Lecture Notes in Networks and Systems 435

Vishal Goyal
Manish Gupta
Seyedali Mirjalili
Aditya Trivedi *Editors*

Proceedings of International Conference on Communication and Artificial Intelligence

ICCAI 2021

Springer

Lecture Notes in Networks and Systems

Volume 435

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.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

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

Vishal Goyal · Manish Gupta · Seyedali Mirjalili ·
Aditya Trivedi
Editors

Proceedings of International Conference on Communication and Artificial Intelligence

ICCAI 2021

Springer

Editors
Vishal Goyal
Department of Electronics
and Communication Engineering
GLA University
Mathura, India

Manish Gupta
Department of Electronics
and Communication Engineering
GLA University
Mathura, India

Seyedali Mirjalili
Torrens University Australia
Adelaide, SA, Australia

Aditya Trivedi
Atal Bihari Vajpayee-Indian Institute
of Information Technology
and Management
Gwalior, Madhya Pradesh, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-19-0975-7 ISBN 978-981-19-0976-4 (eBook)
https://doi.org/10.1007/978-981-19-0976-4

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

Preface

The 2nd International Conference on Communication and Artificial Intelligence (ICCAI-2021) was organized to address various issues to flourish the creation of intelligent solutions in the future. It is a multidisciplinary conference organized with the objective of bringing researchers, developers, and practitioners together from academia and industry with interest in advanced start of the art in communication and artificial intelligence. The theme is whole heartedly concerned with innovating and inspiring the researchers adopt the implementation outcomes.

Technological developments all over the world are dependent upon globalization of various research activities. Exchange of information and innovative ideas is necessary to accelerate the development of technology. Keeping this ideology in preference, the 2nd International Conference on Communication and Artificial Intelligence (ICCAI-2021) was organized at GLA University, Mathura, Uttar Pradesh, India, on November 19–20, 2021. During the time of COVID-19 pandemic, it was very beneficial to organize the conference on virtual platform. The conference displayed the collaboration of existing and upcoming researches for a befitting use in the future. It was an exceptional platform to the researchers to meet and discuss the solutions, scientific results, and methods in solving intriguing problems with folks who are actively working in the evergreen fields.

The International Conference on Communication and Artificial Intelligence has been organized with a foreseen objective of enhancing the research activities at a large scale. Technical Program Committee and Advisory Board of ICCAI-2021 include eminent academicians, researchers, and practitioners from abroad as well as from all over the nation.

In ICCAI-2021 proceedings, selected manuscripts are subdivided into six tracks, namely communication system, control system and IoT, machine learning, image and signal processing, soft and cloud computing, and VLSI design and its applications. ICCAI-2021 received more than 197 submissions. A sincere effort has been made to make it an immense source of knowledge by including 58 manuscripts in this proceedings volume. The selected manuscripts have gone through a rigorous review process and are revised by authors after incorporating the recommendations of the reviewers.

The conference was highly successful. The presented papers maintained high promise suggested by the written abstracts, and the program was chaired in a professional and efficient way by the session chairmen who were selected for their international standing in the subject. The number of delegates was also highly gratifying, showing the high level of international interest in the scopes of the conference.

We are thankful to our chief guests Prof. Deepak Das and Prof. Vikram Sarabhai, Distinguished Professor, Ex. Director, SAC/ISRO/DOS, and our guests of honor Shri Prasanth Perugupalli, CPO, nference, India, and Shri Raghubansh Gupta, Senior Manager, nference, India, for enlightening the participants with their knowledge and insights. We are thankful to our keynote speakers/session chair Prof. Sabah Mohammed, Lakehead University, Ontario, Canada; Prof. Nischal Verma from IIT Kanpur; Prof. Manav Bhatnagar, IIT Delhi; Prof. J. C. Bansal from SAU, New Delhi; Dr. Himanshu Bhushan Baskey, Scientist, DRDO, DMSRDE, Kanpur; Prof. Ahmed Esmail Shalan from Central Metallurgical Research and Development Institute, Egypt; Dr. Asit Tiwari from Harman International, Germany; Dr. Akhil Ranjan from NTU, Singapore; Dr. Vilas Gaidhane from BITS-Pilani, Dubai campus, Dubai; Dr. T. C. Shami, Ex. Scientist, DRDO; Dr. Mukesh Saraswat from JIIT, Noida; Dr. Mayank Srivastava from NIT Jamshedpur, Dr. Sangeeta Singh from NIT Patna, Dr. Mahesh Soni from Lancaster University, Lancaster; Shri Vivek Garg from Valeo Schalter und Sensoren GmbH; and Dr. Ravi Prakash Dwivedi from VIT University, Chennai, for giving their time and showering the participants with immense knowledge. The organizers wish to thank Dr. Aninda Bose, Senior Editor, Springer Nature New Delhi, India, for their support and guidance.

We are also thankful for delegates and the authors for their participation and their interest in ICCAI-2021 as a platform to share their ideas and innovative research work. In addition, we extend our heartfelt gratitude to the reviewers and technical program committee members for showing their concern and efforts in the entire process. We are indeed thankful to everyone directly or indirectly associated with the conference organizing team for leading it toward success.

