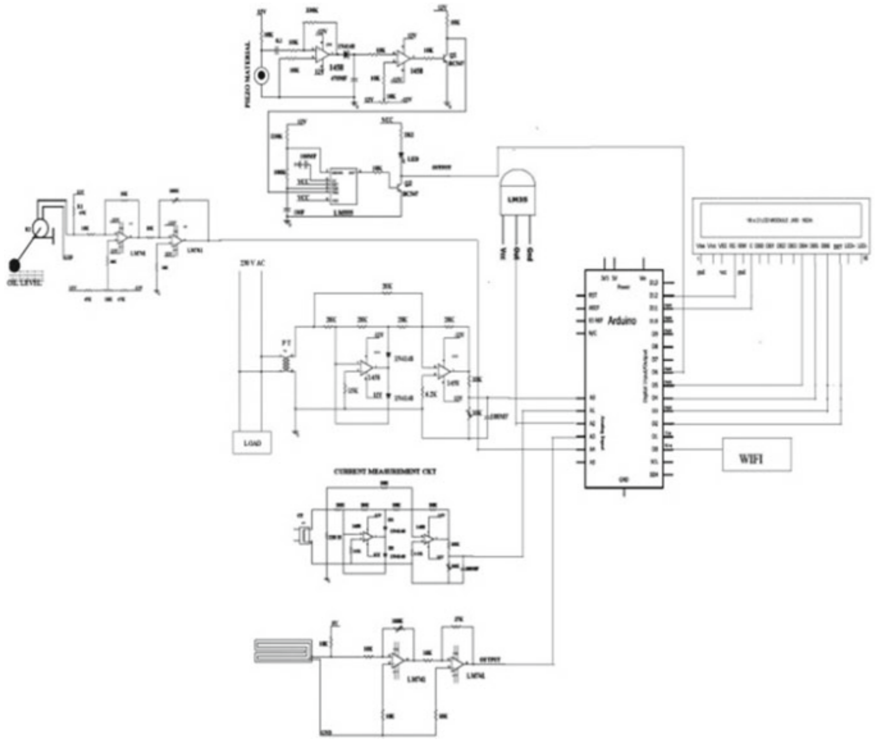


Fig. 2 Simulation output

2.2 Bridge Rectifier

The full wave bridge rectifier produces twice the voltage output amount then the conventional half wave rectifier. Due the presence of diode around 1.4 V of the input voltage has been dropped. IN4007 junction diode used here has 0.7 V drop since two diode are used in the full wave bridge rectifier total of 1.4 V is dropped.

2.3 Filter

Rectifier output is a variable DC voltage with ripples, and hence, it must be smoothen to a constant DC by using filter circuit. Filter circuit consists of capacitor in parallel with load resistor.

Charged to a peak value of the rectified waveform. When the voltage from rectifier exceeds the capacitor voltage, capacitor discharges the voltage to load resistor and the cycle repeats.

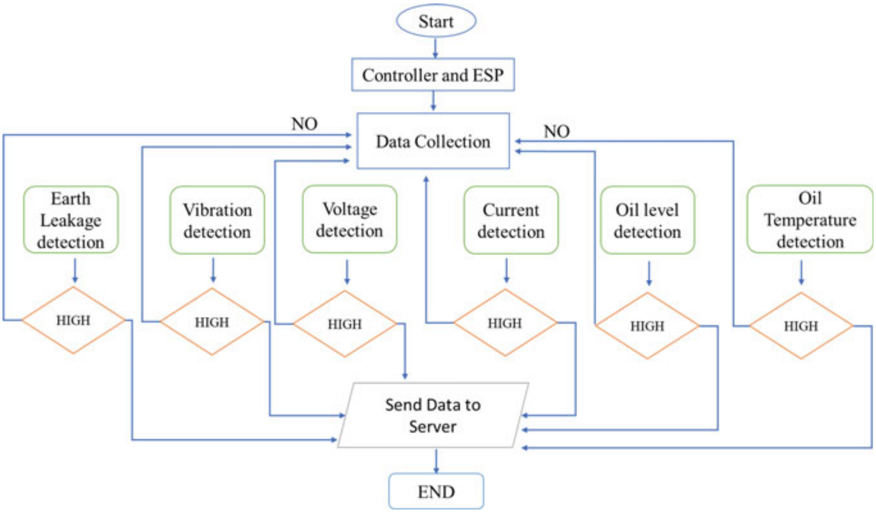


Fig. 3 Flowchart

2.4 IC Voltage Regulators

The voltage regulator consists of series of integrated circuits as shown in Fig. 3. A single IC consists of comparator, amplifier, overload protection coil and control circuits. It regulates fixed positive and negative voltage or adjustable set voltage. Here, we use IC7805 regulator which regulates variable DC voltage to a constant positive DC, 5 V.

2.5 Schematic Explanation

This circuit is designed in such a way that it could monitor the supply voltage. The potential transformer step downs the supply voltage that has to be monitored. Usually, precision rectifier is used here; the step down voltage is rectified. In order to have a circuit behaving like an ideal diode or rectifier, the precision rectifier is a configuration obtained with an op-amp. With the help of variable resistor VR1, the output of rectified voltage is adjusted to 0–5 V. The C1 capacitor filters the output rectified voltage that is given to ripples. After the filtration of the corresponding DC voltage by capacitor C1, the DC voltage is given to ADC or other related circuit.

2.6 Current Measurement

This circuit consists of current transformer and shunt resistor. The current transformer reduces the high supply current to a low value, which is then changed to low voltage by a shunt resistor linked in parallel. Precision rectifier is used to rectify the converted voltage to behave like an ideal diode. The precision rectifier is a configuration obtained with operational amplifier.

2.7 Power Factor Measurement Circuit

Power factor is a ratio of real or true power in watts to apparent power in KVA. It can also be defined as cosine of an angle between the current and voltage.

Power factor is measured using the circuit as shown in Fig. 4. The potential transformer monitors the line voltage and current transformer monitors current. The potential transformer work is to step down the voltage to 6 V from 440 V or 230 V AC. The 6 v signal is fed to zero crossing detector which generates sine wave to square wave signal.

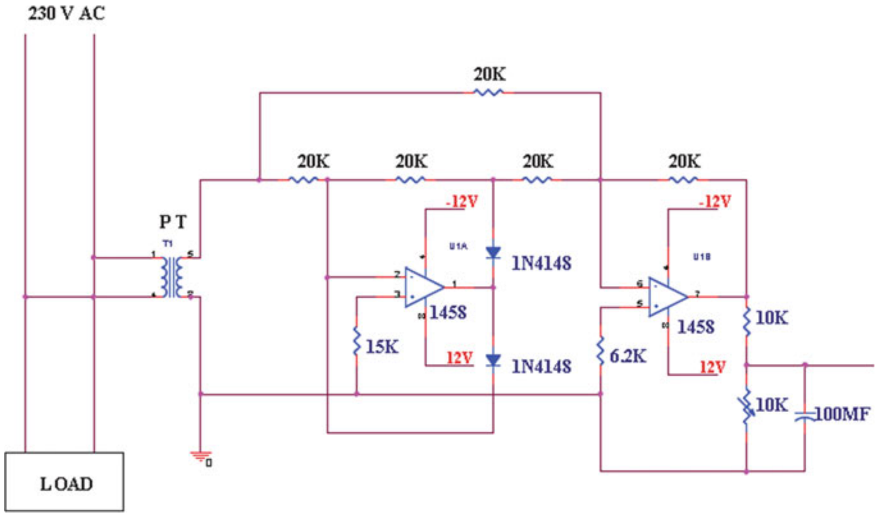


Fig. 4 Voltage regulator circuit

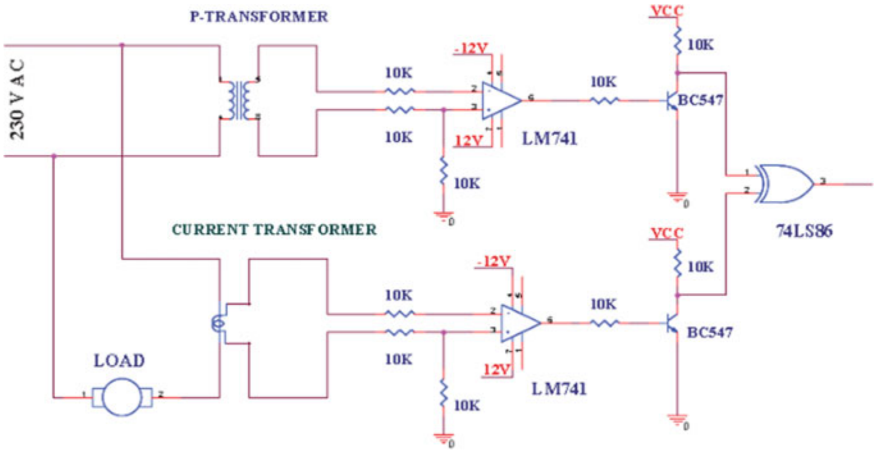


Fig. 5 Power factor measurement

2.8 Vibration Sensor

The vibration circuit senses the mechanical vibration. It has Piezo electric plate as shown in Fig. 5 is a special type of sensor that senses the vibration. It converts the mechanical vibration to electrical signal, and the signal is in the range of milli volt and is fed to comparator to produce pulses. This pulse is given to controller.

2.9 Level Measurement Using Float

Transducer is used to measure the level of oil in the sealed tank (float is one type of sensor) as shown in Fig. 6. Resistance values vary with respect to float according to the water range the principle of potential divider form is used for working of this method.

2.10 Oil Level Measurement Circuit

The float varies depending on the oil level. Float heterogeneous means that the resistance value is also heterogeneous. The output also increased as the resistance value increased. The resistance value and the output are directly proportional. The output analog voltage is fed to the ADC to convert to digital signal. The corresponding digital signal microcontroller then takes over the process. The ADC value increases as the level increases. Only oil level can be measured by controller or processor.

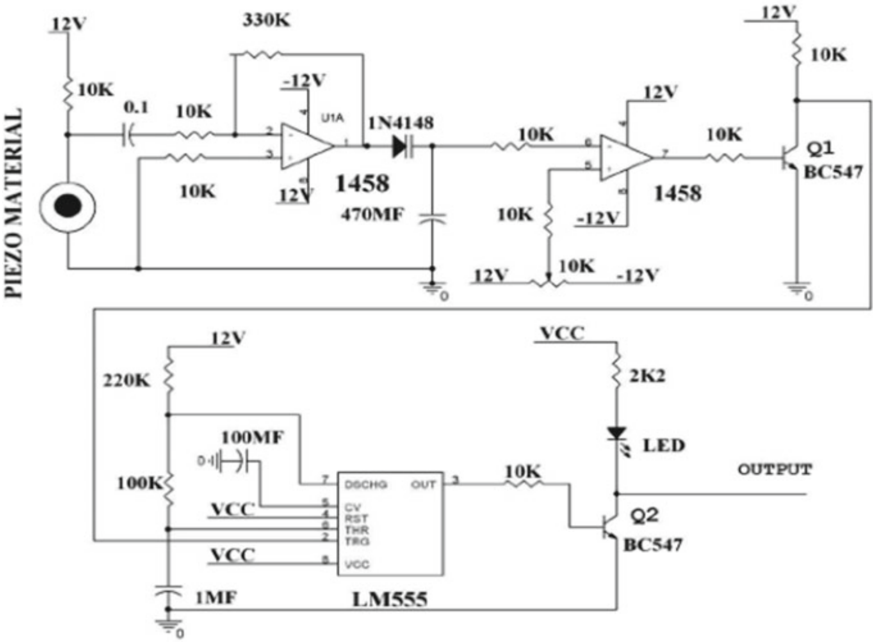


Fig. 6 Vibration sensor circuit

2.11 Earth Leakage Detection

In this circuit, the earth leakage sensor is used to find the earth leakage or liquid content. Earth leakage sensor is a sensitive variable resistor. Resistance value increases with respect to decrease in water particle content.

2.12 Thing speak

This platform (thing speak) supplies a variety of services completely aims to build IoT purpose. It has the capacity of real-time data gathering, visualize the collected data are displayed as charts, capability to build plugins and has applications which will collaborate web services, APIs and other Social Network. 'Thing Speak Channel' is a core element of Thing speak device. Channel stores the output data and has following elements as shown in Fig. 7, *8 fields for storing data of any type,*3 location fields and 1 status field.

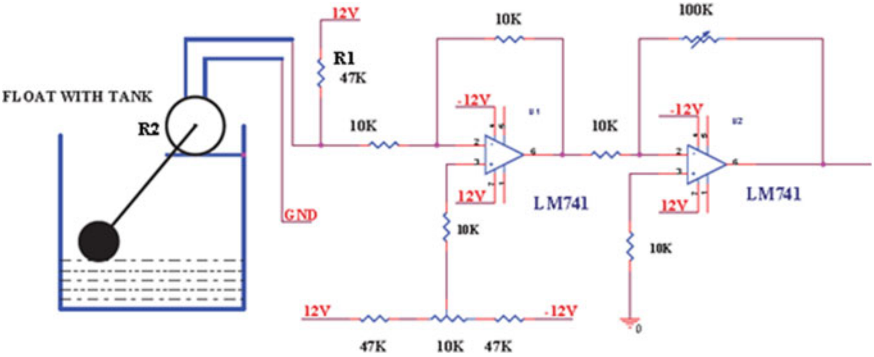
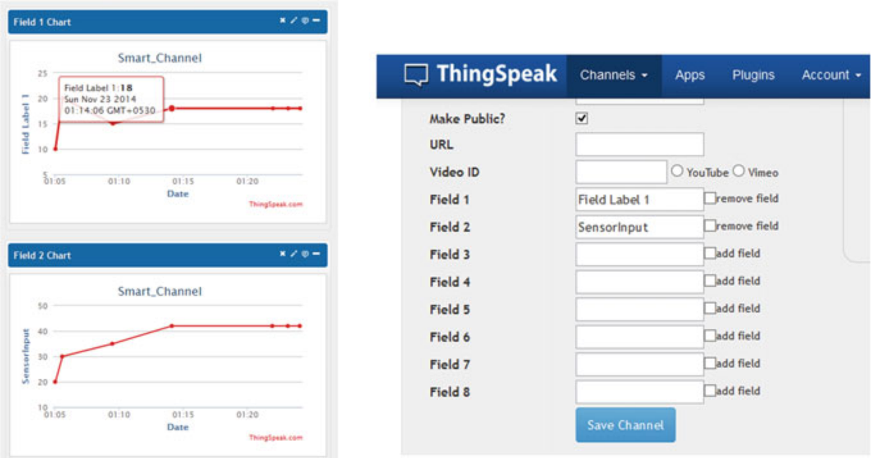


Fig. 7 Oil level measurement

2.13 The Internet of Things

IoT is a system which connects and talks to all devices as shown in structure of IoT in Fig. 8. The embedded operating system has a capability to communicate with the internet or with the neighboring devices. Internet of Things application is a basic block of IOT where all 'devices' talk to each other.

By using Thing Speak' one can analyse, monitor and counteract the parameters in IOT application platform. After continuous updates, the charts in the private view tab for each of the fields will be as shown in Fig. 8. Each highlighted points give the value, and the time at which the value was taken. We can get the details by placing the pointer on the highlighted points along the graph.



Make Public? ☑

URL

Video ID ○ YouTube ○ Vimeo

Field 1 Field Label 1 □ remove field

Field 2 SensorInput □ remove field

Field 3 □ add field

Field 4 □ add field

Field 5 □ add field

Field 6 □ add field

Field 7 □ add field

Field 8 □ add field

Save Channel

Fig. 8 Think speak-channel

2.14 Future Scope

As far now, measured data's are monitored remotely and in future can make the device system to communicate in both directions and can have advanced control on the measured quantities.

3 Conclusion

IoT based online monitoring of transformer is helpful comparing to monitoring it manually and more reliable as it is impossible to measure the oil level at all times, oil temperature rise, ambient temperature rise, and load current manually. As abnormalities occur, it will automatically send data to the system in EB department, and they can take necessary action. Figure 9 shows the snapshot of the hardware (Fig. 10).