Mathura, India

Prof. Vishal Goyal
Dr. Manish Gupta
Prof. Seyedali Mirjalili
Prof. Aditya Trivedi

Contents

Editors and Contributors

About the Editors

Dr. Vishal Goyal is currently working as Professor, Department of Electronics and Communication Engineering, GLA University, Mathura. He has a very rich experience of more than 20 years in teaching undergraduate and postgraduate classes. He has published more than 50 research papers publications in different International Journals and in Proceedings of International Conferences of repute. He has organized 4 international conferences technically sponsored by IEEE. He has also worked as Member of organizing committee in several IEEE international conferences in India and abroad. He has delivered numerous invited and plenary conference presentations and seminars throughout country and chaired the technical sessions in international and national conferences in India. He is also member of reviewer board for *IEEE Transactions on Neural Networks and Learning Systems*, *International Journal of Swarm Intelligence*, *International Journal of Measurement Technologies and Instrumentation Engineering*, *Transactions of the Institute of Measurement and Control*, *Journal of the Chinese Institute of Engineers*, *International Journal of Power Electronics and Drive Systems*. He has been appointed member of board of studies as well as in syllabus committee of different private Indian universities and member of organizing committee for various national and international seminars/workshops. He is Executive Committee Member of IEEE Computational Intelligence Society, IIT Kanpur. He is an active senior member of IEEE, Life member of ISTE, Member IFAC (International Federation of Automatic Control), Life Member Indian Science Congress Association.

Dr. Manish Gupta is presently associated with Department of Electronics and Communication Engineering, Institute of Engineering and Technology, GLA University, Mathura (UP) India. He received his Bachelor of Engineering (B.E.) in Electronics Engineering from Jiwaji University, Gwalior (MP) India, in 2000. He did his M. Tech. degree from Uttar Pradesh Technical University, Lucknow (UP), India, in year 2006 and Ph.D. from Rajasthan Technical University, Kota (Raj.). He has a very rich experience of more than 20+ years in teaching undergraduate and post-graduate classes. He has published 40+ research papers publications in different International Journals and in Proceedings of International Conferences of repute. He had served as the member of organizing committees for various international conferences. Moreover, he was also designated the reviewers various refereed journals. He is a life member of ISTE, Member IET. His research interest includes Communication systems, Image and Signal Processing.

Prof. Seyedali Mirjalili is the director of the Centre for Artificial Intelligence Research and Optimization at Torrens University Australia. He is internationally recognized for his advances in Swarm Intelligence and Robust Optimization. He has published over 200 publications with over 30,000 citations and an H-index of 50. As the most cited researcher in robust optimization, he is in the list of 1% highly- cited researchers and named as one of the most influential researchers in the world by Web of Science. He is a senior member of IEEE and an associate editor of several journals including *Neurocomputing*, *Applied Soft Computing*, *Advances in Engineering Software*, *Applied Intelligence*, and *IEEE Access*. His research interests include robust optimization, engineering optimization, multi-objective optimization, swarm intelligence, evolutionary algorithms, and artificial neural networks. He is working on the application of multi-objective and robust meta-heuristic optimization techniques as well.

Prof. Aditya Trivedi is Professor in the Information and Communication Technology (ICT) Department at ABV-Indian Institute of Information Technology and Management, Gwalior, India. He has about 20 years of teaching experience. Before joining ABV-IIITM in December 2006, he was Reader in the Department of Electronics and CSE at M.I.T.S., Gwalior. He received his bachelor degree in Electronics Engineering from the Jiwaji University. He did his M.Tech. (Communication Systems) from Indian Institute of Technology (IIT), Kanpur. He obtained his doctorate (Ph.D.) from IIT Roorkee in the area of Wireless Communication Engineering. His teaching and research interest include digital communication, CDMA systems, Signal processing, and networking. He has published around 60 papers in various national and international journals/conferences. He is a fellow of the Institution of Electronics and Telecommunication Engineers (IETE). He has completed one AICTE project. He is a reviewer of IEEE and Springer journals. In 2007, he was given the IETEs K. S. Krishnan Memorial Award for best system-oriented paper.

Contributors

Chirag Agrawal PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, Madhya Pradesh, India

Mohammed Nadir Bin Ali Daffodil International University, Dhaka, Bangladesh

Zulfikar Alom Department of Computer Science, Asian University for Women, Chattogram, Bangladesh

R. Angeline SRM Institute of Science and Technology, Ramapuram, India

G. Arockia Selva Saroja Department Electronics and Communication Engineering, Noorul Islam Centre for Higher Education, Tamil Nadu, India

J. Arun Pandian Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India

B. A. Ashwin SRM Institute of Science and Technology, Ramapuram, India

S. V. Aswin Kumer Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

Zeyar Aung Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates

Mohammad Abdul Azim Department of Computer Science, Asian University for Women, Chattogram, Bangladesh

Abhishek Badholia Department of Computer Science and Engineering, School of Engineering and Information Technology, MATS University, Raipur, India

Yasmeen Bano School of Management Studies, Sathyabama Institute of Science and Technology (SIST), Chennai, India

Atul Bansal Department of Electronics and Communication Engineering, GLA University, Mathura, India

Antara Titikhsha Barua Department of Computer Science, Asian University for Women, Chattogram, Bangladesh

S. Bharathi Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

Sitadevi Bharatula Department of ECE, School of Engineering, Amrita Vishwa Vidyapeetham, Chennai, India

Surendra Bhosale Department of EE, Veermata Jijabai Technological Institute, Mumbai, India

Touhid Bhuiyan Daffodil International University, Dhaka, Bangladesh

Rahul Bibave Department of Electrical Engineering, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune, India

Amit Kr. Chaturvedi CSE Department, Government Engineering College, Ajmer, India

Amit Kumar Chaturvedi CSE Department, Government Engineering College, Ajmer, India