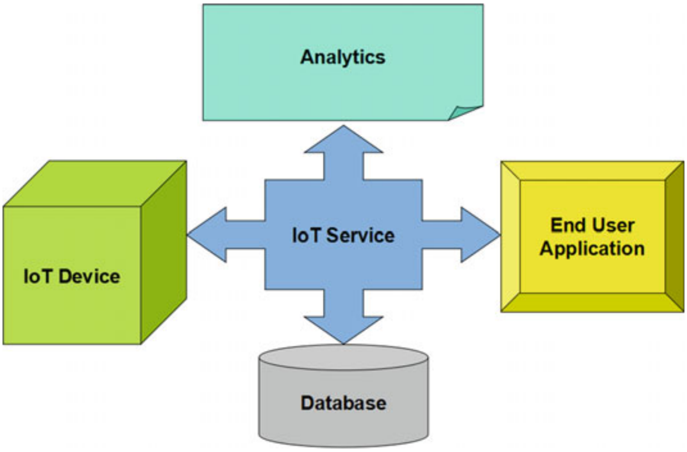


Fig. 9 IOT-structure

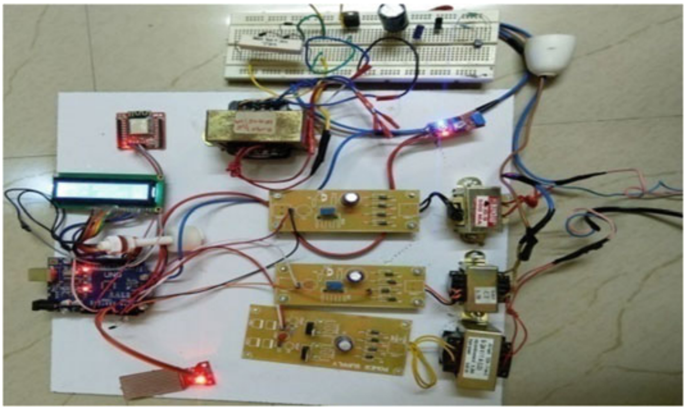


Fig. 10 Hardware snapshot

References

1. K. Vadirajacharya, A. Kharche, H. Kulakarni, V. Landage, Transformer health condition monitoring through GSM technology. Int. J. Sci. Eng. Res. **3**(12), 1 (2012)
2. A. Kumar, A. Raj, A. Kumar, S. Prasad, B. Kumar, Method for monitoring of distribution transformer. Dr. M.G.R, Educational and Research Institute, University, Chennai 600095
3. P. Karpagam, M. Chindamani, S. Anitha, A. Dhanalakshmi, Intelligent load shedding for distributed generating system. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **5**(10), 8213–8219 (2016)
4. A.Z. Loko, A.I. Bugaje, A.A. Bature, Automatic method of protecting transformer using microcontroller as an alternative to the fuse protection technique. Int. J. Tech. Res. Appl. **3**(2), 23–27. e-ISSN: 2320-8163
5. S.R. Karpe, S. Shelar, S. Garkad, S. Lakade, Fault detection and protection of transformer by using microcontroller. Int. J. Modern Trends Eng. Res.
6. R. Murugan, R. Ramasamy, Failure analysis of power transformer for effective maintenance planning lectric utilities. Eng. Fail. Anal. **55**, 182–192 (2015)
7. A.C. Nishant, Failure analysis of a power transformer using dissolved gas analysis—a case study. Int. J. Res. Eng. Technol. **3**(5), 300–303 (2014)
8. D.M. Mehta, P. Kundu, A. Chowdhury, V.K. Lakhiani, DGA diagnostics save transformers—case studies, in *2015 International Conference on Condition Assessment Techniques in Electrical Systems, CATCON 2015—Proceedings* (2016), pp. 116–120
9. M. Mostafaei, F. Haghjoo, Flux-based turn-to-turn fault protection for power transformer. IET Gener. Transmis. Distrib. **10**(5), 1154–1163 (2016)
10. C. Meyyappan, C.S. Ravichandran, 2021 Performance analysis of fault-tolerance techniques towards solar fed cascaded multilevel inverter, in *International Conference on Computer Communication and Informatics (ICCCI)* (2021), pp. 1–7. https://doi.org/10.1109/ICCCI50826.2021.9402705

A Hybrid Framework Using the Boosting Technique for Efficient Rain Streak Removal During Daylight

Subbarao Gogulamudi, V. Mahalakshmi, and Indraneel Sreeram

Abstract Nowadays, images and videos with a rainy footprint will degrade the quality of images. The removal of such rain streaks from the images is considered as a challenging task. The layer priors and recurrent convolutional neural networks (RCNNs) are combined and applied for developing the best quality images. Here, the layer priors focus on superimposing the rain streak layer with respect to the background image and RCNN is focused on motion blur and eliminating such blur from the images. The final output of a valid and best quality image is obtained by using the combined principle of these two approaches. The proposed model is further enhanced by using boosting technique, which has an aim to fasten the principle and minimize the complexity. A quality image is achieved by separating rain streaks from the desired rainy image and focusing more on the background image. The integration of these approaches (layer priors + RCNN using boosting) along with boosting technique results in delivering a faster output with enhanced accuracy and performance. In the future, utilizing any of the innovative image enhancement approaches, it may be possible to improve the image quality without rain streaks.

Keywords Layer priors · RCNN · Boosting · Hybrid framework · Accuracy · Performance · And quality image

1 Introduction

Recently, different methodologies and tools are developed to remove rain streaks from the images and video sequences. When recording a video during a rainstorm, the sequence of images may include rain drops, reducing the quality and visibility of the captured images and videos. In order to increase the visibility of images, the rain streaks should be removed effectively. Some specific research studies have been

S. Gogulamudi (✉) · V. Mahalakshmi
Department of CSE, Annamalai University, Cuddalore, Tamilnadu, India

I. Sreeram
Department of CSE, St. Ann's College of Engineering and Technology, Chirala, Andhra Pradesh, India

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_73

developed recently to remove the streaks and produce clear images. The following are the few relevant approaches (Table 1).

The steps to be performed in order to achieve the theme of the hybrid approach using boosting technique are as follows:

(1) Read the data from data set.
(2) Apply layer priors.
(3) Apply RCNN.

Table 1 Existing approaches and their details

Approach	Purpose	Advantages	Disadvantages
Dictionary learning	Based on predefined atoms, it could process the task	Fast	May result in corruption of pixels
TAWL	Works based on appearance, wideness, and location	Flexible on different resolutions and frame rates	Still possible to enhance the quality
Self-learning	Decomposes the image, applies the logic over each part, and enhances the background image	Applies to videos, and efficiency is more	Needs to work on dynamic scenes
Layer priors	Single image will be layer-wise decomposed, and then the rain streak gets superimposed based on Gaussian model	Applies on multiple orientations and scales	Needs to work on many images in the sense videos
Deep learning	It works on steps like absolute residual and detailed layers	Applies on videos as well as images	Complexity is more
RCNN	It focuses on motion blur caused by rain streaks and works based on angle and length parameters	Applies on videos as well as images	Complexity is more
K means clustering	It works based on segmentation, i.e., number of approximating the clusters, and evaluates the accuracy	Applies on videos as well as images	NA
Hierarchical approach	It works on low-frequency part image and next-level layers in the process till quality image is obtained	Applies on videos as well as images	NA

(4) The output of layer priors with boosting and output of RCNN with boosting will be performed to achieve the appropriate accuracy and performance.
(5) Show the graphs of hybrid approach against individual existing approaches.
(6) Showcase the quality through the results.

2 Literature Review

In this section, many studies that resemble the theme of the rain streak removal are explored. Recently, the kernel-guided convolutional neural network is applied with efficient architectures in order to preserve the contrast, and by considering the source mentioned in [1], the texture while removing streaks from the rain will be observed. The two parameters considered here are length and angle. The disadvantage of getting over- and under-rain results has been obtained by using earlier deep neural networks.

As per the source specified in [2], the DerainNet approach performs advantageously when compared with the other similar approaches in terms of computational efficiency and image quality. The expected results are obtained without requiring any deep- or wide-sized neural network. With respect to the work given in [3], the BLSRN is proposed that helps in few activity separations in other layers such as fence, reflection, and shadowing. The efficiency and extraordinary performance of this BRN are explored and compared with other de-rained models. As the focused study explored in [4], the image is decomposed and forms the sub-net which would extract reflection and illumination images from the given image. The multi-stream network autoencoder is modeled which recurrently enhances the nightmare rainy images. Here, the sub-nets are helpful in removing the streaks from illumination-enhanced image. With regard to [5], it deals with J4R(Joint Recurrent Rain removal and reconstruction net integrates degradation classification, spatial texture, and temporal coherence that reconstructs background details. This model measures both the occlusions and rain streaks. As the source specified in [6], the semi-supervised approach is proposed that uses supervised rain images with or without synthesized rain. This helps to track of residuals between the input and output networks.

As per the study focused in [7], the proposed approach called conditional generative adversarial networks uses adversarial loss as additional component to the loss function. This helps to regulate the system and produces better results compared to other existing methods. With regard to demonstration of work in [8], the TAWL based on appearance, wide, and location uses the features on different resolutions and frame rates, and removes the rain from the video sequences. The two categories such as synthetic and real type are applied in order to demonstrate the results against other approaches. With respect to the information provided in [9], the SLDNet is applied where the first-phase temporal correlations are focused on clean video and extracts main structures where second-phase temporal consistency is focused on adjacent frames and extracts structural details. These two are jointly applied as motion estimation and rain region detection. In the aspect of source specified in [10], the L0 gradient considers nonzero gradients of an image and uses salient edges. This decreases the details of low gradients as well as consequential quality from the given image. The histogram adjustment technique is applied to remove the streaks and produce better results. As per the demonstration of [11], the existing methods such as dictionary learning and low rank structures when applied leaves the too many streaks or over-smooth could be done in the background images. The proposed patch-based prior method would use multiple orientations as well as the scales. Removing the rain streaks by using the defined method is better than other existing methods. With regard to data provided in [12], the image is decomposed into low- and high-frequency parts by using bilateral filter, where high-frequency part is decomposed into rain and non-rain components. Hence, it is possible to remove rain component and preserve the actual details. With respect to the demonstration of work in [13], various operations over the image like in-painting and de-noising on the spare representation of those images are based on the predefined dictionary. The performance of K-SVD and sparse approaches is compared, and similar performance will be identified w.r.t. the state-of-the-art approaches. As per the demonstration of [14], the image gets decomposed into LF and HF parts, where HF part is further analyzed to obtain rain-free image, where multiple guided filters are applied in order to remove the rain streaks and minimum operation is applied in order to enhance the image quality. With regard to the information provided in [15], the three priors are considered after decomposing the image into background B and rain streak R, and the final result obtained would be a quality image. With regard to source specified in [16], the process remains dependent on four terms such as sparsity norm, fidelity, and two directional smoothness regularization terms, which leads to develop a novel approach called alternating directional method of multipliers (ADMM). In the connection of source given in [17], the rain map is demonstrated w.r.t. adjacent frames like previous and next frames based on HSV color space saturation and filtering of low pass in order to fasten the output when compared to traditional approaches. With regard to the demonstration of work in [18], the rain streak should be removed from the rain images based on several rain streak removal approaches. Their accuracies and performances are compared for producing the clear quality images. As per the discussion made in [19], the significance of IoT along with deep learning is demonstrated for the optimization of output by using gradient descent feature, global search technique, and stochastic analysis to perform iterations by using swarm intelligence. In the view of studies from the sources [20–31], the deep learning network with features decomposition and composition, which aims to extract the clean background and rain layered images.

From the aspect of [30], the detection of premature bosom irregularity in the images related to personal healthcare systems, is discussed and the role modality is explored in processing of the system. Regarding the demonstration of information provided in [31], the green environment is guaranteed in smart systems and consumer electronics domain using ANFIS. The efficient management and quick data transfer are guaranteed over nonlinear outputs obtained from case study environment and make progressing toward sustainable environment. Many factors are considered in making greener system. From the view of [32], the iterative method in the first step for the direct inverse, where the computational cost is expensive as well as difficult in choosing the hyper parameters. The novel integration of residual and multi-resolution decomposition learning performs well when compared to other approaches and reconstructs image with high resolution.

The aforementioned studies are more focused on many existing state-of-the-art approaches to remove rain streaks from the rainy images or videos, and it also requires smart devices for detecting, processing, and communicating the information in order to produce the result as expected by increasing the automation and minimizing the human efforts.

3 Proposed Approach

The proposed theme is divided into different modules such as the application of layer priors to decompose the HF part into the frames and extract the clear quality image by superimposing the rain streak image against the background image, wherein the application of RCNN aims to eliminate the motion blur that exists in the input image in order to obtain the visible quality image and fasten the process in order to obtain the output in a quicker time by using the boosting approach.

The three modules defined in this proposed theme are as follows:

(A) **Application of Layer Priors**: The input image is decomposed into background and rain streak images and expressed in the form $O = B + R$. The equations for solving the H, $\{B, R\}$, and {GB and GR} are specified in pseudo-code of layer priors algorithm. The resulted RGB model is converted into YUV model and removes the streak pixels from Y(YUV) space.

(B) **Application of Boosting**: It makes weak learners into strong learner. It includes various techniques such as AdaBoost, gradient boosting, and XGBoost. It repeats certain steps such as next image by overcoming and refining the previous image mean square errors. It includes the benefits such as flexibility to overfitting, easy interpretation, and strong prediction of true scene image.

(C) **Application of RCNN**: It ensures high accuracy and best visual coherence. It does not depend on segmentation and corrects its own errors. It models the input image based on complex spatial dependencies with low inference cost (Fig. 1).

The pseudo_Procedure Layers_Priors(image_dataset[], guided_filter):

Input: Input image, and G_B and G_R Gaussian variables

Output: $B \leftarrow O$; $R \leftarrow 0$; $\omega \leftarrow \omega_\circ$ where B is background image after removing the rain streaks

Do the following steps till final B and R layer images are formed with no noise

Step 1: Based on LOSSO problem, the shrinkage operator simplifies the vectors and matrices over the elements

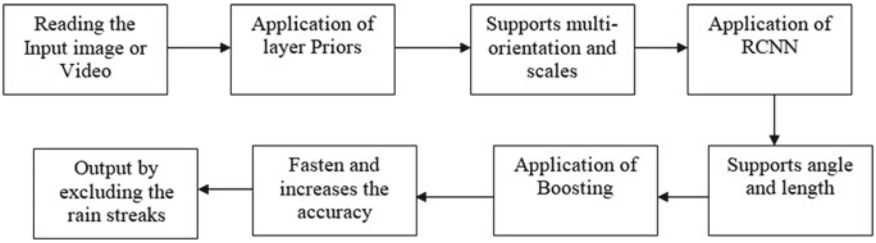

Reading the Input image or Video
Application of layer Priors
Supports multi-orientation and scales
Application of RCNN
Supports angle and length
Application of Boosting
Fasten and increases the accuracy
Output by excluding the rain streaks

Fig. 1 Theme of hybrid approach block diagram

Wise of $\mathbf{H}^{(t+1)} = \arg\min_{\mathbf{H}} \alpha \|\mathbf{H}\|_1 + \omega \left\| \nabla \mathbf{B}^{(t)} - \mathbf{H} \right\|_F^2$ into a resulted form as $\mathbf{H}^{(t+1)} = S_{\alpha/2\omega}\left[\nabla \mathbf{B}^{(t)}\right]$. The H could be updated for finite iterations.

Step 2: Solve {B, R} based on:

$$\begin{aligned}\left\{\mathbf{B}^{(t+1)}, \mathbf{R}^{(t+1)}\right\} = \underset{\mathbf{B},\mathbf{R}}{\operatorname{argmin}} \|\mathbf{O} - \mathbf{B} - \mathbf{R}\|_F^2 + \beta \|\mathbf{R}\|_F^2 \\ + \omega \sum_i \left(\left\| \mathcal{P}(\mathbf{B}_i) - \mathbf{g}_{\mathbf{B}_i}^{(t)} \right\|_2^2 + \left\| \mathcal{P}(\mathbf{R}_i) - \mathbf{g}_{\mathbf{R}_i}^{(t)} \right\|_2^2 \right) \\ \text{s.t.} \quad \forall i \quad 0 \le \mathbf{B}_i, \mathbf{R}_i \le \mathbf{O}_i\end{aligned}$$

This could minimize L2 problem.

Step 3: Solve the G_B and G_R Gaussian variables

$$\mathbf{g}_{\mathbf{B}_i}^{(t+1)} = \underset{\mathbf{g}_{\mathbf{B}_i}}{\operatorname{argmin}} \, \omega \left\| \mathcal{P}\left(\mathbf{B}_i^{t+1}\right) - \mathbf{g}_{\mathbf{B}_i} \right\|_2^2 - \gamma \log \mathcal{G}_B(\mathbf{g}_{\mathbf{B}_i}).$$

Step 4: $W = 2 * W$.

Repeat Step 1 to Step 4 till maximum convergence is obtained w.r.t. saturation (Fig. 2).

The Pseudo_Procedure Boosting (Output_Layer_Priors_imgset[], Output_RCNN_imgset[]).

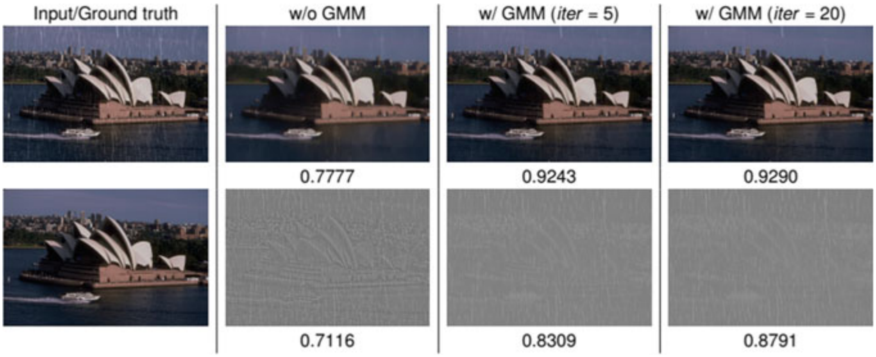


Fig. 2 Effect of GMM over the input image; detain images are shown at bottom of each image

Pseudo_Procedure XGBoost(Dataset[][])

Aim: It is coined by Taingi Chen. It is based on gradient boosting and other variations and trains the model by efficiently making use of available resources.
Input: image Dataset
Output: True scene image is to be guaranteed

- The target variable y to be predicted from initial model F0.
- A new model h1 is used to fit residuals from the previous step.
- New model F1 is formed by combining F0 and h1, which is boosted version of F0. The mean square error from F1 is lower than F0 is defined as $F_1(x) <- F_0(x) + h_1(x)$
- New model F2 to be formed to improve the performance of F1 and after residuals of F1 is defined as $F_2(x) <- F_1(x) + h_2(x)$.
- This process is repeated for n iterations until residuals are minimized and generalized as $F_m(x) <- F_{m-1}(x) + h_m(x)$.
- Additional learners are used to bring down the errors without affecting the functions formed in previous steps.
- For an example, F0(x) for given dataset is defined to minimize loss function or MSE in this case as $F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$ $argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma) = argmin_\gamma \sum_{i=1}^{n} (y_i - \gamma)^2$.
- The boosting function is defined by taking differential w.r.to γ as $F_0(x) = \frac{\sum_{i=1}^{n} y_i}{n}$.
- In this process, the gradient of loss function is defined iteratively as $r_{im} = -\alpha \left[\frac{\partial(L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$, where α is the learning rate .

The boosted model Fm(x) for each hm(x) on each step using multiplicative factor γm is defined as $F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$.

The Pseudo_Procedure RCNN (output_Layer_Priors_imgset[]):

```
Step 1: Convolution operation to be performed
Step 2: Performs the ReLU operation over Step 1
Step 3: Performs pooling
Step 4: Flattening to be performed
```

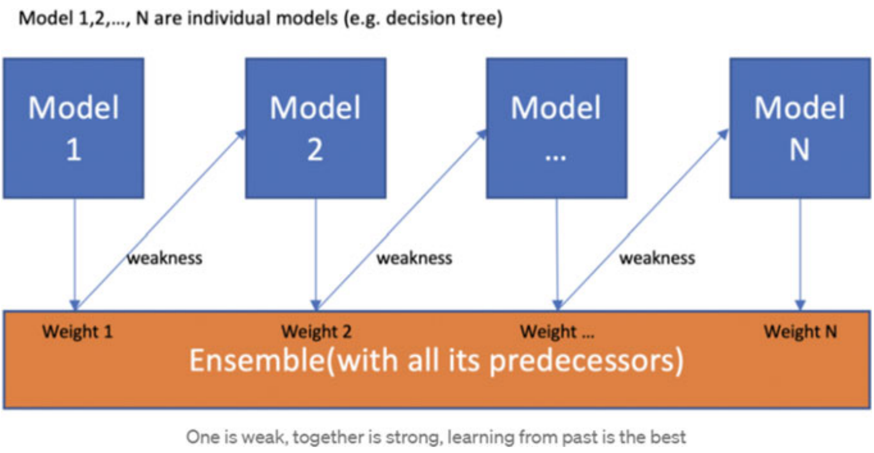

Model 1,2,..., N are individual models (e.g. decision tree)
Model 1
Model 2
Model ...
Model N
weakness
weakness
weakness
Weight 1
Weight 2
Weight ...
Weight N
Ensemble(with all its predecessors)
One is weak, together is strong, learning from past is the best

Fig. 3 Boosting method theme block diagram

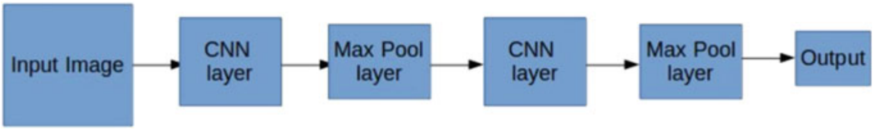


Fig. 4 Simple theme of recurrent convolutional neural network

Step 5: Full connection to be provided.

From Fig. 4, the input image is transformed into convolutional network, and pooling is applied till true scene to be obtained by minimizing the error rate from previous to next images during the extraction (Fig. 3).

From Fig. 5, the true scene image is extracted by applying the convolutional neural network steps one after the other in order to get the quality image by removing all blurs represented by rain streaks here in this case.

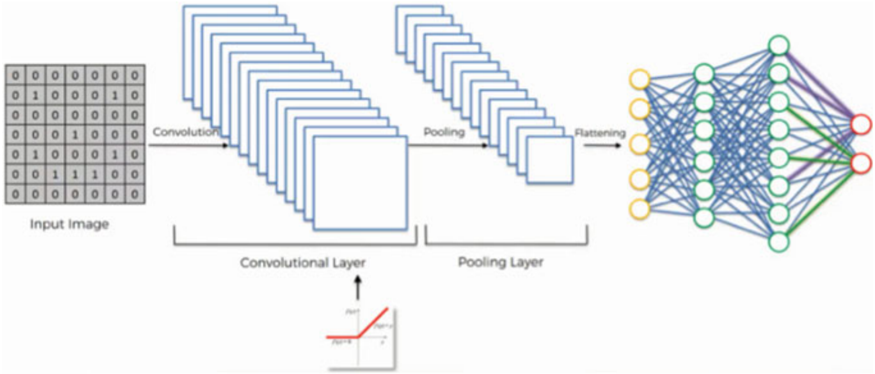
The flowchart of hybrid approach for rain streak removal is defined as the below form (Fig. 6).

4 Results

The first module when applied over the input image results in the following output and statistical table when taken against few existing approaches (Fig. 7):

The following table shows the data information based on SR and LRA against layer priors based on Gaussian approach (Table 2).

After applying the boosting technique, the values are hiked as follows (Tables 3, 4, and 5).



0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

Fig. 5 Detailed theme of recurrent convolutional neural network

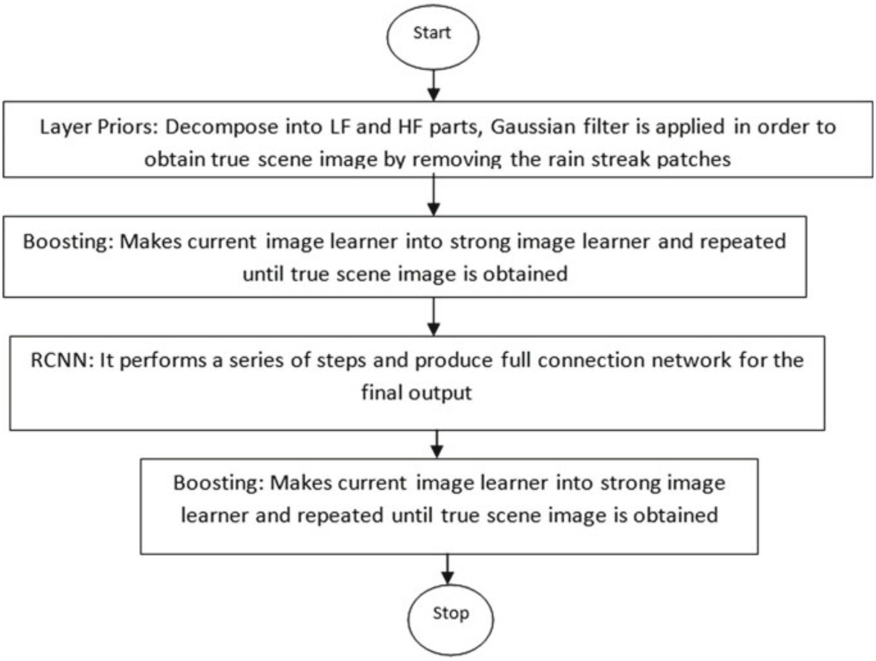

Start
Layer Priors: Decompose into LF and HF parts, Gaussian filter is applied in order to obtain true scene image by removing the rain streak patches
Boosting: Makes current image learner into strong image learner and repeated until true scene image is obtained
RCNN: It performs a series of steps and produce full connection network for the final output
Boosting: Makes current image learner into strong image learner and repeated until true scene image is obtained
Stop

Fig. 6 Flowchart of rain streak removal hybrid approach

When the RCNN is applied over the following input image, the expected sequence of activities performed is shown in the below scene (Fig. 8).

After applying the boosting technique, the values are hiked as follows.

The accuracy is defined in the following graph as the average of layer priors accuracy and RCNN accuracy with boosting feature over them (Fig. 9).

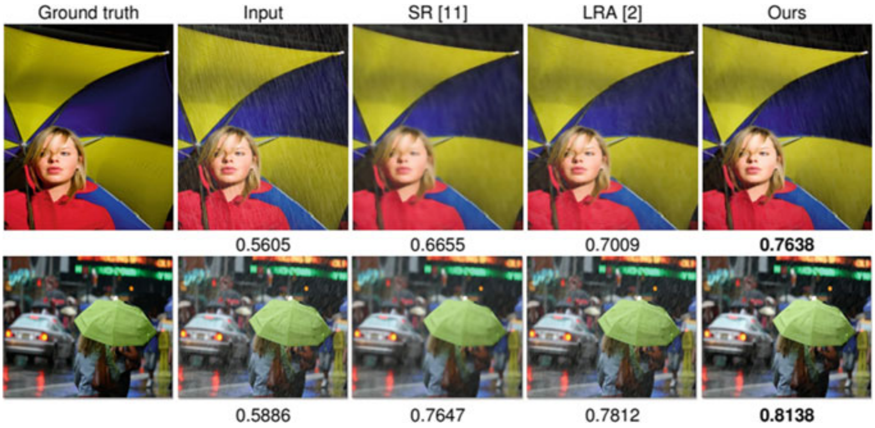


Fig. 7 Quality of background image extraction based on SR and LRA

Table 2 Data set comparison of layer priors based on Gaussian against other approaches

SR [11]	0.73	0.79	0.83	0.76	0.62	0.73	0.82	0.77	0.73	0.74	0.63	0.76
LRA [2]	0.83	0.87	0.79	0.85	0.88	0.90	0.92	0.82	0.87	0.83	0.85	0.81
Ours	**0.88**	**0.93**	**0.93**	**0.93**	**0.90**	**0.95**	**0.96**	**0.90**	**0.91**	**0.90**	**0.86**	**0.92**

The performance of the hybrid approach with boosting technique against individual approaches performed separately is also shown (Fig. 10).

5 Conclusion

Layer priors, where the HP part has further analyzed the Gaussian modeling in order to extract true scene image by eliminating the rain streaks, and RCNN, where a series of steps are applied in order to extract true scene image by removing the rain streaks by refining existing image to the new image, are the two approaches considered in this study. Their accuracies are noted and analyzed by using boosting technique. The results proved that 99% of accuracy be guaranteed in obtaining the true scene without rain streaks. Hence, the integrated selection of approaches such as layer priors and RCNN has produced fruitful output as per expectations. In addition, the type of boosting technique applied here is XGBoosting, which has fastened the performance of hybrid approach.

Table 3 Data set comparison of layer priors based on Gaussian against the identical method with boosting

Layer priors with Gaussian modeling	0.88	0.93	0.93	0.93	0.9	0.95	0.96	0.9	0.91	0.86	0.92
Layer priors with Gaussian modeling using boosting	0.95	0.99	0.99	0.99	0.96	1	1	0.96	0.97	0.94	0.98

Table 4 Data set comparison of RCNN against the identical method with boosting

	Pic1	Pic2	Pic3	Pic4	Pic5	Pic6	Pic7	Pic8	Pic9	Pic10	Pic11	Pic12	Aug
RCNN	0.91	0.95	0.95	0.98	0.95	0.98	0.98	0.98	0.96	0.93	0.94	0.95	0.986
RCNN using boosting	0.97	1	1	1	1	1	1	1	1	0.99	1	1	1

Table 5 Accuracy obtained by the hybrid approach using boosting

Layer priors with Gaussian modeling	92
Layer priors with Gaussian modeling using boosting	98
RCNN	96
RCNN using boosting	100
Average of layer priors and RCNN with boosting	99

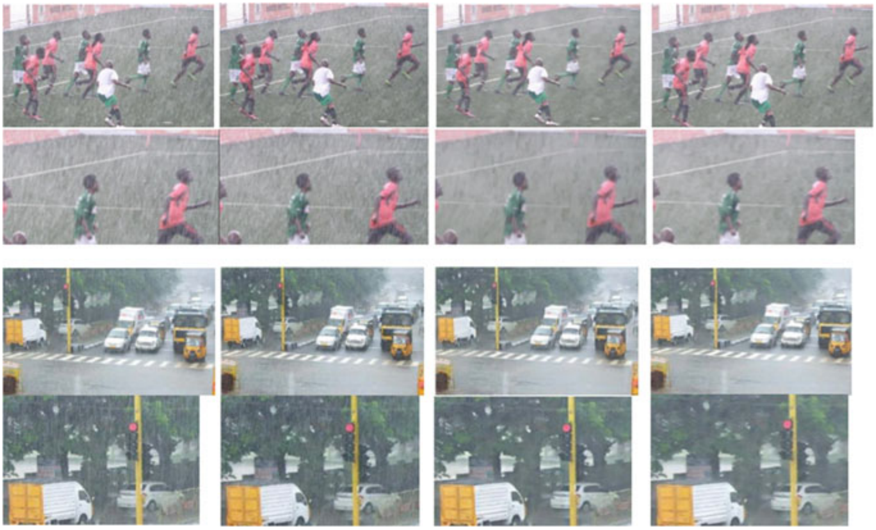


Fig. 8 Effect of RCNN against other existing approaches

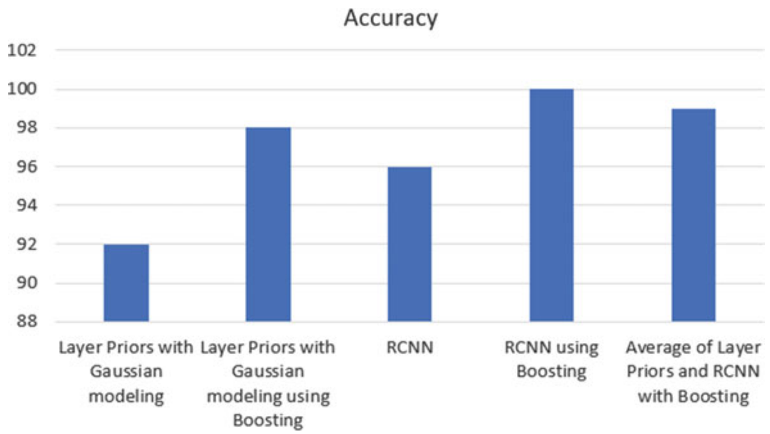


Fig. 9 Accuracy of hybrid approach against identical approaches without boosting

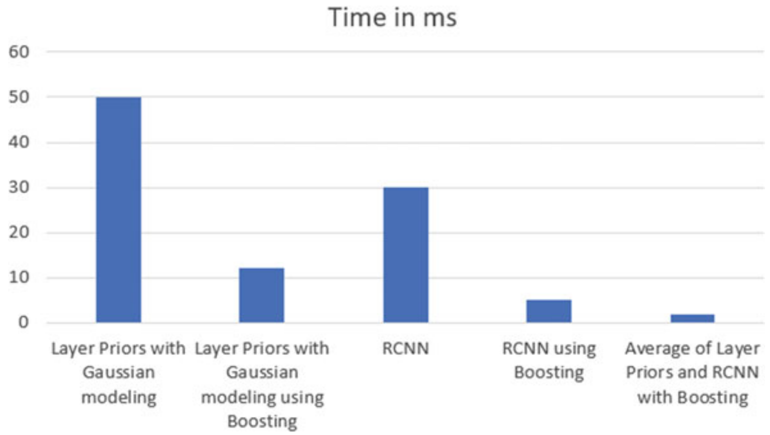


Fig. 10 Performance of hybrid approach against identical approaches without boosting

References

1. Y.-T. Wang, X.-L. Zhao, T.-X. Jiang, L.-J. Deng, Y. Chang, T.-Z. Huang, Rain streak removal for single image via Kernel guided CNN, August 2018. https://doi.org/10.1109/TNNLS.2020.3015897
2. X. Fu, J. Huang, X. Ding, Y. Liao, J. Paisley, Clearing the skies: a deep network architecture for single-image rain removal, February 2017. https://arxiv.org/pdf/1609.02087.pdf
3. D. Ren, W. Shang, P. Zhu, Q. Hu, D. Meng, W. Zuo, Single image Deraining using bilateral recurrent network. IEEE Trans. Image Process. https://csdwren.github.io/papers/2020_tip_BRN.pdf
4. Z. Shi, Y. Feng, M. Zhao, L. He, A joint deep neural networks-based method for single nighttime rainy image enhancement. Neural Comput. Appl. 1913–1926 (2020). https://doi.org/10.1007/s00521-019-04501-5