K. Chethan Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysuru, India

Omdev Dahiya School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

Sandeep Dalal Department of Computer Science and Applications, Maharshi Dayanand University, Rohtak, India

Priyanka Darbari Department of CEA, GLA University, Mathura, UP, India

Bhushan Deore Department of EE, Veermata Jijabai Technological Institute, Mumbai, India

Arpan Desai Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India

Sharmishta Desai Maharashtra Institute of Technology, World Peace University, Pune, India

Smita Deshmukh Terna Engineering College (TEC), Mumbai University, Navi Mumbai, India

Somesh Dhamija GLA University, Mathura, UP, India

M. K. Dhananjaya Computer Science Department, Bhagwant University, Ajmer, India

Amita Dhankhar Department of Computer Science and Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, India

Vishesh Dharaiya Department of Materials Science and Metallurgical Engineering, Indian Institute of Technology, Hyderabad, Telangana, India

T. Dhikhi SIMATS, Chennai, India

P. Durgadevi Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

V. Joseph Emmanuvel Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

Mohit Gangwar Department of IT, IET, Dr. Rammanohar Lohia Avadh University, Ayodhya, Uttar Pradesh, India

Anjana Goel Department of Biotechnology, GLA University, Mathura, U.P, India

M. E. Gokul SRM Institute of Science and Technology, Ramapuram, India

Vishal Goyal Department of Electronics and Communication Engineering, GLA University, Mathura, India

Devbrat Gupta Department of Electronics and Communication Engineering, GLA University, Mathura, India

Manish Gupta GLA University, Mathura, India

R. K. Gupta Madhav Institute of Technology and Science, Gwalior, M.P, India

Mohammed Shamsul Hoque Daffodil International University, Dhaka, Bangladesh

Mohamed Emran Hossain Daffodil International University, Dhaka, Bangladesh

Shaik Jakeer Hussain School of Electronics, VFSTR, Guntur, India

Dinh Tran Ngoc Huy MBA, Banking University HCMC, Ho Chi Minh city, Vietnam;
International University of Japan, Niigata, Japan

Mohd Kamran Ikram Department of Electrical Engineering, Aligarh Muslim University, Aligarh, UP, India

Nisha Joseph Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu, India

Dheeraj Kalra GLA University, Mathura, India

K. Kanchanadevi Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India

J. Karthikeyan Vellore Institute of Technology, Vellore, India

M. Kasiselvanathan Department of Electronics and Communication Engineering, Sri Ramakrishna Engineering College, Coimbatore, India

Rahul Katarya Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

Gurjit Kaur Delhi Technological University, Delhi, India

A. R. Kavitha Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

Utkal Khandelwal GLA University, Mathura, India

D. R. Kirubaharan Department of Mathematics, PRIST Deemed to be University, Thanjavur, India

Sunanda Kulshrestha Department of Biotechnology, GLA University, Mathura, U.P, India

Anusha Kumar Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, Pilani, India

Devendra Kumar Department of Electronics and Communication, GLA University, Mathura, India

Divesh Kumar Department of ECE, GLA University, Mathura, U.P., India

Jitendra Kumar Department of Electronics and Communication Engineering, GLA University, Mathura, India

Manish Kumar GLA University, Mathura, India

Manoj Kumar Department of CEA, GLA University, Mathura, UP, India

Smit Kumbhani Department of Computer Science and Engineering, Marwadi Education Foundation's Group of Institutions, Rajkot, Gujarat, India

Brijesh Kundaliya Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India

J. Lakshmi Prasanna Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

Payal Mahajan Department of Electrical Engineering, Jamia Millia Islamia, New Delhi, India

Ronitt Mehra Delhi Technological University, Delhi, India

Aman Mishra GLA University, Mathura, UP, India

Puneet Mishra Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, Pilani, India

Yatin Mishra GLA University, Mathura, UP, India

Divya Mohan Department of Computer Science and Engineering, Satyabhama University, Chennai, Tamil Nadu, India

Dhanshree More Department of EE, Veermata Jijabai Technological Institute, Mumbai, India

B. S. Murthy Department of Telecommunications, Ministry of Communications, New Delhi, India

D. Murugan Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu, India

Mohd Mustaqeem Department of Computer Science, Aligarh Muslim University, Aligarh, UP, India

Mohammad Nadeem Department of Computer Science, AMU, Aligarh, UP, India

Thi Hang Nguyen Thai Nguyen University, University of Information and Communication Technology, Thai Nguyen, Vietnam

Samarth Nigam Delhi Technological University, Delhi, India

Rajat Pandey Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India

Killol Pandya Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India

Aneesya Panicker GLA University, Mathura, India

Chinmaya Panigrahy Thapar Institute of Engineering and Technology, Patiala, Punjab, India

Nishtha Parashar GLA University, Mathura, UP, India;
Madhav Institute of Technology and Science, Gwalior, M.P, India

Vaibhav A. Parjane Department of Computer Science and Engineering, Bhabha University, Bhopal, Madhya Pradesh, India

Upesh Patel Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India

Ambika Patidar Maharashtra Institute of Technology, World Peace University, Pune, India

Rina S. Patil Department of Computer Science and Engineering, Bhabha University, Bhopal, India

Anu Pius Department of Mathematics, PRIST Deemed to be University, Thanjavur, India

J. Prasad Department of Electronics and Communication Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

Gudapati Ramyasri School of Electronics, VFSTR, Guntur, India

M. Ravi Kumar Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

Rahul Rishi Department of Computer Science and Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, India