5. J. Liu, W. Yang, S. Yang, Z. Guo, Erase or fill? Deep joint recurrent rain removal and reconstruction in videos, https://www.icst.pku.edu.cn/struct/Projects/J4RNet_files/CVPR-Rain-Video-main-v4.0.pdf
6. W. Wei, D. Meng, Q. Zhao, Z. Xu, Y. Wu, Semi-supervised transfer learning for image rain removal. IEEEExplore, https://openaccess.thecvf.com/content_CVPR_2019/papers/Wei_Semi-Supervised_Transfer_Learning_for_Image_Rain_Removal_CVPR_2019_paper.pdf
7. P. Hettiarachchi, R. Nawaratne, D. Alahakoon, D. De Silva, N. Chilamkurti, Rain streak removal for single images using conditional generative adversarial networks. Appl. Sci. **11**(5), 2214 (2021). https://doi.org/10.3390/app11052214
8. M.R. Islam, M. Paul, Rain streak removal in a video to improve visibility by TAWL algorithm. Electr. Eng. Syst. Sci. Image Video Process. (2020). arXiv:2007.05167v1
9. W. Yang, R.T. Tan, S. Wang, J. Liu, *Self-Learning Video Rain Streak Removal: When Cyclic Consistency Meets Temporal Correspondence.* (IEEE, 2020). https://openaccess.thecvf.com/content_CVPR_2020/papers/Yang_SelfLearning_Video_Rain_Streak_Removal_When_Cyclic_Consistency_Meets_Temporal_CVPR_2020_paper.pdf
10. G. Ananthi, T. Jenitha, S. Amutha, Rain streak removal using L0 gradient minimization technique. Eur. J. Molecular Clin. Med. **7**(2) (2020). ISSN 2515-8260. https://ejmcm.com/article_3139_937c84b5490a184bf6d9598c96fee0ab.pdf
11. Y. Li, R.T. Tan, X. Guo, J. Lu, M.S. Brown, *Rain Streak Removal Using Layer Priors.* (CVPR, 2020). https://doi.org/10.1109/CVPR33180.2016
12. L.-W. Kang, C.-W. Lin, Y.-H. Fu, Automatic single-image-based rain streaks removal via image decomposition. IEEE Trans. Image Process. (2011). https://doi.org/10.1109/TIP.2011.2179057
13. A. Kumar, S. Kataria, Dictionary learning based applications in image processing using convex optimisation, http://home.iitk.ac.in/~saurabhk/EE609A_12011_12807637_pdf
14. A. Gautam, Dr. K. Raj, Rain removal in digital images using guided filter. Int. J. Inform. Technol. Electr. Eng. **7**(5) (2018). http://www.iteejournal.org/Download_oct18_pdf_2.pdf
15. L. Zhu, C.-W. Fu, D. Lischinski, P.-A. Heng, Joint bi-layer optimization for single-image rain streak removal, in *2017 IEEE International Conference on Computer Vision.* https://doi.org/10.1109/ICCV.2017.276
16. Y. Wang, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, T.-X. Jiang, Rain streaks removal for single image via directional total variation regularization, in *ICIP 2019*. 978-1-5386-6249-6/19/2019 IEEE
17. K. Park, M. Kim, H. Lim, S. Yu, J. Paik, Fast rain removal using rain map and temporal filtering for consumer imaging applications. 978-1-5386-6095-9/18/2018 IEEE
18. K. Priyanka, Nagave, S.R. Mahadik, Automatic rain streak removal: review. Int. J. Novel Res. Dev. **2**(5) (2017)
19. A. Mahesh, Dr. P. Manimegalai, Optimizing artificial intelligence system using stochastic diffusion search in deep learning network. J. Adv. Res. Dyn. Control Syst. **10**(6) (2018)
20. S. Li, W. Ren, J. Zhang, J. Yu, X. Guo, Fast single image rain removal via a deep decomposition-composition network, April 2018. https://arxiv.org/pdf/1804.02688.pdf
21. Dr. P. Tumuluru, Dr. S. Hrushikesava Raju, CH.M.H. Sai Baba, S. Dorababu, B. Venkateswarlu, ECO friendly mask guide for corona prevention. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022047
22. CH.M.H. Sai baba, Dr. S. Hrushikesava Raju, M.V.B.T. Santhi, S. Dorababu, Er. Saiyed Faiayaz waris, International currency translator using IoT for shopping. IOP Conf. Ser. Mater. Sci. Eng. **981**, 4. https://doi.org/10.1088/1757-899X/981/4/042014
23. N. Sunanda, S. Hrushikesava Raju, S.F. Waris, A. Koulagaji, Smart instant charging of power banks. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022066
24. R. Mothukuri, Dr. S. Hrushikesava Raju, S. Dorababu, S.F. Waris, Smart catcher of weighted objects. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022002
25. M. Kavitha, Dr. S. Hrushikesava Raju, S.F. Waris, Dr. A. Koulagaji, Smart gas monitoring system for home and industries. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022003

26. Dr. S. Hrushikesava Raju, Dr. L.R. Burra, S.F. Waris, S. Kavitha, IoT as a health guide tool. IOP Conf. Ser. Mater. Sci. Eng. **981**, 4. https://doi.org/10.1088/1757-899X/981/4/042015
27. Dr. S. Hrushikesava Raju, Dr. L.R. Burra, Dr. A. Koujalagi, S.F. Waris, Tourism enhancer app: user-friendliness of a map with relevant features. IOP Conf. Ser. Mater. Sci. Eng. **981**, 2. https://doi.org/10.1088/1757-899X/981/2/022067
28. M. Kavitha, S. Srinivasulu, K. Savitri, P.S. Afroze, P. Akhil, V. Sai, S. Asrith, Garbage bin monitoring and management system using GSM. Int. J. Innov. Technol. Explor. Eng. **8**(7), 2632–2636 (2019)
29. M. Kavitha, K. Anvesh, P. Arun Kumar, P. Sravani, IoT based home intrusion detection system. Int. J. Rec. Technol. Eng. **7**(6), 694–698 (2019)
30. M. Kavitha, P.V. Krishna, V. Saritha, *Role of Imaging Modality in Premature Detection of Bosom Irregularity in Internet of Things and Personalized Healthcare Systems* (Springer, Singapore, 2019), pp. 81–92
31. H. Wang, Sustainable development and management in consumer electronics using soft computation. J. Soft Comput. Paradigm (JSCP) **1**(1), 56 (2019). https://doi.org/10.36548/jscp.2019.1.006
32. T. Vijayakumar, Posed inverse problem rectification using novel deep convolutional neural network. J. Innov. Image Process. (JIIP) **2**(3), 121–127 (2020). https://doi.org/10.36548/jiip.2020.3.001

Detection of DoS and DDoS Attacks Using Hidden Markov Model

M. Balaji Bharatwaj, M. Aditya Reddy, Thangavel Senthil Kumar, and Sulakshan Vajipayajula

Abstract This is the era of cloud computing. Major companies use the internet to access resources like GPU and NAS. A reliable and good network is very essential for any company. DNS attacks are more prominent in these segment of networks. DoS and DDoS minimize the potential of the resources that are immediately available to the company. They use majority of the resources and important tasks carried out in the company will get delayed because of hogging these resources. In this paper, we have proposed a scheme to detect DoS and DDoS using HMM algorithm. We have taken KDD-CUP dataset for our research. We have started the paper by giving an introduction about why networking is important and how DoS and DDoS affect the companies. Next, we have enumerated the sub classes of DNS attacks followed by explaining Markov Models and HMM. Then, we have proposed our scheme to tackle these DoS and DDoS attacks using HMM, followed by presenting our results and performance metrics.

Keywords Uniform resource locator · Denial of service · Distributed denial of service · Internet protocol · Graphic processing unit · Network attached storage · Markov chains · Clusters · Internet control message protocol · Hidden markov model

M. Balaji Bharatwaj · M. Aditya Reddy · T. Senthil Kumar (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: t_senthilkumar@cb.amrita.edu

M. Balaji Bharatwaj
e-mail: cb.en.u4cse16607@cb.students.amrita.edu

M. Aditya Reddy
e-mail: cb.en.u4cse17431@cb.students.amrita.edu

S. Vajipayajula
Chief Architect - Security Analysis, STFM, IBM security, Bangalore, India
e-mail: svajipay@in.ibm.com

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_74

1 Introduction

In today's world, every organization depends on networking. Internet is not only used for conducting video calls, sending E-Mails, and it is also used to access some of the very expensive resources that a company has to offer. The reason it is accessed via internet connection is because a commoner cannot afford to buy expensive GPU's or NAS to carry out majority of the tasks. These expensive hardware is the key to a company's success for faster product delivery and in turn growth of the company. But every single day companies battle DoS and DDoS attacks [1]. These attacks create friction and difficult for employees to access the hardware and complete an assigned task [2].

The primary notion of these attackers is to deprive the potential of the resources. When these attacks are performed, the resources tend to perform abnormally. Either the resources malfunction and deliver incorrect results, or these attacks hog the entire resource, and wait time in the job queue is increased exponentially. The attacks have the potential to create a blackout. This results in temporarily shutting down the organization till those resources work efficiently.

Another reason for these attacks happen is to create a back-door access to these resources. During a blackout or temporary shutdown of the company, IT experts reset these resources by changing the pass-code. With spoofing, these pass-codes can be gained and used to access the infrastructure of the company. Not only these attackers can use these machines, but also they have the ability to launch any attack on the company [3].

Everyday Cyber Security experts are trying new methods to permanently stop these attacks. The major problem for these experts is to find a way to identify these kinds of attacks. The objective of this research paper is to implement a simple algorithm/methodology that can detect DoS and DDoS attacks. This algorithm is also easily deployable on a web server that works with less load on the resources. The novelty of this paper is the implementation is not complex, and the model doesn't use any generic machine learning algorithm that has low-detection rate compared to our proposed model. In this paper the followings are the sections. Section 2.3—Markov Model and Markov Chains, Sect. 2.4—Hidden Markov Model, Sect. 3—Proposed Scheme, Sect. 4— Results and Discussion, and finally Sect. 5— Conclusion.

2 Literature Review

This section contains the detailed analysis of various algorithms used by other researchers to detect DoS and DDoS. We have enumerated the types of DNS attacks that are available currently, various types of machine learning algorithms in the following section. We have also explained about Markov Models and Markov Chains and Hidden Markov Model.

2.1 Sub Classes of DNS

The KDD-CUP dataset contains DNS [4] network packets. Even though the target variable is ‘attack.’, they are categorized into 5 sub classes. The sub classes are

1. Neptune Attack
2. Land Attack
3. Teardrop Attack
4. Ping of Death Attack
5. Smurf Attack.

Neptune Attack Neptune attack [5] or SYN flood attack utilizes the three way handshake method for TCP connections. The attacker will send numerous SYN packet with sequence numbers for each packet. The server will respond to those malicious SYN packets with ACK packet and waits for SYN-ACK packet for all those requests. Since these are malicious, and the attacker will never send a SYN-ACK [6] in return, the server will wait for a long time unable to create a time out for the requests. This will fill up the Transmission Control Block (TCB) table causing the server not to take up any more requests from legitimate clients [7].

Land Attack Local Area Network Denial (LAND) [8] attack creates a spoofed TCP SYN packet where it has identical source, destination address and ports. When the targeted machine tries to reply, it gets into a infinite loop because of identical source and destination address, causing a crash in victim’s machine.

Teardrop Attack Teardrop attack [9] works based on discrepancy on the offsets sent via those packets on the header. When the offset of one packet does not match with the next packet, it presents a problem of assembling those packets. This causes the victim computer to crash as it fails to assemble the packets properly based on the packet’s offset.

Ping of Death Attack Ping of Death attack [10] is sending a packet larger than the recommended or accepted size (65535 bytes) to the target machine. The target machine unable to process the packets will become unstable and crash. This happens due to the fact that machine could not re-assemble packets larger than the accepted size, which creates a buffer overflow. This does not happen in modern machines. Ping of Death attack was utilizing an exploit back in 1990.

Smurf Attack A Smurf attack [11] is a sub class of DoS attack, where the attacker will send an ICMP [12] echo requests to a broadcast address that has large number of connections using a spoofed source address. The spoofed address is the address of the victim. Since the methodology of ICMP is to return with a acknowledgement, there will be huge traffic, and this will create an instability to the victim [13].

2.2 Clustering Algorithms

The algorithm that creates clusters based on certain properties or clusters by learning the properties in the data is called clustering. In simple words, these algorithms create groups based on similarities. There are two main types of clustering algorithms which are defined as follows.

Supervised Clustering When a clustering algorithm performs clustering based on the labels or feature name provided in the dataset, it is called a supervised algorithm [14]. It is simple, accurate in clustering. In supervised clustering, the learning will be done before predicting the data. There is no interim learning by the algorithm. The output classes are known beforehand.

Unsupervised Clustering When a clustering algorithm performs clustering on a data that does not have labels or features in the dataset, it is called an unsupervised algorithm [15]. It is relatively complex than a supervised algorithm, but it is faster, and the learning happens in real time. The user cannot predict the number of output classes. It is less accurate than the supervised algorithm [16].

2.3 Markov Model and Markov Chains

Markov Model is a stochastic model [17] which works using sequential data or ordered data. The output of a particular event at a particular time is dependant on the previously generated output. Mathematically, output O at a time t will depend on the output at time t_{-1} [18].

The formula for Markov Model can be depicted as follows

$$\prod_{i=1}^{n} P(q_i|q_{i-1}) \tag{1}$$

With Markov model the following goals are achieved.

1. Pattern Recognition.
2. Prediction and Estimation.
3. Learning more about the sequential data.

The prediction of the output is done by Markov Model. But in order to visualize the prediction as a graph, transition matrix is created. A transition matrix is a bi-directional graph which depicts the probability of going from one state to another. It has states under consideration. These states are connected by arrows, and it shows the probability of going from state Sx to Sy. Another name that transition matrix is called Markov Chains [19].

The mathematical equation of Transition matrix

$$P_{ij} = P(X_{t+1} = j|X_t = i) \tag{2}$$

Consider a classic example of weather in a particular day. There are three weather types under consideration. They are Sunny S_1, Rainy S_2, Cloudy S_3 The probability from one state to another are called as Transition Probabilities. Let's consider a sequence of states (Cloudy, rainy, sunny, sunny, cloudy …). Therefore, in a given day, the weather can be any one of these three weather types. The sequence of this weather conditions can be represented as a state sequence, which is depicted as $q_1, q_2, q_3, \ldots, q_n$ where $q_i \epsilon (Sunny, Rainy, Cloudy)$. The probability of next day's weather can be computed using the Markov property as shown in the equation below.

$$\prod_{i=1}^{n} P(q_i | q_{i-1}) \tag{3}$$

2.4 Hidden Markov Model

Hidden Markov Model (HMM) [20] is also a stochastic model. The difference between Markov model and HMM is that the states in the HMM are hidden. Those states can be identified with certain observations. In the previous example, we can find the transition probability between two states. Let's consider, the weather in a particular day is not known. The weather can be identified with some clues that the person can observe. For example, if the room is not well lit by natural light, then there is a chance that it can be cloudy or rainy. If a person coming to your place carries an umbrella, then we can decide that it's rainy. With these observations, a particular state can be determined.

Let's assume that the days have passed, and the observation sequence is $O = o_1, o_2, \ldots, o_n$ where $o_i \epsilon (Umbrella, withoutumbrella)$. As mentioned earlier, Q is the state sequence. The HMM property is defined as shown in equation below.

$$p(q_1, \ldots, q_t | o_1, \ldots, o_t) = \frac{p(o_1, \ldots, o_t | q_1, \ldots, q_t) p(q_1, \ldots, q_t)}{p(o_1, \ldots, o_t)} \tag{4}$$

A module or web service where HMM is used for malware detection can be deployed anywhere on the internet. This can also be used as a real-time service in companies where it will detect whether an incoming packet is a Malware or not, and simultaneously learn about the packets using HMM. HMM has the potential to understand and interconnect several features in the dataset and for any incoming packet, and analyse with less error rate. This is very vital for any company who uses internet as their primary gateway to access the resources [21].

3 Proposed Scheme

User's browser The user's browser is going to be the entry point to access the model that is deployed in the server. The user will enter details, share their dataset to the web browser. On the website, input fields will be given where user can share these

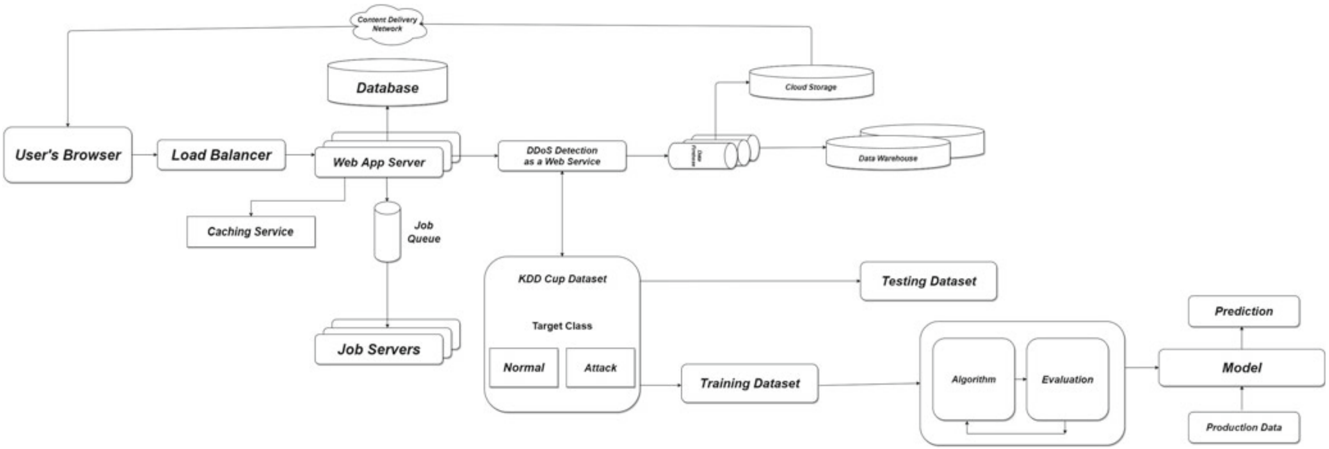


Fig. 1 Architecture diagram of our methodology

details, and at the backend, the script will take care of assigning the values to the variables designated for further action that need to be taken on the dataset.

Load Balancer Load balancer is an entity that acts as a mediator between User Interface and Backend. It efficiently sends incoming requests to multiple servers that are available. Load balancer also takes care of not overloading jobs to a particular server. It makes sure all the servers work at their optimal and efficient level to process the incoming requests [22].

Job Queue This entity takes care of tracking the jobs that needs to be executed. These jobs are user requests. When user submits a request, it will be added to the job queue for further processing. The job queue will maintain the order at which each job need to be processed.

Job Servers This is the entity that processes the requests made by the user. It analyses the request and decides what are the steps that needs to be taken. After analysis, it further creates a framework to proceed for further implementation of the user's request [23].

Database Database is the heart of any web service. It stores all related files. It contains executable files, scripts to process user's requests, metadata, schema for databases.

When a user inputs data in the given fields, the data is passed on to the load balancer, where it efficiently takes a choice of a job queue to send the incoming request. Job queue holds the request until all the previous requests has been processed. Then, the user's request is sent to the job server. Job server does the following tasks in order [24]

1. Check whether all the required fields have a value provided. Further it checks the values contains the same type as the target variable
2. Analyse the request in hand. It understands the request given by the user. In this example, it is to find DoS and DDoS attacks in the dataset given by the user.
3. The process of training and testing the dataset is initiated. Training the dataset is done to understand the dataset, and testing is done to verify whether the understand is correct.
4. Output is generated. In the output, it enumerates the total number of packets that are normal, and total number of packets that are attack packets. It also states how efficient the results are in the form of a confusion matrix.

Our aim is to find out the DoS and DDoS packets using Hidden Markov Model. We require a dataset that has a lot of features, so that we can pick the features that we want instead of being deficit in features. For this project we used KDD-CUP dataset that contains DoS and DDoS packets which are categorized as 'attack.' and 'normal'. This contains all the features required for us to proceed with our research. Further the features present in the dataset are well interconnected which contributes to whether a particular packet is a DoS packet or normal packet [25]. This approach requires very little pre-processing and directly applying the algorithm, followed by prediction. The HMM algorithm is a continuous and supervised algorithm. HMM is based on time series, i.e., the resultant value at time X will be based on the values at time X-1 and X-2 and so on. It is a supervised algorithm because each features are labeled and there is a target label for clustering.

3.1 Dataset

We used the KDD-CUP dataset [26] for our analysis. This dataset is a part of The Third International Knowledge Discovery and Data Mining Tools Competition conducted at University of California, Irvine. The dataset is used for the competition. But for our research, the dataset contains data of both the DoS packets and Normal Network packets. This dataset also has numerous features that contributes to the type of the packet. The entire feature set has been enumerated under Table 1. The obtained dataset is unlabeled [27]. With reference to the website, the labels were available for the dataset.

Since the dataset has numerous features, some of the features are selected for the type of attack under discussion. The reason behind this decision is to make sure the end result is accurate as possible. Features that don't contribute to the attack will contribute to incorrect results. In the results and discussion section, results are discussed with all of the features available in the dataset, and with specific features to DoS and DDoS attacks [28].

The reason of choosing this particular dataset because of its exhaustive feature set and the amount of data itself. Having more data refined the results and brought near accurate results. Compared to other datasets like CICDDOS-2019 and others, this dataset stands out in terms of the amount of data available and multiple features to compare [29]. It also helps refine the learning process of the models used in our paper.

3.2 Implementation

The dataset contains several features including the sender's IP, destination IP, packet length and others. For each data, there is a target label that differentiates between a attack network packet and normal network packet. Given below is a detailed explanation of the step-by-step process of our implementation of our proposed scheme.

Feature Selection and Dataset Split In order to give optimal results, we chose multiple features from the dataset. These features are well connected to categorize whether a packet is normal or attack. The dataset is split into training and testing dataset. 80% of the dataset constitutes to training set, and remaining 20% constitutes to testing set. The reason for splitting this way is to find the perfect balance between understanding the dataset and not to avoid over fitting. The features used can be viewed in Table 1.

Training and testing the model HMM is a very complex model. The reason behind its complexity is stemming from the fact that it takes the past data into consideration to render out the present data. In our lab conditions, training the data consumed a lot of time. But we cannot omit the fact that the model understands the data better than generic machine learning models. The training function is set to the default parameters because changing the value of parameters did not bring any significant

Table 1 Features of KDD-CUP dataset

Feature name	Feature description
Duration	Length of the connection (in seconds)
Protocol_type	Type of protocol (TCP, UDP)
Service	Network service on the Destination (Eg., HTTP, telnet, etc.)
Flag	Status of the connection (normal/error)
src_bytes	Number of data bytes from source to destination
Wrong_fragment	Number of “wrong” fragments
Urgent	Number of urgent packets
Hot	Number of “hot " indicators
num_failed_logins	Number of failed login attempts
Logged_in	1 for successful login 0 for unsuccessful login
num_compromised	Number of “compromised” conditions
num_root	Number of “root” accesses
num_file_creations	Number of file creation operations
is_hot_login	1 if login belongs to “hot list” 0 otherwise.
_guest_login	1 if its a “guest” login 0 otherwise
count	number of connections to the same host as the current connection in the past 2 s
srv_count	Number of Connections to the same service as the current connection in the past 2 s
srv_serror_rate	Percentage of connections that have “SYN” errors (with same connections)
rerror_rate	Percentage of connections that have “REJ” errors (with same connections)
srv_diff_host_rate	Percentage of connections to different hosts
Dst_host_same_srv_rate	Percentage of connections to the same service (Host Level)
dst_host_srv_serror_rate	Percentage of connections that have “SYN” errors (Host level)
dst_host_srv_rerror_rate	Percentage of errors having “REJ” errors (Host Level)

changes in the output of the algorithm. The strength of HMM is understanding the past data to decide the outcome for the present data. We used the remaining 20% of the data to test whether the model understood the training data. It was unprecedented. It was ahead of all the generic algorithms that was used before.

Following section will reveal the comparison between various generic algorithms and HMM.

4 Results and Discussion

4.1 Parameters

While implementing, we used the default values for Logistic Regression, KNN, and HMM. The first reason is that the default values for the parameters showed some unprecedented results. The second reason is that changing the values of the parameters did not show any significant increase in the True Negative Rate value.

We did performance analysis for 2 cases. One with all the features of the dataset, and another with 24 features that contributes to the DNS attacks. We have arrived with some interesting results, and we have concluded that HMM performs better than the Logistic Regression and KNN.

4.2 Performance Analysis

Logistic Regression is one of the generic algorithms used in this paper for comparison. Logistic Regression classifies a dataset into 2 target classes. In this case, it is either a normal network packet or attack packet (DNS). Consider the values for Logistic Regression from Table 2. When considering all features, we get the True Negative Rate of 80.04%. The difference in performance between the two kinds of feature selection is 0.63%. This is stemming from the fact that Logistic Regression is not powerful enough to understand the features. This trend also follows with the KNN algorithm, but there is virtually no increase or decrease when using specific features to detect DNS attack. But, when it comes to HMM algorithm, we can see there is a 12.07% increase in DNS attack detection when using all the features, and 9.57% increase when using features specific to DNS attacks. The increase attributes to the fact that HMM learns about present as well as past network packets to see the trend. This increase is unprecedented in the security industry (Table 3).

Table 2 Performance analysis of all the algorithms (% shows the percentage increase with respect to previous algorithm

Performance analysis									
Feature selection	LR			K-Nearest neighbour			Hidden markov model		
With all features	TPR	TNR	%	TPR	TNR	%	TPR	TNR	%
	19.93	79.40	–	19.13	80.04	**0.64%**	7.88	92.11	**12.07%**
With features specific to DNS (1)	19.93	80.03	–	19.92	80.03	**0%**	10.39	89.60	**9.57%**

Table 3 Performance analysis with Varied Training and testing dataset splits for LR, KNN, HMM [28]

	ALGORITHMS			
Dataset split	Metrics	LR (%)	K-NN (%)	HMM (%)
60% Training 20% Testing	TPR	19.96	19.87	7.78
	TNR	80.01	80.11	92.21
75% Training 25% Testing	TPR	19.96	19.72	7.99
	TNR	80.01	80.26	92.00
80% Training 20% Testing	TPR	19.78	19.98	8.05
	TNR	80.19	80.01	91.94

5 Conclusion

This paper demonstrated how HMM outperform the other two machine learning algorithms under consideration. Detecting DoS and DDoS is not easy with standard machine learning algorithms. It requires understanding of the interconnection of the features and analysing the past data. HMM got an accuracy of 92.11%. HMM was able to detect more packets than generic machine learning algorithms. There is a definite possibility that attackers will try to find a solution where it is hard for any machine learning algorithms to detect these packets, but we as researchers should continuously work on refining the process to make internet a safe place. This model can be deployed in a cloud and can be used as a web service, for variety of datasets available with the developer. The framework and the design for this web service have been shown as a architecture diagram in Fig. 1.

Acknowledgements This work is part of the project work Malware detection using FPGA, Sandboxing and Machine Learning which is supported by IBM Shared University Research Grant.

References

1. B. Joshi, A. Santhana Vijayan, B.K. Joshi, Securing cloud computing environment against DDoS attacks, in *2012 International Conference on Computer Communication and Informatics* (IEEE, 2012)
2. R. Manoj, M. Maruthi, G. Vivek, T. Senthil Kumar, A survey: artificial neural networks in surveillance system, in *IJCA Proceedings on International Conference on Innovation in Communication, Information and Computing 2013 ICICIC* **2013** vol. 1, pp. 19–22 (2013)
3. S. Murugesan, T. Senthil Kumar, U.S. Priyanka, K. Abinaya, Towards an approach for improved security in wireless networks. Int. J. Comput. Appl. ICICIC **2013**(1), 9–13 (2013)
4. Project Resource Files https://github.com/mbalaji777/ibmproject

5. I. Ahmad, A.B. Abdullah, A.S. Alghamdi, Application of artificial neural network in detection of DOS attacks, in *Proceedings of the 2nd international conference on Security of information and networks* (2009)
6. R. Arends, et al. DNS security introduction and requirements. RFC 4033 (Proposed Standard) (2005)
7. Y. Xu, H. Sun, F. Xiang, Z. Sun, Efficient DDoS detection based on K-FKNN in software defined networks. IEEE Access **7**, 160536–160545 (2019). https://doi.org/10.1109/access.2019.2950945
8. D. Kshirsagar, A. Rathod, S. Wathore, Performance analysis of DoS LAND attack detection. Persp. Sci. **8**, 736–738 (2016)
9. W. Eddy, TCP SYN flooding attacks and common mitigations. RFC 4987, August, 2007
10. F. Yihunie, E. Abdelfattah, A. Odeh, Analysis of ping of death DoS and DDoS attacks, in *2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT)* (IEEE, 2018)
11. G. Jinhua, X. Kejian, ARP spoofing detection algorithm using ICMP protocol, in *2013 International Conference on Computer Communication and Informatics* (IEEE, 2013)
12. P. Solankar, S. Pingale, R. Parihar, Denial of service attack and classification techniques for attack detection. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **6**(2), 1096–1099 (2015)
13. G. Palaniappan, S.S., B. Rajendran, et al., Malicious domain detection using machine learning on domain name features. Host-based features and web-based features. Proc. Comput. Sci. **171**, 654–661 (2020). https://doi.org/10.1016/j.procs.2020.04.071
14. S. Fine, Y. Singer, N. Tishby, The hierarchical hidden Markov model: analysis and applications. Mach. Learn. **32**(1), 41–62 (1998)
15. C.F. Eick, Zeidat, N., Z. Zhao, Supervised clustering-algorithms and benefits, in *16Th IEEE International Conference on Tools with Artificial Intelligence* (IEEE, 2004)
16. J. Cheng, J. Yin, Y. Liu, Z. Cai, C. Wu, Detecting distributed denial of service attack based on multi-feature fusion. Secur. Technol. **132–139**(2009). https://doi.org/10.1007/978-3-642-10847-1/17
17. V. Das, V. Pathak, S. Sharma et al., Network intrusion detection system based on machine learning algorithms. Int. J. Comput. Sci. Inf. Technol. **2**, 138–151 (2010). https://doi.org/10.5121/ijcsit.2010.2613
18. S. Anuraj, P. Premalatha, T. Gireeshkumar, High speed network intrusion detection system using FPGA. advances in intelligent systems and computing **187–194**(2015). https://doi.org/10.1007/978-81-322-2517-1/19
19. S. Kumar, Smurf-based distributed denial of service (ddos) attack amplification in internet, in *Second International Conference on Internet Monitoring and Protection (ICIMP 2007)* (IEEE, 2007)
20. X. Xu, Y. Sun, Z. Huang, Defending DDoS attacks using hidden markov models and cooperative reinforcement learning. Intell. Secur. Inf. 196–207. https://doi.org/10.1007/978-3-540-71549-8/17.
21. H. Polat, O. Polat, A. Cetin, Detecting DDoS attacks in software-defined networks through feature selection methods and machine learning models. Sustainability **12**, 1035 (2020). https://doi.org/10.3390/su12031035
22. T. Oo T. Phtu, Analysis of DDoS detection system based on anomaly detection system, in *International Conference on Advances in Engineering and Technology (ICAET'2014) March 29-30, 2014* (Singapore, 2014) https://doi.org/10.15242/iie.e0314146.
23. S. Smys, DDoS attack detection in telecommunication network using machine learning. J. Ubiquitous Comput. Commun. Technol. (UCCT) **1**(01), 33–44 (2019)
24. S. Dong, M. Sarem, DDoS attack detection method based on improved KNN with the degree of DDoS attack in software-defined networks. IEEE Access **8**, 5039–5048 (2020). https://doi.org/10.1109/access.2019.2963077
25. W. Wang, X. Ke, L. Wang, A HMM-R approach to detect L-DDoS attack adaptively on SDN controller. Future Internet **10**(9), 83 (2018). https://doi.org/10.3390/fi10090083/2
26. N. Grira, M. Crucianu, N. Boujemaa, Unsupervised and semi-supervised clustering: a brief survey. A Rev. Mach. Learn. Tech. Proc. Multimedia Content **1**, 9–16 (2004)

27. S. Alhaidari, A. Alharbi, M. Zohdy, Detecting distributed denial of service attacks using hidden markov models. Int. J. Comput. Sci. **15**, 9–15 (5) (2018). https://doi.org/10.5281/zenodo.1467645.
28. V. Bolon-Canedo, N. Sanchez-Marono, A. Alonso-Betanzos, Feature selection and classification in multiple class datasets: an application to KDD Cup 99 dataset. Expert Syst. Appl. **38**(5), 5947–5957 (2011)
29. S.M. Kannan Mani, M. Balaji Bharatwaj, N. Harini, A scheme to enhance the security and efficiency of MQTT protocol. Smart Innovat. Syst. Technol. **79–93**(2020). https://doi.org/10.1007/978-981-15-5971-6/9

M. Balaji Bharatwaj is an incoming Cyberseucity student at University of Maryland, College Park, MD, USA. His research interests are Cyberseurity and Pentesting. He completed his Bachelors in Computer Science and Engineering at Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore.He did the project under the mentorship of Dr. Senthil Kumar. He is part of the project titled “Malware detection using FPGA, Sandboxing and Machine Learning under the IBM Shared University Research (SUR) scheme.

M. Aditya Reddy is currently pursuing Computer Science and Engineering at Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore. His research interests are in Machine learning and Cybersecurity. His research work includes implementing machine learning models (KNN, Logistic Regression and HMM), to detect DoS and DDoS attacks.

Thangavel Senthil Kumar currently serves as an Associate Professor at the Department of Computer Science and Engineering at Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore. His research interests include Video Analytics, Big Data Analytics, Intrusion Detection Systems and Deep learning. He completed his B. Tech. (Computer Science and Engineering) from Sethu Institute of Technology, Madurai. He then completed his M. Tech. (Distributed Computing Systems) from Pondicherry Engineering College, Pondicherry. He completed his Ph.D. in Information and Communication Engineering from Anna University, Chennai. He is part of the project titled "Malware detection using FPGA, Sandboxing and Machine Learning under the IBM Shared University Research (SUR) scheme.

Sulakshan Vajipayajula is a Chief Architect of Security Analytics in the IBM Security Systems division. He was known for the work in Watson for Cyber Security where he was the architect and took the concept to the product offering. He exclusively works on development of AI and ML enabled security analytics platform deployable in multi-cloud environments specializing in Graph analytics and NLP in the W4CS. He worked in the IBM Security CTO office specializing in Network Security, AI, Fraud detection , Cloud Security and Analytics . He is a Lab Advocate for key Customers, enterprise integrator, and avid technical evangelist. Sulakshan has a Masters degree in Computer Science from New Jersey institute of Technology.

Multimodal Biometric Systems, its Security Issues, Research Challenges and Countermeasures—Technical Review

M. Gayathri, C. Malathy, Hari Akhilesh Chandrasekar, and Prabhakaran Mathialagan

Abstract A biometric system is an amalgamation of algorithms that uses an individual's unique biological data innate physical features (fingerprints, retina, face, ears, etc.) or behavioural features (signature, voice, keystroke dynamics, etc.) or both to authenticate the individual. This essentially makes a biometric system function as a pattern recognition system that helps an individual in any kind of adversarial environment. Hence, like any conventional authentication system, it is vulnerable to the attacks of malicious adversaries that try to manipulate the data and thus decline the security of the system by compromising its robustness. These possible attacks have not been considered while developing most biometric authentication systems. Our work describes multimodal biometrics, its feature extraction methods, research challenges and focuses on the major security issues in both unimodal and multimodal biometric systems. This paper summarizes some robust multimodal systems, their countermeasures and algorithms that have been devised to tackle these security challenges.

Keywords Biometrics · Multimodal biometrics · Biometric fusion · Security · Attacks · Countermeasures · Encryption

M. Gayathri (✉) · C. Malathy · H. A. Chandrasekar · P. Mathialagan
SRM Institute of Science and Technology, SRM Nagar Kattankulathur, Kanchipuram, Chennai, TN 603203, India
e-mail: gayathrm2@srmist.edu.in

C. Malathy
e-mail: malathyc@srmist.edu.in

H. A. Chandrasekar
e-mail: cr3487@srmist.edu.in

P. Mathialagan
e-mail: pm3960@srmist.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_75

1 Introduction

Security of data is considered to be vital or present need in all fields due to the advancements in digitization. Biometric authentication involves the process of verification of an individual that depends on the unique characteristics of an individual in order to verify that there are who they say they are. A biometric authentication system uses methods that compare a captured biometric data to stored and confirmed template data that is already verified in a database. If both the templates of captured data and template in the database match, then the verification of the individual is accepted as positive. In other words, only the people who are authorized to access the data can gain access to sensitive data.

Pattern recognition is "the act of taking in raw data and taking an action based on the category of the pattern. Pattern recognition techniques are currently used in several security applications such as biometrics based person recognition, spam filtering and intrusion detection in computer networks, with the goal to discriminate between a legitimate and a malicious pattern class."[1]. Therefore, a computation performed at the core of a biometric system can be considered to be pattern recognition, since it has to process if the data of the person seeking authorization is similar to the enrolled initially data of the person. The accuracy of this recognition can be largely and negatively affected by adversaries when they try to manipulate results by illegitimately gaining access to the Biometric System. With the growing usage of data in real world, there is also a rise in data theft which makes the storage of data in the biometric database itself insecure. Confidentiality is protecting information from being accessed by unauthorized parties. If the confidentiality is breached, the data can not only be used to gain unauthorized access to the system but also illegally cannot be used outside the system. This work explains a basic overview of how a biometric system works and then go on to ponder upon the security issues in the authentication process as well as the data storage process. We also list some robust countermeasures to avoid the breach of this security. Biometric traits can be classified based on the physiological and behavioural aspects of human nature. Face, iris, fingerprint, retina, hand geometry depicts the physiological traits of the individual. The way the individual walks (Gait), way he types (keystroke dynamics) refers to the behavioural traits of individual. The Fig. 1 refers to the classification of biometrics and Fig. 2 represents the various traits available.