V. Sahaya Sakila Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

R. Sandeep Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysuru, India

M. Sangeetha Lakshmi Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

Chella Santhosh Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

S. Saranya Computer Science and Engineering, K.Ramakrishnan College of Engineering, Tiruchirappalli, India

M. S. Saravanan Saveetha School of Engineering, SIMATS, Chennai, India

Mohammad Sarfraz Department of Electrical Engineering, Aligarh Muslim University, Aligarh, UP, India

Paresh Chandra Sau GLA University, Mathura, India

G. Sekar Department of Electronics and Communication Engineering, Sri Ramakrishna Institute of Technology, Coimbatore, India

Karama M. K. Senan Department of Electrical Engineering, Aligarh Muslim University, Aligarh, UP, India

M. S. Shamanth Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysuru, India

Annu Sharma Computer Science Department, Bhagwant University, Ajmer, India

Avnish Sharma GLA University, Mathura, India

Gaurav Kumar Sharma GLA University, Mathura, India

Kalpana Sharma Computer Science Department, Bhagwant University, Ajmer, India

Neha Sharma Department of Electronics and Communication, Ujjain Engineering College, Ujjain, M.P, India

Rohit Shrivastava Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpat, Chennai, Tamil Nadu, India

Aasheesh Shukla Department of Electronics and Communication Engineering, GLA University, Mathura, India

Janpreet Singh School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

Mayank P. Singh Department of Electrical Engineering, GLA University, Mathura, India

Rashmi Singh Department of Management, GLA University, Mathura, India

Shreyaskar Pratap Singh GLA University, Mathura, Uttar Pradesh, India

Thakurendra Singh Department of ECE, GLA University Mathura, Mathura, India

Kamna Solanki Department of Computer Science and Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, India

Amrita Soni Department of Electronics and Communication, Ujjain Engineering College, Ujjain, M.P, India

Anand Kumar Soni Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpat, Chennai, Tamil Nadu, India

S. P. Sri Krishna SRM Institute of Science and Technology, Ramapuram, India

P. Srilaxmi Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

Khusboo Srivastava GLA University, Mathura, UP, India

Pallavi Srivastava Department of Electronics and Communication Engineering, GLA University, Mathura, India

Preeti Tarkar Department of Management, GLA University, Mathura, India

Sharika Tasnim Department of Computer Science, Asian University for Women, Chattogram, Bangladesh

E. J. Thomson Fredik Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

Akhilesh Tiwari Madhav Institute of Technology and Science, Gwalior, M.P, India

Divya Tiwari Terna Engineering College (TEC), Mumbai University, Navi Mumbai, India

V. K. Tomar Department of ECE, GLA University Mathura, Mathura, India

Ajay Kant Upadhyay Dayalbagh Educational Institute, Agra, India

Anju Upadhyay Department of Electrical Engineering, GLA University, Mathura, U.P, India

Trushit Upadhyaya Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India

Pham Van Tuan National Economics University, Ha Noi, Vietnam

Prathibha Varghese Electronics and Communication Department, Noorul Islam Centre for Higher Education, Tamil Nadu, India

S. Vasantha School of Management Studies, Technology and Advanced Studies (VISTAS), Vels Institute of Science, Chennai, India

K. Vengatesan Department of Computer Science and Engineering, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune, India

P. Vigneshwaran Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpat, Chennai, Tamil Nadu, India

K. P. Vishal Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysuru, India

P. William Department of Information Technology, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune, India

Amit Yadav Department of Electronics and Communication, RBSETC, Agra, India

Indresh Yadav Department of Electrical Engineering, GLA University, Mathura, U.P, India

Madhuri Yadav Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

Sujit Kumar Yadav GLA University, Mathura, Uttar Pradesh, India

S. Yuvaraj Department of Electronics and Communications Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India

Adnan Zafar Department of Computer Science, AMU, Aligarh, UP, India

Zaheeruddin Department of Electrical Engineering, Jamia Millia Islamia, New Delhi, India

Communication System

3 × 3 Split-Ring Resonators Array-Inspired Defected Ground Plane Antenna for 2.4/5.5 GHz Wireless LAN Applications

Upesh Patel, Trushit Upadhyaya, Arpan Desai, Rajat Pandey, Killol Pandya, and Brijesh Kundaliya

Abstract Homogeneous split-ring resonator array-inspired antenna for the Wireless LAN (WLAN) is presented in this article. The overall dimension for the designed antenna is 0.40λ mm^2 ×0.32λ mm^2 (at lower operating frequency). This antenna consists of the defected ground plane structure for the gain and bandwidth improvement. 3 × 3 split-ring resonator array is formed as defected ground plane. For the fabrication purpose, cost-effective and standard thickness of 1.6 mm, FR-4 substrate is used. The resonance frequencies are at 2.4 and 5.5 GHz with the values of gain are 1.06 and 2.14 dBi, respectively. The impedance bandwidths for both the frequencies are 7.5% (2.31–2.49 GHz) and 21.8% (4.9–6.1 GHz). The omnidirectional radiation patterns, overall gain, and efficiency of around 79% make the antenna most commercially suitable for the WLAN devices.

Keywords Array structure · Split-ring resonators · Defected ground plane · WLAN

1 Introduction

The antenna, generally used as a transducer is an essential component of wireless communication technology. The recent decades have been recognized as increasing demand for major wireless technologies like; laptops, palmtops, mobile phones, and other portable devices. In all such devices, the antenna is a vital element from the communication point of view and this antenna must have good performance parameters in order to survive in communication technology market. Nowadays, the most widely used Wireless LAN (WLAN) protocols are IEEE 802.11b/g, which utilizes the operating frequency of 2.4 GHz and IEEE 802.11a which utilizes the 5.5 GHz ISM band (5.15–5.82 GHz). The antenna should have the ability to perform in both the frequency bands for the better feasibility and flexibility[1].