2 Components of a Biometrıc System

A basic Biometric System has four major parts to carry out the basic operations. The type of algorithms deployed in each of the parts depends on the trait(s) being used and the type of data processed. Figure 3 depicts the components used in biometric system.

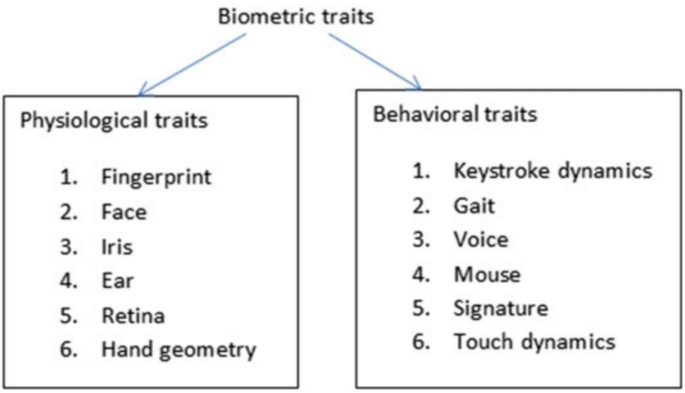


Fig. 1 Classification of traits

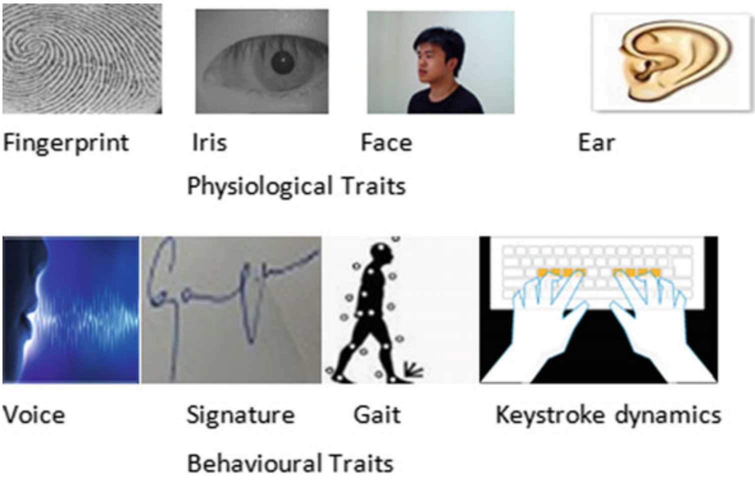


Fig. 2 Different physiological traits and behavioural traits

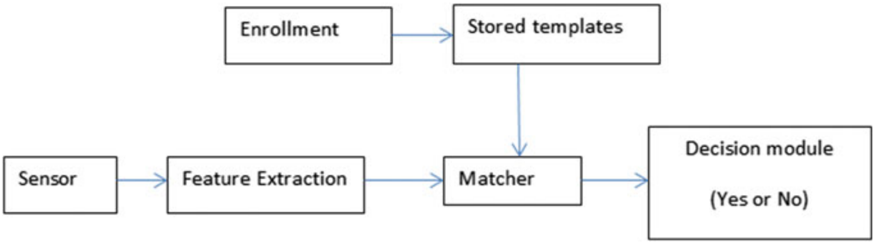


Fig. 3 The components of a simple biometric system

The sensor module, feature extraction, matching module and decision-making module are the four fundamental components. Additional processes add up with additional features.

1. Sensor module: This module is for acquiring the input for further processing of biometric data.
2. Feature extraction: This module focuses on the collected data, to extract important feature vectors to generate a template.
3. Matching module: This module is helpful in comparing feature vectors acquired live with the enrolled user template, to focus on the authentication part.
4. Decision-making module: This module ensures about the final decision about if the user's identity is matched with the enrolled data of the user is made. The user is either given or denied permission for authentication.

3 Multimodal Biometric Systems

Multimodal biometric system is a system which involves more than one biometric trait for user authentication and recognition. The lack of security breaches in unimodal systems has paved the way for multimodal systems where it is difficult to forge as the multimodal system is complex. This system has sensor modules, feature extraction modules, matching modules and decision modules. Two biometric traits, three biometric traits and even four biometric traits are combined together for user recognition using different levels of fusion. Multimodal systems use the combination of two or more traits. The multimodal systems acquire the data from different sources, for example, the multi biometric systems are designed based on the sources from multiple sensors (different sensors for the same biometric trait, eg: using capacitance and optical sensor for the same fingerprint), multiple algorithms (extract features by two different algorithms for the same trait, eg: minutiae and texture based approach for same fingerprint), multi samples (refers to the acquiring of many samples of the same trait eg: many samples of same finger), multiple instances (refers to take multiple instances of traits, eg: all ten finger instance) and multiple traits (different modalities of different traits, eg: face, fingerprint) [2] (Fig. 4).

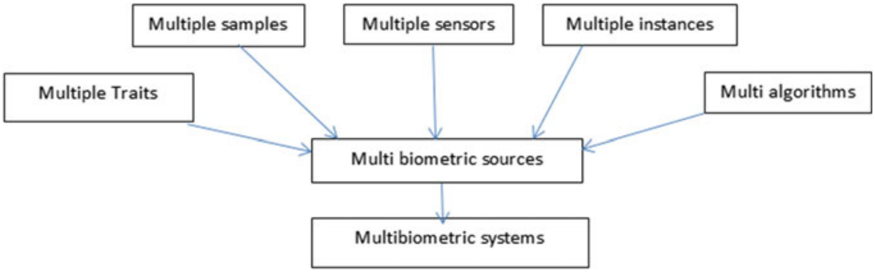


Fig. 4 Multibiometric systems

4 Types of Fusion

In multimodal Biometrics system, the data is acquired using various types of sensors and the acquired data is being cleaned by the preprocessing techniques and the required features are obtained and saved in the form of templates in database. For fusing the multiple data's obtained from different types of modes, few fusion techniques are available. Based on the specific requirement of the application the fusion method is decided. The fusion methodology is carried out in two ways (i.e.) fusion being carried before the matching module phase and fusion carried after matching module phase. Feature level fusion and sensor level fusion are the methods of fusion carried out before the matching module phase, match level fusion and rank level fusion, decision level fusion are carried out after the matching module phase [2].

Feature level fusion

Features are extracted from the various biometrics systems. Feature sets extracted from the various sources are combined into single feature vector and it is stored as a template in database. The space requirement is the limitation of feature level fusion. The space requirement of feature vector generated after fusion is large. It is termed as "curse of dimensionality". Normalization is done on the various features to obtain the final single feature vector.

Sensor level fusion

The data obtained from direct source of variety of sensors are fused together. This is not suitable fusion method in many cases. Multiple sensors may have different resolutions in their equipment which may not match during fusion phase. Sensor level fusion helps in capturing multi samples of single biometric trait for the same person.

Rank level fusion

The fusion is done after the matching process based on the ranks assigned to each user data. The fused score is calculated and ranks are assigned. The methods involved are Borda count, Logistic Regression and Highest rank. It is carried out in identification mode.

Decision level fusion

The matcher of different biometric subsystems decides the best match at the output individually. The decision level fusion is done on the decisions of individual subsystems by various methods like AND Rule, OR Rule, Majority voting, Bayesian Decision, weighted Majority voting, Dempster-Shafer Theory of Evidence and behaviour knowledge space.

Match score level/measurement level/confidence level

The match score is a score that tells about the similarity measure of traits, these scores are fused together again to arrive at a final decision of the match level is called the match score level. There are many rules adopted in match score level fusion product rule, sum rule, max rule, min rule, median rule. Techniques like density based score fusion or transformation based score fusion or classifier based score fusion is carried out after normalizing the score levels by decimal scaling or double sigmoid functions.

5 Machine Learning and Neural Networks in Multimodal Biometrics

The fusion of biometrics can take place using machine learning algorithms, which in turn that ensure better recognition and better security. Popular Supervised Algorithms used for fusion at score levels as proposed in [3] are:

Support Vector Machines (SVM): The input score's dimension can be expanded using a kernel function and then according to the kernel function we can choose optimal cost functions and variance parameters. After which training of the SVM can take place. During which the cost function and variance parameter can be tuned to ensure the classification of the input as either imposter or genuine with high accuracy.

Fuzzy Expert System (FES): We can use a suitable FES to classify the input as an imposter of genuine using a TSK FES. We can use segmentation to create three dimensional fuzzy rules. The output of each of the fuzzy rules would be a linear function of the classifiers in the unimodal systems. Then an estimation function can be used to compare the obtained result with that of a threshold to tell whether the input is from an imposter or is from a genuine user.

Gaussian Mixture Model (GMM): Using multiple gaussian distributions as different subclass we can model a Bayesian Classifier to classify the given input as either imposter or imposter of genuine. The probability density function for the same can be taken as a weighted sum of Gaussians. The weights can be tuned after experimentation.

Artificial Neural Network (ANN): We can use a Neural Network with the input neurons as the number of unimodal outputs and with one output neuron to tell if the inputs are from imposter or from an authorized user. There can be different numbers of hidden layers and neurons to process the data. We can use different algorithms to change the bias of each neuron, an example would be the back-propagation algorithm.

Apart from these Convolutional Neural Networks can be used at decision level fusion or feature fusion as shown in [4] where they use a set of CNN layers to extract features and use that to tell whether the user is fake or genuine at different levels. This not only proves that there are multiple techniques that can be used but that these techniques can be used across different fusion levels to build a better multimodal biometric system.

6 Feature Reduction Methods

Analyzing a lot of features especially one with high dimensionality can prove to be difficult especially in a biometric system. Also, as dimensions increase it leads to “curse of dimensionality problem” when Machine Learning Methods are used, etc. Hence to make the system faster we can use reduce the number of dimensions of the features. The process by which dimensions of a biometric feature is reduced is known as feature reduction. In general, the techniques that can be used to reduce dimensions can be classified as two, linear and non-linear [5]. The most commonly used are highlighted below:

Linear Reduction:

Principal Component Analysis (PCA): It is a technique that is unsupervised in nature and used for linear dimension reduction. It uses eigen vectors to create projections of multiple dimensions onto the first few dimensions. It usually focuses on maximizing variance.

Linear Discriminant Analysis (LDA): It is a supervised technique wherein the dimensions are reduced using eigen vectors but it also considers the subspaces or classes while reducing the dimensions. It focuses on reducing the variance and maximizing the distance between classes.

Independent Component Analysis (ICA): Though somewhat similar to PCA, it distinguishes itself on independent components. It does not impose orthogonality as in PCA. To use this as a dimension reduction method we need to ensure that it is under completed. One such example is shown in [6].

Non-Linear Reduction:

IsoMap: Isomap is a non-linear technique reduction technique where geometric properties of data are preserved. It does this by collecting points from which a neighbourhood graph is created. Now it computes the shortest path between the nodes and then applies multidimensional scaling to the obtained distance matrix. Hence reducing the dimensions as required.

Locally Linear Embedding (LLE): The local characteristics and properties technique of the data is tried to be preserved in this technique. This is achieved by first identifying the nearest neighbours to each point and then reconstructing weights to maximize the distance between all non-adjacent points. Now we compute the vectors based on the weights and bottom eigen values. By following the steps, we are able to reduce the dimensions.

Laplacian Eigenmaps: This technique starts similar to Isomap and LLE where we choose nearest neighbours after which a weighted graph is created. After this, a kernel is constructed and then eigen decomposition of this kernel is done, which is the required feature with its dimension reduced [7].

7 Modalities Combined

The Tables 1, 2 and 3 describe the combination of more than one modality for the biometric fusion with different fusion levels and methodologies adopted (Table 4).

Table 1 Different levels of fusion

Fusion types	Operation	Process	Methods adopted
Feature level	Feature vector extracted	Carried before matching	Normalization
Sensor level	İmage level fusion	Carried before matching	Mosaicing
Rank level	Ranks are assigned for fused score	Carried after matching	Borda count, logistic regression
Decision level	Decisions assigned by individual matchers	Carried after matching	AND rule, OR rule
Score level	Match score is obtained	Carried after matching	Sum rule, min rule

Table 2 Two modalities combined

Biometric fusion	Methodology adopted	Fusion level	Database used
Face and hand	Supervised descend method for face. Sparse representation classifier	Feature level fusion	Own database collected at Aalborg university (69 subjects)
Fingerprint and online signature	Summation rule And Euclidean distance calculation	Feature level fusion	ICDAR 2001
Face and Iris	2D LOG Gabor filter and spectral	Hybrid (feature and score level fusion)	CASIA Iris database

Table 3 Three modalities combined

Biometric fusion	Methodology adopted	Fusion level	Database used
3Dface, soft biometrics (full body image), face image	FVF, SVM	Sensor level fusion	QUIS-CAMPI
Ear print, fingerprint and palmprint	Mean Extrema based confidence weighting Technique (MEBCW)	Score and feature level	AMI, FVC2002, IITD_Right
Iris, vein and fingerprint	Backtracking search optimization algorithm	Score level fusion	FVC2006, VERA, IITD_PolyUris

Table 4 Four modalities combined

Iris, face, palm print and signature	Encoded DWT And Principal Component Analysis	Feature level fusion	CASIA

8 Limitations of Biometric System Causing Threat to the Security

Unimodal Biometric Systems:

These systems make usage of only one biometric trait for the purpose of verification. Some of the major authentication issues and their causes are mentioned in the following sections.

Noise in sensed data:

Defective and inappropriately maintained sensors in the biometric system may lead to the development of noise in it. Dirt accumulation or fingerprints remnants resulting in noisy fingerprint data can form an example. When there is noise in the test data, it cannot be compared correctly for honest users, with the corresponding templates saved from before, leading to false negatives. There are also chances of false positives- imposters' test samples getting matches with the template in the database. Thus, all these lead to a significant lessening in the efficiency of the authentication of the system [1, 8].

Intra-class variations:

These kinds of variations are mostly caused due to the difference in environmental conditions surrounding the user when the data is being collected during the authentication test and when the user originally submitted the biometric samples. Examples can include illumination changes during a facial recognition biometric system, usage of different sensors during the two processes or natural changes that occurred to the user's biometric traits (raging, accident, etc.) [9]. Large rates of intra-class variation can lead to the decrease of the Genuine Acceptance Rate (GAR) of the biometric system.

Intra-class similarities:

When the biometric samples overlap, in feature space, it leads to multiple classes or individuals [1]. This is called Inter-class similarities. This scarcity of enough individuality in the feature sets of the biometric traits lead to a growth in FAR (False Acceptance Rate) of the system [1]. Thus, it becomes imperative for the presence of an upper limit on the number of users to be authenticated by a Biometric System.

Non Universality:

Each person on whom the biometric system has been deployed, for authentication must be able to have the features of the corresponding biometric trait correctly recognized. When this gets defied, a state of non-universality is reached within the biometric system. It may not be capable of detecting the analyzing the required biological data from a percentage of users. An example would be that of the National Institute of Standards and Technology (NIST) reporting the "inability to extract correct minutia features from the fingerprints of two percent of the population (manual workers with many cuts and bruises on their fingertips, people with hand related disabilities, etc.), due to the poor quality of the ridges" [10]. Non universality gives rise to the FTE! (FTE!) (failure to enrol) rate. "No Biometric trait is truly universal."[11].

9 Attacks Agaınst Bıometrıc Systems

Hill Climbing Attack.

The attacker obtains information from the matcher after a synthetic sample is fed and uses the same for generating a better synthetic sample or the attack can target the feature extraction module where the data generated resembles an original acquired biometric. So, we can tell that the attack is performed in an iterative manner. This type of attack unlike the presentation or masquerade attack doesn't require the attacker to have knowledge of the biometric system. Powerful approaches that are used in this attack are the "Hooke-Jeeves approach", "Simultaneous Perturbation Stochastic Approximation (SPSA)", "Nelder-Mead (NM)" [12]. This attack can generally be mitigated by using suitable complex template protection and authentication schemes.

Presentation Attack.

Presentation attack is an attack where an object or image is used to impersonate a biometric to gain access or is done to evade recognition. This is usually done using a Presentation attack instrument. These instruments can vary from biometric to biometric a few examples would be a fingerprint overlay, 3D silicone Mask, image of fingerprint vein, etc. Since this attack is being conducted directly on the sensor any sort of digital authentication or protection schemes are proven to be useless. However, these attacks can be prevented by detecting the liveliness of the biometric. Liveliness can be anything that a physical human body produces that the device or inanimate object cannot produce. A simple example of this would be detecting short wave infrared waves produced by our body using a "Short Wave Infrared Range (SWIR)" band detector [13].

Morphing Attack.

A variant of the presentation attack is the morphing attack where multiple samples of different subjects are combined to form a synthetic sample and this enables the hacker to gain access to the data [14]. This has gained quick popularity due to efficiency and hence is identified as an individual attack on its own. The process of morphing varies from biometric to biometric, an instance of face recognition in which morphing takes place by four steps namely correspondence where landmarks are used, warping where the distortion in the images are removed and they are aligned, blending and finally post processing where image quality is improved, correction of colour gradients take place, etc. There are multiple libraries and software through which morphing can take place. Higher quality morphing usually manual in nature leads to a greater chance of being able to gain access. The lack of benchmarks along with overfitting of existing morph detection systems make detection and mitigation of morphing attacks hard [15]. That being said liveliness detection in combination with complex template protection schemes can possibly mitigate the attack.

Inversion Attack.

Attacks done by reversing or inversing a biometric template are called inversion attacks. These attacks obtain a template, reverse it to find the original biometric and present it to the system to gain access to the system. However, the attacker requires detailed knowledge of the system that he has targeted to ensure that his attack is successful. Since biometric systems usually give more than one chance for the user to authenticate himself, the attacker can continue to give multiple attempts at this till he gains access. Depending on the way the inversion algorithm works many samples can be produced to help the attacker with his multiple attempts. These types of attacks can be mitigated by ensuring that the template produced by the system is irreversible and unlinkable. While the former means that the template shouldn't generally produce the original biometric on being inverted, the latter means that even if one biometric gives multiple templates, one shouldn't be able to trace back its original source by any means [16].

Replay Attack.

These are attacks that are done by sending the same information as the original. These are generally obtained by eavesdropping on the communication channel between the biometric scanner and processor [17]. Other methods of replaying exist; an example would be replaying recorded passphrase audio to an audio based biometric system as said in [18]. At times replay attacks directly take place by inserting data directly onto the processing system rather than actually sending the same data through the sensor. As mentioned in [17] mitigation of such attacks is possible by adding random fake features and ensuring a strong sharing protocol. The true way to ensure replay attack doesn't take place is actually biometric dependent as such replay attack detection differ from biometric to biometric and good detection of replay attack will result in stopping the attack.

Side Channel Attack.

Side Channel Attack is an attack where attackers obtain information without actually attacking the biometric system. Side Channel Attacks use measurable physical effects to reveal secret or sensitive information. Combinations of multiple side channel attacks can take place to ensure greater success in obtaining vital information. We can prevent these attacks by eliminating leakage of nonessential information and trying to remove or reduce the interdependence between the secret data and actual data exposed to the outside world. Other physical measures such as "Shielded Circuits", etc. can also be taken to prevent this attack. Though software countermeasures exist they may prove to be computationally costly. Two software countermeasures that are specific to biometrics are blinding and masking. The principle that is common to both of them is randomizing the input data which is collected. Apart from which "unlinkability" and "irreversibility" may also prove to be useful in mitigating such attacks. Many biometric systems fail to consider this attack into consideration and this can prove to be fatal as this attack is an upcoming attack [19]. Different attacks and its preventive measures are shown in Table 5.

Table 5 Various types of attacks and their preventive measures

Attack type	Description	System knowledge required by attacker	Preventive measures
Hill Climbing Attack	A synthetic sample is fed and information obtained from the matcher is used by attacker to make the sample better. This process is repeated till successful breach	No	Suitable Complex Protection Schemes and Authentication Schemes can be used
Presentation Attack	A presentation attack instrument is used to present an impersonated biometric	Yes	Detecting Liveliness of Biometrics can be used
Morphing Attack	Multiple samples are combined or morphed together to form a new sample. This new sample will be used by the attacker to gain access	Yes	Liveliness detection with complex template protection schemes can mitigate attack
Inversion Attack	A Template is reversed to generate the original biometric and this generated biometric is given for authentication	Yes	Ensuring that the template is unlinkable and irreversible can stop this attack
Replay Attack	Attackers Presents information obtained from the original either through eavesdropping or other means, to gain access	No	Addition of random fake features to the original information and strong sharing protocols can mitigate this attack
Side Channel Attack	Physical effects of the system are measured to gain information about the biometric system. When multiple such instances provide the necessary information, the attacker is able to gain access using the obtained information	Yes	Physical countermeasures to avoid leakage, reducing interdependence on information between the physical world and the system, ensuing unlinkability and irreversibility can mitigate the attack

10 Data Securıty Breach Countermeasures

Data that is enrolled in a biometric system is stored predominantly as images that are encrypted. Unless data is stored in a distributed storage system, it is always in danger of data breach. Some algorithms to keep the data secured are listed below:

Visual Cryptography: This algorithm was first introduced in the 1990s but has had several advancements since then. This is a method by which data that can be perceived by human eyes is protected. During enrolment, this method is used to break up the data into different parts which are in turn stored in different storages including servers and user's end. To be able to access all the parts the decryption key is required which can be accessed only by an authorized person. If only partial data is present, it is almost useless as nothing can be done to obtain the full data.

Blockchain: Data that can be exposed through storage in normal database or do to due to channel interception can be prevented using blockchain as a storage alternative [20]. It can also be used to avoid an attempt by the attacker where he tries to insert a fake template in the biometric database itself [21]. Since blockchain generally uses "InterPlanetary File System (IPFS)" to store data in a distributed manner, data can be split and stored [20]. It is generally recommended to encrypt the template before storing in IPFS to make it immutable across the system. After storage in the IPFS the returned hash can be stored in a blockchain such as Ethereum to ensure better access and security. For verification, the hash obtained from Ethereum can be cross verified by reuploading the biometric to IPFS and checking the hash value with that of in the Ethereum Network [21].

There also exists algorithms by which one can keep the data safe even if a data breach occurs, these are known as template protection schemes. So, data can be stored safely even if not stored in a distributed manner through its recommended to combine both distributed storage methods along with strong template protection schemes. Some of the common Template Protection Schemes are:

Fuzzy vault for Biometric Encryption: Biometrics that are obtained are stored in an encrypted format so as to protect them from data leakage. Generally, such encryption takes place either in the form of key generation, key binding or key release. Fuzzy vault is a scheme implemented for the key binding type of biometric encryption. In the fuzzy vault, data is generally encrypted using a polynomial expression which is generated from a crypto key given by the user. This polynomial expression is used to generate chaff points from the biometric, now these generated chaff points are appended to the original template to form a template in the format of a polynomial system with both chaff and original points. This is repeated till a good ratio typically 10:1 is achieved. Now these points in a *x*–*y* plane is the fuzzy vault. During the authentication phase instead of repeating the whole process, one can compare the closest matching vault member to the given biometric to get authenticated and hence generating the initial polynomial expression which can be used by the system for its further use such as decrypting data encrypted by the biometric. However, depending on the complexity of the polynomial one can perform a statistical attack followed by a brute force attack to gain access to the original biometric, so generally, this type of encryption needs to have a very strong polynomial to ensure utmost security [22].

Decoying using Honey templates: This more than protecting the data is more about deceiving the attacker by having multiple templates which are fake (Honey templates) along with the original (Sugar template) in the server (Honeypot). This collection of fake and original templates (Sweet templates) is not only used to prevent attacks but also detect attacks but also lure and detect attackers. This in general when combined with an effective data structure such as bloom filter, which in turn is manipulated using a random feature generator can result in a stronger secure system with efficient storage space usage [23]. When this combination is used properly as in [23] one can generate a system that is not only hard to crack and detects attack one can strengthen the biometric template protection by ensuring irreversibility and unlinkability.

Cancellable biometrics: Biometrics once stored are unchangeable, i.e. per person, there is only one biometric which can be created from a trait, thus making revoking impossible. So, if by any chance biometric is leaked it will cause a huge problem. This can be rectified using cancellable biometrics. Cancellable biometrics is created by adding intentional and systematic distortions repetitively. In case the biometric is exposed, then just by changing distortion characteristics, one can easily map a biometric to a new template. The main aim of cancellable biometrics is to ensure "diversity", renewability", "Non-invertibility" and "Performance". Generally, to attain cancellable biometrics distortions are applied by transforming the template at Signal Level or Feature Level. At signal level, the distortion is performed on the data obtained by the sensor directly after collection. A few popular techniques at this level is grid morphing, block permutations. At feature level, distortions are made at the set of features. Algorithms applied at this level can be classified broadly as biometric salting, wherein an external independent key is used to create the salt which is added to the feature set and non-invertible transformation wherein a set of algorithms ensure that distortion takes place based on formulae and input variables, so in case of generation of new template one can simply change the input variables. However Cancellable biometrics may be vulnerable to attacks such as stolen tokens, replay attacks, hill climbing attacks, etc [24].

Feature Transformation: Biometrics heavily rely on features to do anything right from encryption to granting access, etc. These features are usually specific to certain types of biometric for instance minutiae is a feature that is available in fingerprints. Now the attackers who try to crack a biometric system for unwarranted needs try multiple attacks and many of the attacks depends on the attacker somehow managing to get features there are similar to these or try to obtain a template. These template features, when fed back in, will grant them access however by transforming features of a biometric at the processing level even if the attacker obtains a template one cannot gain access due to the template actually being a modified version of the original template. Apart from protecting the system feature transformation can ensure revocability, diversity, security and good performance along with unlinkability [25]. A example of this type of protection scheme is showed in [25] where minutiae of fingerprints are moved based on their orientation with respect to a singular reference point and user specific keys.

Table 6 Storage of data in a secure manner

Storage Algorithm	Description
Visual Cryptography	Data that can be perceived by human eyes is broken up and stored at various storage spaces. These can be only accessed using a particular decryption key, without which only partial data is available. Partial Data is almost useless for usage
Blockchain	Data can be encrypted, split and stored in blockchain, generally since blockchain store data using IPFS data can be stored there and the returned hash can be stored in a Blockchain like Ethereum. Since blockchain is used here, leakage or modification of data is very hard hence ensuring that attacker cannot access the data

Hardware Based Encryption: Even though the world is moving more towards a digital oriented approach, physical features always have something to offer. In such a case for encrypting our biometrics, we can use physical devices such as smart tokens and Physically Unclonable Functions (PUF) devices to ensure security. Taking the example of PUF, even while creating it there are variances even if the same person is creating both of them. This is mainly due to factors like atmospheric pressure, source material, etc. A popular type of PUF is the oxide rupture PUF wherein the PUF is generated by the randomness from natural gate oxide during the manufacturing process of the IC. The drawback of this method is that if by chance the PUF or Smart Token is lost and there is no other workaround implemented there are chances of losing access permanently in the biometric system. Two different storage algorithms and its storage measures are elaborated in Table 6. Various characteristics of data protection and its countermeasures are elaborated in Table 7.

11 Inference from the Survey

It is seen that a multimodal system based on many modalities uses more feature level fusion. This is crucial since many attacks target features of the biometrics and hence the attacks that can take place against a multimodal system reduce. With that being said if a strong template protection scheme is chained in with this, then the chances of an attack bypassing the system may go down significantly. Furthermore, combinations of template protection schemes can also be applied to make the system even stronger.

A simple example might be that against Hill Climbing attack, a two or more-modality biometric system along with Feature Transformation Template Protection Scheme and Blockchain for storage can lead to the mitigation of hill Climbing attack. This is because apart from guessing out what the templates used are, the attacker must also guess how many templates were used. To worsen it for the attacker, since the template is stored in blockchain it cannot be obtained by eavesdropping neither can it be replayed. Even if the attacker somehow manages to get his hands on the

Table 7 Countermeasures which help even if data is leaked

Template Protection Algorithms	Description	Weakness	Efficient against
Fuzzy vault for Biometric Encryption	Encryption of obtained biometric data is done using key binding. A polynomial expression is used to generate chaff points, which in turn are added to the template. Now during verification phase, the closest members can be obtained and those points can be used to authenticate the user	One can perform statistical attack followed by a brute force attack to gain access to the original biometric Hill Climbing Attack can also be used to break the system	Presentation Attack, Morphing Attack, Side Channel Attack, Inversion Attack
Decoying using Honey templates	Fake templates can be inserted along with the original encrypted template to confuse attackers. This can also be used for attack detection	Iterative form of statistical attack using templates obtained from database can make attackers gain access Also prone to repeated hill climbing attacks	Replaying attack using templates obtained from database breaches or by eavesdropping
Cancellable biometrics	Intentional and systematic distortions are induced by the system to make the biometric unlikable and irreversible. This also ensure in case of leak the biometric data can be replaced without having to use another biometric, i.e. allows generation of multiple templates from one biometric	System will be vulnerable to stolen tokens, replay attacks, hill climbing attacks	Morphing Attack, Side Channel Attack, Inversion Attack
Feature Transformation	Transformation of features of the biometric takes place in this countermeasure. The features obtained from original template are modified based on different algorithms, hence making the template unlikable, irreversible and diverse in nature	System will be vulnerable to stolen tokens, replay attacks, hill climbing attacks	Morphing Attack, Side Channel Attack, Inversion Attack

(continued)

Table 7 (continued)

Template Protection Algorithms	Description	Weakness	Efficient against
Hardware Based Encryption	Physically Unclonable Functions (PUF) device is used to encrypt the biometric template or is used for two factor authentications. PUF is uniquely made and vary due to their creation process and creation environment. Hence, they cannot be duplicated or obtained easily by an attacker	If by chance the PUF is lost then users lose access permanently to the biometric system. Also, if the attacker gets hold of the PUF then a replay attack can be taken place	Presentation Attack, Morphing Attack, Side Channel Attack, Inversion Attack, Hill Climbing Attack

template, due to feature transformation he might not be able to figure out the original biometric. Thereby in this simple example, we can say that we have put up multiple blockades to the attacker.

12 Conclusion

This extensive report on biometric systems, multimodal biometric systems, types of attacks and countermeasures inspire new techniques for enhancing security apart from strengthening existing systems and reduce privacy breaches around the world. This paper provides a platform to learn about multimodal biometrics and biometric security. This work may lay a solid foundation for researchers in analyzing the multimodal systems with respect to their endeavours or application. This can also be eventually helpful in general Data Privacy related work in any product space.

References

1. S.G. Bhable, A survey of security of multimodal biometric systems. Int. Journal of Eng. Res. Appl. **5**(12), 67–72 (Part—4) (Dec 2015) ISSN: 2248-9622
2. A.K. Jain, A.A. Ross, K. Nandakumar, Introduction to biometrics. ISBN 978–0–387–77325–4, DOI https://doi.org/10.1007/978-0-387-77326-1, (Springer, New York, Dordrecht, Heidelberg, London)
3. I.G. Damousis, S. Argyropoulos, Four machine learning algorithms for biometrics fusion: A comparative study. Appl. Comput. Intell. Soft Comput. **2012**, Article ID 242401, 7 (2012) https://doi.org/10.1155/2012/242401

4. M. Hammad, Y. Liu, K. Wang, Multimodal biometric authentication systems using convolution neural network based on different level fusion of ecg and fingerprint. IEEE Access **7**, 26527–26542 (2019). https://doi.org/10.1109/ACCESS.2018.2886573
5. I. Rida, N. Al-Maadeed, S. Al-Maadeed et al., A comprehensive overview of feature representation for biometric recognition. Multimed. Tools Appl. **79**, 4867–4890 (2020). https://doi.org/10.1007/s11042-018-6808-5
6. D. De Ridder, R.P. Duin, J. Kittler J, Texture description by independent components, in *Structural, Syntactic, and Statistical Pattern Recognition* (Springer, 2002), pp 587–596
7. J. Wang, Laplacian eigenmaps, in *Geometric Structure of High-Dimensional Data and Dimensionality Reduction* (Springer, Berlin, Heidelberg, 2012). https://doi.org/10.1007/978-3-642-27497-8_12
8. S. Bashir, S. Sofi, S. Aggarwal, S. Singhal, Unimodal & multimodal biometric recognition techniques a survey. IJCSN Int. J. Compute. Sci. Netw. **4**(1) (Feb 2015)
9. M.D. Garris, C.I. Watson, C.L. Wilson, Matching performance for the US- Visit IDENT system using flat fingerprints. Technical report, 7110, National Institute of Standards and Technology (NIST) (July 2004)
10. B. Biggio, Adversarial pattern classification. PhD thesis, University of Cagliari, Cagliari (Italy) (2010)
11. C. Wilson, A.R. Hicklin, M. Bone, H. Korves, P. Grother, B. Ulery, R. Micheals, M. Zoepfl, S. Otto, C. Watson, Fingerprint vendor technology evaluation 2003: Summary of results and analysis report. Technical Report NISTIR 7123, National Institute of Standards and Technology (NIST) (June 2004)
12. E. Maiorana, G.E. Hine, P. Campisi, Hill-climbing attacks on multibiometrics recognition systems. IEEE Trans. Inf. Forensics Secur. **10**(5), 900–915 (2015)
13. R. Tolosana, M. Gomez-Barrero, C. Busch, J. Ortega-Garcia, Biometric presentation attack detection: beyond the visible spectrum. IEEE Trans. Inf. Forensics Secur. **15**, 1261–1275 (2020)
14. U. Scherhag et al., Biometric systems under morphing attacks: assessment of morphing techniques and vulnerability reporting, in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)* (Darmstadt, 2017), pp. 1–7
15. U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, C. Busch, Face recognition systems under morphing attacks: a survey. IEEE Access **7**, 23012–23026 (2019)
16. M. Gomez-Barrero, J. Galbally, Reversing the irreversible: a survey on inverse biometrics. Comput. Secur. **90**, 101700 (2020), ISSN 0167-4048
17. I. Hazan, O. Margalit, L. Rokach, Securing keystroke dynamics from replay attacks. Appl. Soft Comput. **85**, 105798 (2019), ISSN 1568-4946
18. R.K. Das, H. Li, Instantaneous phase and excitation source features for detection of replay attacks, in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (Honolulu, HI, USA, 2018), pp. 1030–1037
19. J. Galbally, A new Foe in biometrics: A narrative review of side-channel attacks. Comput. Secur. **96**, 101902 (2020)
20. O. Delgado-Mohatar, J. Fierrez, R. Tolosana, R. Vera-Rodriguez, *Blockchain and Applications*, Vol. 1010, ISBN: 978-3-030-23812-4
21. M.A. Acquah, N. Chen, J.-S. Pan, H.-M. Yang, B. Yan, Securing fingerprint template using blockchain and distributed storage system. Symmetry **12**(6), 951 (2020)
22. M. Khalil-Hani, M.N. Marsono, R. Bakhteri, Biometric encryption based on a fuzzy vault scheme with a fast chaff generation algorithm. Future Gener. Comput. Syst. **29**(3), 800–810 (2013), ISSN 0167-739X
23. E. Martiri, M. Gomez-Barrero, B. Yang, C. Busch, Biometric template protection based on Bloom filters and honey templates. IET Biometrics **6**(1), 19–26 (2017). https://doi.org/10.1049/iet-bmt.2015.0111
24. H. Kaur, P. Khanna, Biometric template protection using cancelable biometrics and visual cryptography techniques. Multimed. Tools Appl. **75**, 16333–16361 (2016)
25. S.S. Ali, I.I. Ganapathi, S. Prakash, P. Consul, S. Mahyo, Securing biometric user template using modified minutiae attributes. Pattern Recogn. Letters **129**, 263–270 (2020), ISSN 0167-8655

Diabetes Prediction by Artificial Neural Network

R. Ranjitha, V. Agalya, and K. Archana

Abstract Diabetes is a syndrome caused by the hyperglycemia of multiple chronic combined with the variation of carbohydrate, fat and protein metabolism, which impact the improper discharge of insulin and the proper usage of insulin in the human body or both. Diabetes affects more than 463 million people globally. By 2020, 88 million people in Southeast Asia are suffering from this illness. According to a report by the International Federation (IDF), India has 77 million people out of 88 million affected individuals. Therefore, diabetes is one of the growing health concerns in India and has no persistent heal. Therefore, rapid diabetic perception is essential, and it can be done inexpensively through the computation method. The research is carried out for the detection of diabetes by artificial neural network (ANN). Here, the prediction is based on back propagation algorithm of an ANN model for diabetes analysis. For training and testing, the dataset was obtained from the UCI machine learning repository's Pima Indian Diabetes Dataset (PIDD). The network was built with different neurons at various epochs and observed that the accuracy reaches up to 99.23%.