U. Patel (✉) · T. Upadhyaya · A. Desai · R. Pandey · K. Pandya · B. Kundaliya
Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India
e-mail: upeshpatel.ec@charusat.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_1

The ordinary antenna has its typical gain and bandwidth requirement for optimal performance along with low manufacturing costs. A traditional microstrip patch antenna suffers from issues such as single resonance frequency, lower gain, lower impedance bandwidth, and linear polarization [2]. There are numerous techniques available for radiation characteristics enhancement of antenna. Out of several well-known methods, defective ground plane technique has gained significant popularity due to its structural design and ease in fabrication [3–5]. DGS offers extra freedom of tuning and opens up additional dimensions in terms of variation of characteristics impedance of antenna. Improvement in the antenna characteristics in terms of bandwidth and gain values is highly required. There are multiple techniques available in literature; however, Defected Ground Structure (DGS) is most widely used in case of array antenna for the coupling issue and reduction of cross-polarization. The defected ground structures are either unit cells or strategically cascaded periodic and/or non-periodic structures etched out in ground plane or directly placed under the transmission line [6, 7]. However, as per the advanced research, placement of defected structure near the feed line lead to a better output. The ground plane is engineered by placing symmetric or asymmetric slits and cuts to make it defected ground plane. This perturbs the surface currents flowing in antenna due to introduction of slit resistance, capacitance, and inductance [8–10]. Due to this change, the overall equivalent circuit parameters and hence antenna resonance performance has altered. In addition, effective parameter extraction and fabrication is easier compared to photonic band gaps (PBG) and Electromagnetic band gaps (EBG) inspired antennas [11]. For the miniaturization of antenna, along with the DGS, split-ring resonators based designs are also shown a major concern for active and passive microwave devices and components [12, 13].

Split-ring resonator (SRR) [14] is the most widely used mu-negative material in the design of antenna and other microwave components. A unit cell of an SRR structure consists of a metallic ring with a slot at opposite sides in the center. Using a continuous electromagnetic wave the SRR is excited and electric charge is induced in two rings. The gaps on both opposite sides prevent the current flow around the rings. Hence the total capacitance and inductance effect is equivalent to LC circuit. This phenomenon is used in the gain and bandwidth improvement. SRR structure is formed on ground plane which exhibits the capacitance in between the gap of metallic rings and which further helps for the bandwidth improvement at resonance.

The organization of the proposed article is as follows. The design geometry along with the dimension configuration of the designed antenna is presented in Sect. 2. The software simulated results and hardware prototype measured results of the antenna are discussed in the next section. In Sect. 4, the overall conclusion of the article is illustrated followed by the state of art articles used for reference.

2 Design Configuration of Antenna

Figure 1a, b depict the top view and ground plane, respectively. In Fig. 1b, the unit cell element of split-ring resonators is shown with the detailed dimensions. The array of the SRRs is formed on the ground to make ground plane a defected ground plane. Antenna substrate has dimensions of 50 $mm^2 \times 40\ mm^2$. Low-cost FR4 substrate having loss tangent 0.02 and relative dielectric constant 4.4 has been utilized for antenna design. Microstrip patch is of dimension 40 mm × 30 mm is printed on the top of substrate. The antenna is fed through microstrip line of 50Ω impedance and a low loss SMA connecter is used for the feeding at antenna. The physical dimensions

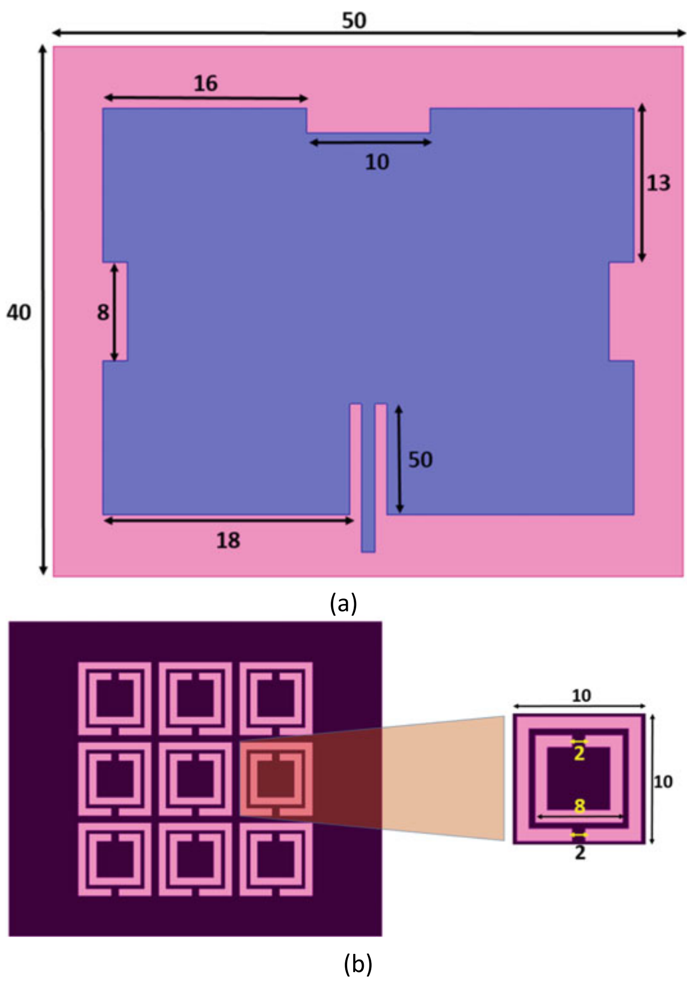


Fig. 1 Design configuration of antenna **a** antenna top view, **b** ground plane with SRR unit cell

of proposed antenna are shown in the below figure (All mentioned dimensions are in mm.).