Keywords Machine learning (ML) · Artificial neural network (ANN) · Pima Indian Diabetes Dataset (PIDD) · Back propagation algorithm · Neuron · Accuracy · Regression

R. Ranjitha (✉) · V. Agalya
Department of EEE, CMR Institute of Technology, Bengaluru 560037, India
e-mail: ranjitha.n@cmrit.ac.in

V. Agalya
e-mail: agalya@cmrit.ac.in

K. Archana
Department of EEE, Cambridge Institute of Technology, Bengaluru 560037, India
e-mail: archana.eee@cambridge.edu.in

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_76

1 Introduction

According to a World Health Organization (WHO) study, 1.6 million individuals die each year as a result of diabetes. Diabetes is one the major common diseases found in recent decades. In India, diabetes is a budding challenge with the estimated 8.7% of diabetic population in the age group of 20–70 years, which is projected to rise to 134.3 million by the year of 2045. Hence, India is often named as the diabetes capital of the world. Especially, in metropolitan cities, the prevalence of diabetes has increased from 2% in 1970 to 20% at present, and the rural areas are also following up [1].

The WHO has classified diabetes into four types, namely Type 1, Type 2, gestational diabetes, and other types such as monogenic diabetes, pancreatic diabetes, and drug included diabetes. The two most important diabetes out of these are Type 1 and Type 2 diabetes. Type 1 diabetes is mainly caused due to the damage of pancreatic beta cells resulting in the deficiency of insulin in the body, thus requiring insulin for survival and good health. Type 2 diabetes is specified by insulin shortage and abnormal secretion of insulin, either of which may predominate. The prevalence of Type 2 diabetes has risen dramatically over the past three decades [2].

Intense hunger, frequent urination, and intense thirst are among the symptoms of high blood sugar in the human body. The level of glucose in the human body is range of 70–99mg/deciliter. If the glucose level is more than 126mg/dl, it indicates diabetes. If the blood glucose concentration is between 100 and 125mg/dl, a person is considered to be prediabetes. Suppose the level of sugar increases, major complications will happen which leads to nerve damage, heart disease, stroke, and kidney failure [3].

The health issues like microvascular and macrovascular complications are caused by the effect of long-term diabetes. Microvascular complications affect small blood arteries, producing problems in the eyes, kidneys, feet, and nerves [4]. Damage to big blood vessels in the kidney, leg, and brain is a macrovascular issue. The early detection of diabetes can avoid many complications. The increased prevalence of diabetes is driving major factors as rapid urbanization, unhealthy diets, and change of lifestyle [5]. Many health care industries have gathered a huge data quantity which consist of records of hospital, results of medical examination, and patient's medical reports. Through the doctor's experience and knowledge, early detection of diabetes can be done. But sometimes, the prediction may be inaccurate and susceptible. Hence, it may lead to improper decision-making.

The hidden data pattern may be unrecognized, and it can also cause an effect on the decision-making process. As a result, patients become deprived from proper treatment. The early detection of diabetes is highly required with better efficiency through the automated computation domain [5]. Among several machine learning algorithms, ANN is mainly used in the medical field. ANN is used for the early detection of diabetes [6]. ANN is chosen to predict diabetes based on the dataset chosen from the PIDD. The network of neural is trained and analysed to predict the capabilities of the neural network based on the unseen data.

2 Literature Survey

Several researches are done on machine learning methods to predict diabetes using Pima Indian Diabetes Dataset (PIDD). This dataset contains 768 records of 9 attributes describing female patients. Suyash srivastava et al. [6] showed an accuracy of 92% through an ANN technique on PIDD. Nesreen Samer et al. [7] implemented NN network in just neural network (JNN) environment. The dataset was collected from the association of diabetic's of Urima which contains 1004 samples of 9 attributes with an accuracy of 87.3% with an average error of 0.0010.

Venkatesan et al. [8] proposed that radial basis function (RBF) is suitable compared with the multilayer perceptron network and the classical regression. Hussam et al [9] proposed an ANN model to predict diabetes which was 98.73% effective in predicting diabetes. Jobeda et al. [10] proposed research with the various ML algorithms for the diabetes prediction on PIDD. They performed the comparison among various algorithms and concluded that NN with two hidden layers provided an accuracy of 88.6%.E. Guldoganthe et al. [11] showed the comparison of multilayer perceptron (MLP) and RBF and suggested that the MLP model has better performance than RBF with the use of PIDD. Detection of diabetes can be done by invasive and non-invasive methods [12]. Karunakaran et al. [13] implemented adoptive functioning classifier and learning techniques are implemented to predict the output.

The observation made from the literature survey is that the various algorithms were implemented for ANN, where the accuracy of training the network was obtained in the range of 88–92%. To increase the accuracy, the present study is proposed along with machine learning methods. The major issues depend on features selection and classifier. In our study, Pearson's correlation method is used to find the logical features. The work is presented for the determination of the individual as diabetic or non-diabetic. The neural network with different neurons at different epochs has been done, and later, the accuracy is calculated.

3 Methodologies

3.1 Sets of Data and Software Tool

The PIDD is taken from the UC Irvine machine repository for our proposed work. The PIDD dataset comprises information on female patients who are at least 21 years old. About 768 patients are represented in the databases, each having their own set of attributes. The nine attributes are mentioned with the detailed description in Table1. The last attribute is used to predict the person as a diabetic or non-diabetic. The prediction of diabetic is represented by the binary values as '1' for diabetes and '0' for non-diabetic. In our work, data mining is done to improve the accuracy of the prediction.

Table 1 Attributes for diabetes data prediction

Attribute	Description	Type
Pregnancy	Number of times pregnant	Numeric
Glucose	Plasma glucose concentration 2 h in an oral glucose	Numeric
BP	Diastolic blood pressure	Numeric
Skin thickness	Triceps skinfold thickness	Numeric
Insulin	2-h serum insulin	Numeric
BMI	Body mass index	Numeric
DPF	Diabetic pedigree function	Numeric
Age	Age in years	Numeric
Class	Diabetic results diagnose	Nominal

Weka 3-8-5 is an open-source machine learning and data mining tool for analysing the performance of a dataset. The data preparation is done with the Weka tool. And, nntool is used to implement the neural network in MATLAB 2021a.

3.2 Data Preprocessing

Preprocessing of data helps to obtain the model with a better accuracy. Quality of data selection can be improved by the various functions such as replace missing values, outlier rejection, and normalization of data, and feature selection in order to improve the quality of data. The PIMA datasets were divided into two categories: diabetic samples (268) and non-diabetic samples (500).

3.2.1 Identification of Missing Value

Dataset is fed to the Weka tool, with the option of replacing missing values. The missing values are replaced with the corresponding mean value. Table 2 shows the number of values missing in the Pima dataset.

Table 2 Identified values in the dataset

Attribute	Missing values identified
Pregnancy	0
Glucose	5
BP	35
Skin thickness	227
Insulin	374
BMI	11
DPF	0
Age	0

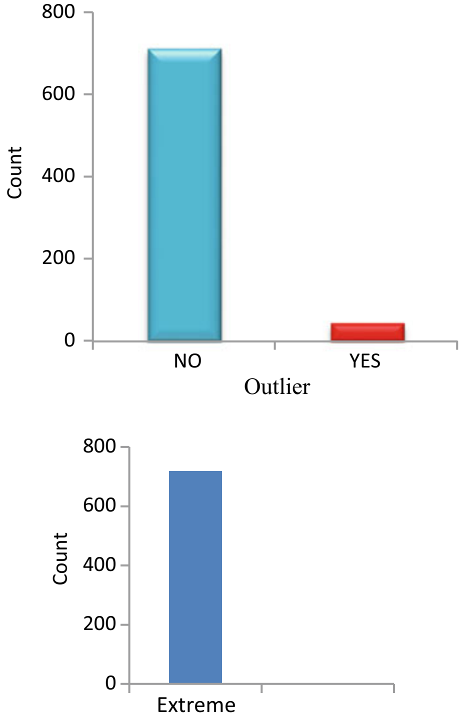


Fig. 1 Outlier and extreme values

3.2.2 Identification and Removal of Outliers

The interquartile range is used to identify the extreme and outlier values from the dataset. The number of extreme values was 0 and outliers were 45, and it is represented as shown in Fig. 1. There were 719 instances after the removal of outliers values from the datasets.

3.2.3 Selection of Feature

In Weka, the correlation co-efficient is found by the correlation filter. The results obtained are represented in Table 3. The value (0.2) is used as a cut-off for relevant attributes. Therefore, the attributes, namely BP, skin thickness, and DPF are removed. The remaining five attributes, namely glucose, BMI, insulin, pregnancy, and age are considered as important.

Table 3 Input and output attributes correlation

Attribute	Missing values identified
Pregnancy	0.718
Glucose	0.4866
Insulin	0.3617
Age	0.2532
BP	0.1609
Skin thickness	0.1009
DPF	0.0479

3.2.4 Normalization

The feature selection is done by normalizing the data from range 0–1, so that the network can be trained well to obtain better accuracy. Table 4 shows the means and standard deviations. After the completion of preprocessing, the total samples were obtained as 716. The 214 patients were diabetic, and the rest were non-diabetic.

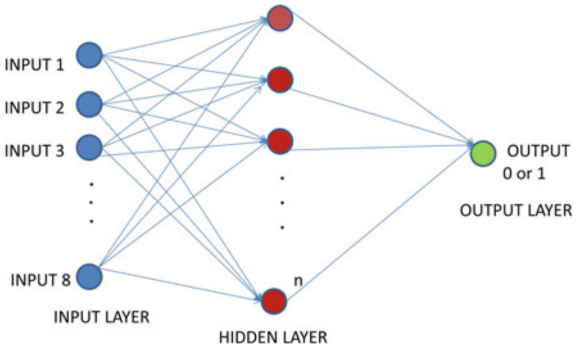
3.3 Iteration of Artificial Neural Network Model

The neural network model was built with different neurons (10,100,250) as shown in Fig. 2 of two hidden layers, and the obtained results are shown in table with the comparison. In NN, the activation function at the layer of input is processed by the

Table 4 Mean and standard deviation of attributes

Attribute	Mean Standard deviation	Standard deviation
Pregnancy	0.229	0.198
Glucose	0.604	0.16
BP	0.494	0.125
Skin thickness	0.214	0.157
Age	0.1609	0.197

Fig. 2 Representation of ANN model



weighted sum of input. The activation functions used at the hidden layers are transig and purelin.

In ANN, the output error is reduced by the optimizer at the time of back propagation method. The optimizer used is stochastic gradient descent. The weight adjustment is initialized by the learning rate with respect to the loss gradient. The 80% of dataset was taken to train the model, and 20% of dataset was chosen to test the model.

3.3.1 Iteration of Neural Network with One Hidden Layer

The hidden layer was placed between the input and output. The five attributes are given with five neurons at the input. Hundred neurons are placed at the hidden layer. The activation function used was ReLU. A sigmoid is used as the activation at the output layer with the learning rate of 0.01.

3.3.2 Iteration of Neural Network with One Hidden Layer

The 250 neurons are fed to the hidden layer which is placed between the input and output. The activation function used was ReLU. The sigmoid is used as the activation at the output. The learning rate used was 0.001. The regression plot obtained after training the ANN model is as shown in Fig. 3.

3.3.3 Iteration of Neural Network with Two Hidden Layers

The two hidden layers were placed between the input and output. The 25 neurons are placed at the first hidden layer and 6 neurons at the second layer with the learning rate of 0.01.

Impact of accuracy with different neurons placed at the hidden layer.

Hidden layer	Number of neurons accuracy prediction status	Accuracy	Prediction status
1	100	77	0
1	250	99.23	1
2	First layer-25 Second layer-10	88.3	0

The status of diabetic prediction is said that the person is diabetic(1) or not diabetic(0). This result is only based on the accuracy level of the tested output of ANN model. If the accuracy of the ANN is obtained in the range of 70–90%, the prediction of diabetes may not be accurate always which will be always pretending to be correct. If the obtained accuracy is between 90 and 100% and the prediction exactly meant, it is a diabetic.

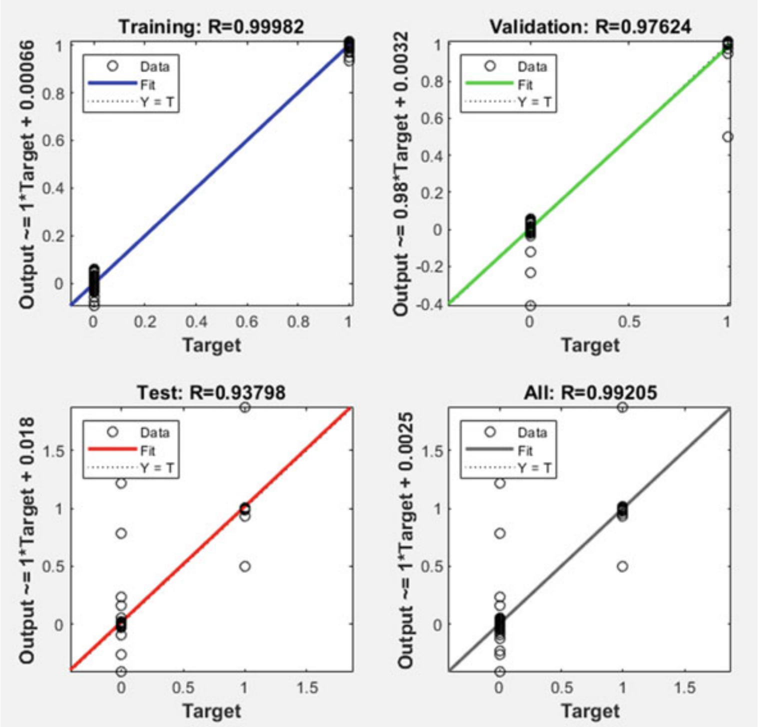


Fig. 3 Regression plot of 250 neurons

4 Conclusion

The detection of diabetes at the early stage is very significant. An ANN system has been designed which is used to detect diabetes with a high accuracy. The data is been preprocessed with the help of Weka tool. Then, the dataset with five attributes is trained with the neural network of one hidden layer and two hidden layers. After training and testing the network, the best accuracy was obtained with 250 neurons placed at the hidden layer of 99.20% for the PIDD.

References

1. ICMR guidelines for management of type 2 diabetes
2. https://www.medicalnewstoday.com
3. https://www.who.int/health-topics/diabetes
4. https://www.diabetes.co.uk/diabetes_care/blood-sugar-level-ranges.html
5. https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/diabetes-long-terms-effects

6. S. Srivastava, L. Sharma, V. Sharma, A. Kumar, H. Darbari, Prediction of diabetes using artificial neural network approach: ICoEVCI (2018, India). https://doi.org/10.1007/978-981-13-1642-5_59
7. N.S. El-Jerjawi, S.S. Abu-Naser, Diabetes prediction using artificial neural network. Int. J. Adv. Sci. Technol. **121**, 55–64 (2018)
8. P.V. Anitha, Application of a radial basis function neural Application of a radial basis function neural. Curr. Sci. **91**(9), 10 (2018)
9. H.H. Harz, A.O. Rafi, M.O. Hijazi, S.S. Abu-Naser, Artificial neural network for predicting diabetes using JNN. Int. J. Acad. Eng. Res. (IJAER) **4**(10), 14–22 (Oct 2020) ISSN: 2643-9085
10. J. Jobeda Jamal Khanam, S.Y. Foo, A comparison of machine learning algorithms for diabetes prediction. ICT Express (2021) ISSN 2405-9595. https://doi.org/10.1016/j.icte.2021.02.004
11. D.Ö.Ü.M.O. Kaya, Computer-aided model for the classification of acute inflammations via radial-based function artificial neural network. J. Cogn. Sys. **6**(1), 1–4 (2021). https://doi.org/10.52876/jcs.913730
12. V. Agalya, S. Sumathi, An assessment of pain-free blood glucose level by noninvasive methods. Int. J. Current Res. Rev. **13**(05), 32–35 (March 2021)
13. P. Karunakaran, Y.B. Hamdan, Early prediction of Autism spectrum disorder by computational approaches to fMRI analysis with early learning b technique. J. Artif. Intell. **2**(04), 207–216 (2020)

Author Index

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*,
Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6