3 Results and Discussions

Unit cell structure and its analysis are shown in Figs. 2 and 3, respectively. A unit cell of negative refractive index, square split-ring resonator, structure was created inside vacuum filled box in software simulator. Two ports were assigned for extraction of transmission and reflection coefficients. Perfect H boundaries were applied to the SRR facing surfaces of unit cell for generating magnetic dipole moment in split-ring resonator.

The return loss of the implemented configuration is shown in Error! Reference source not found.. Continuous red line represents S11 in dB and blue line represents S21 in dB It shows that model is resonating in the lower range of frequency bands.

Ansys High-Frequency Software Simulation (HFSS) is used for the design and simulation of the proposed antenna. The proposed antenna was fabricated using the

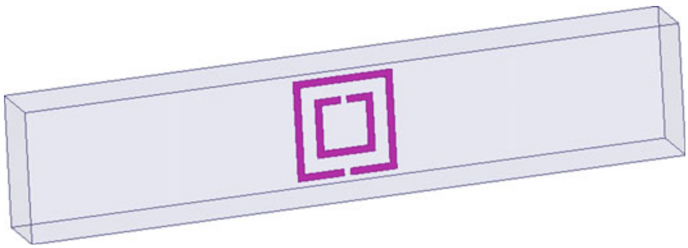


Fig. 2 Square SRR excited in plane wave polarization

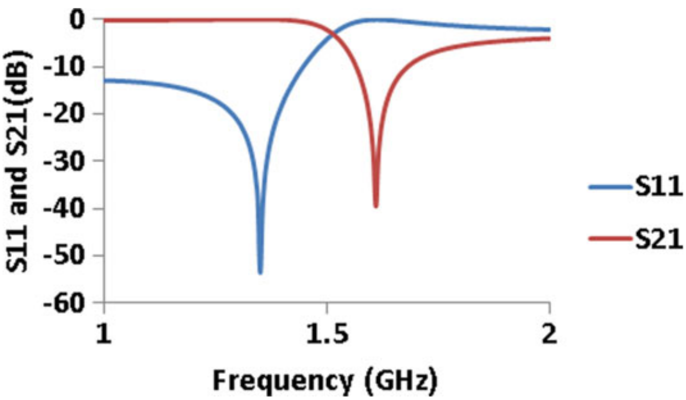


Fig. 3 Return loss of square SRR unit cell

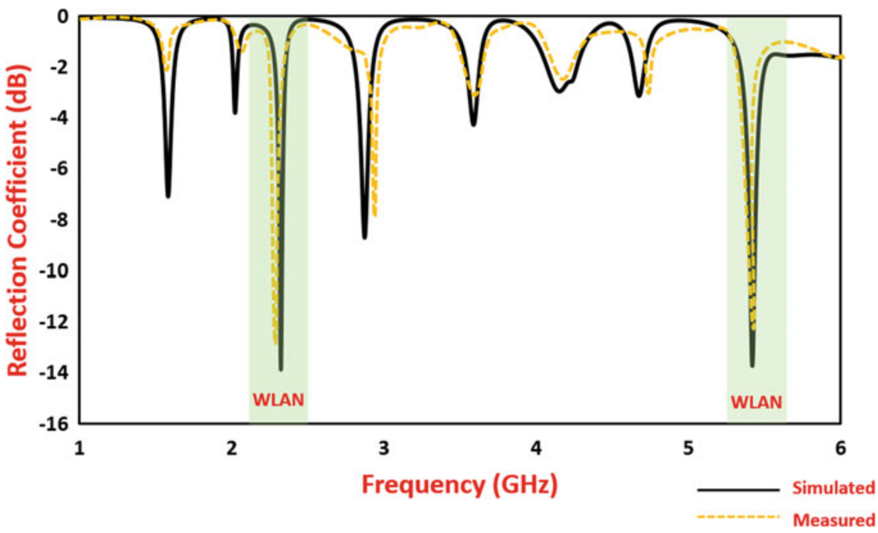


Fig. 4 Reflection coefficient (S_{11}) (simulated and measured)

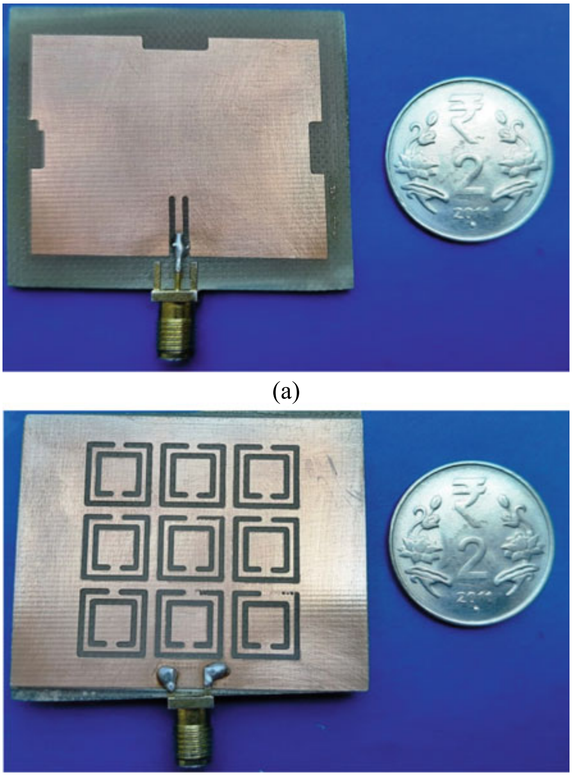
standard PCB fabrication process. Keysight 9912A vector network analyzer was used for the measurement of the return loss of the designed prototype. The return loss graph for the proposed antenna is depicted in Fig. 4.

The measured and software simulated reflection coefficient values are satisfactorily well and more than the standard level of −10 dB for the application frequency bands of WLAN at both the frequencies. The output of the VNA displayed the measured return loss values for the 2.4 GHz, and 5.5 GHz are −13.86 and −14.02 dB, respectively. VSWR values are 1.50 and 1.49 which is <2 shows the authenticity of the design. This antenna has application-specific dual resonant frequencies viz. 2.4 and 5.5 GHz for Wireless LAN. These both bandwidths are used for the practical targeted application. The fabricated antenna has the measured impedance bandwidth values of 7.5% and 21.8%, respectively for the both frequencies.

The software simulated and fabricated prototype measured results of the return loss are seems slightly varied from each other. This prototype was fabricated and SMA connector was soldered using standard techniques. However, the material properties and mechanical processes affect the final results. The fabricated antenna was designed using the software and fabrication prototype is shown in Fig. 5.

To analyze the distribution of the current flow in the fabricated antenna is important for the analysis of the antenna electrical characteristics. The designed antenna has defective ground plane, however, the ration patterns without DGS are shown in Fig. 7.

Figure 6 depicts the analysis of the current distribution at both the resonance frequency of 2.4 and 5.5 GHz, respectively. The current density is higher at the feed part for both the frequency compared to the rest of the area of the conducting patch. It is also clearly visible that the current concentration is higher at the 2.4 than 5.5 GHz resonance.



(a)

(b)

Fig. 5 Fabricated antenna prototype **a** top view, **b** ground plane

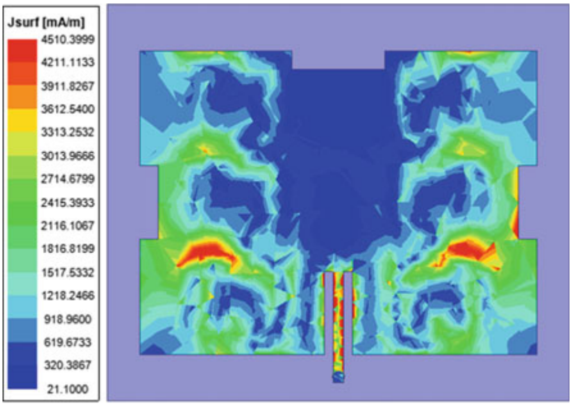
Radiation pattern is the graphical parameter to measure the coverage and gain of the antenna. The radiation patterns of both the frequencies at 2.4 and 5.5 GHz are illustrated in Fig. 8a, b for measured and simulated values. As shown in figure, both the radiation patterns of the E-plane and H-plane are of omnidirectional type and bidirectional and have a good coverage for the WLAN application.

Figure 9a, b shows the radiation pattern measurement setup. The anechoic chamber is used to measure the radiation pattern for the fabricated prototype for both E- and H-plane.

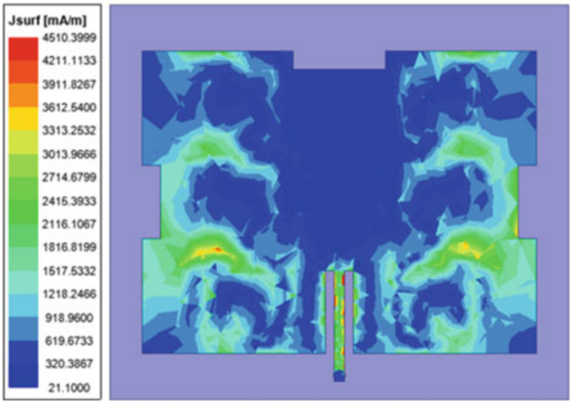
4 Conclusion

The development and analysis using various results and discussion of antenna for WLAN application are proposed and presented. The proposed antenna focuses on higher gain and improved bandwidth due to optimal use of SRR as a defected ground plane in the design. The measured results have good agreement with the software

Fig. 6 Antenna current distribution

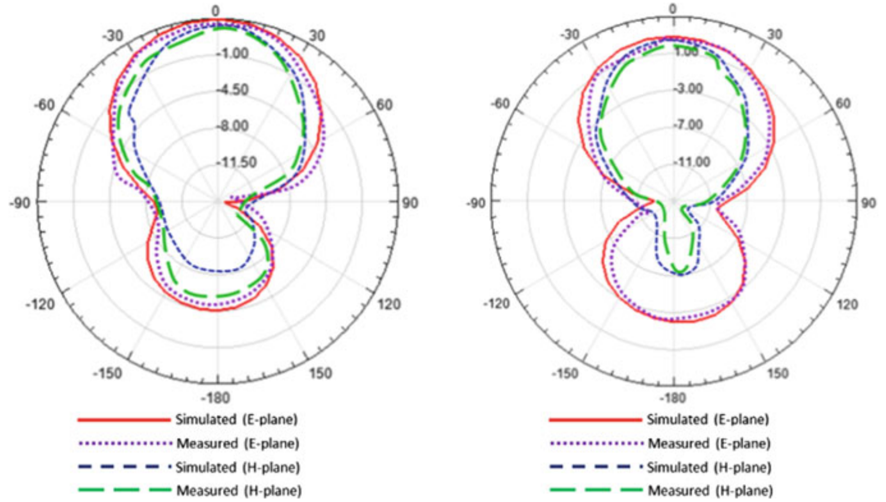


(a) 2.4 GHz



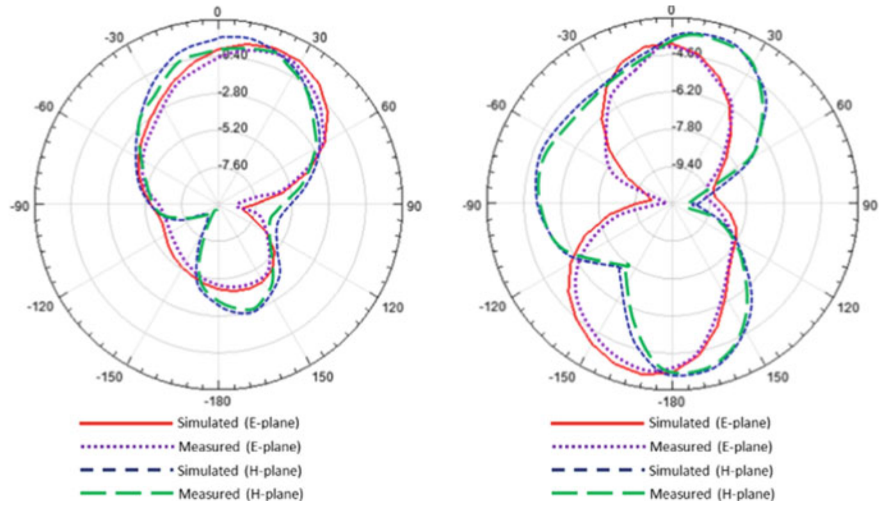
(b) 5.5 GHz

simulated results. The antenna radiates at 2.4 and 5.5 GHz frequencies is highly used for the WLAN application. The depicted average gain values and bandwidth at the resonating frequencies show that the proposed antenna is best suited for the Wireless LAN applications. The ground plane can also be made defective using CSRR structure for the better results of antenna radiation characteristics.



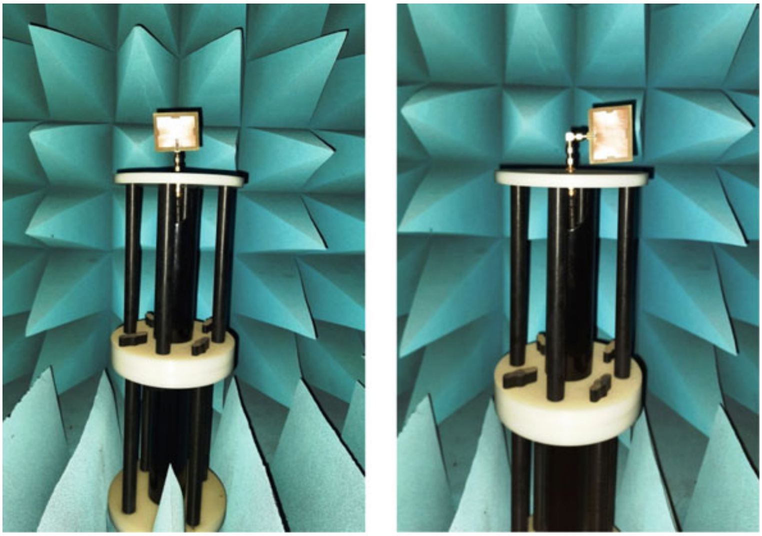
(a) 2.4 GHz (b) 5.5 GHz

Fig. 7 Simulated and measured radiation patterns for E-plane and H-plane without DGS



(a) 2.4 GHz (b) 5.5 GHz

Fig. 8 Simulated and measured radiation patterns for E-plane and H-plane with DGS



(a) (b)

Fig. 9 Radiation pattern measurement at anechoic chamber **a** E-plane, **b** H-plane

References

1. Ren W (2008) Compact dual-band slot antenna for 2.4/5 GHz WLAN applications. Prog Electromagn Res 8:319–327
2. Sah S, Tripathy MR, Mittal A (2018, January 29) Multiband and miniaturized dual layer antenna incorporated with FSS and DGS. Adv Electromagn 7(1):1–6
3. Patel U, Upadhyaya TK (2020) Dual band planar antenna for GSM and WiMAX applications with inclusion of modified split ring resonator structure. Prog Electromagn Res 91:1–7
4. Zulkifli FY, Rahardjo ET, Hartanto D (2008) Radiation properties enhancement of triangular patch microstrip antenna array using hexagonal defected ground structure. Prog Electromagn Res M 5:101–109. https://doi.org/10.2528/PIERM08101601
5. Patel U, Upadhyaya TK (2019) Design and analysis of compact μ-negative material loaded wideband electrically compact antenna for WLAN/WiMAX applications. Prog Electromagn Res 79:11–22
6. Guha D, Biswas S, Antar YMM (2010) Defected ground structure for microstrip antennas. In: Microstrip and printed antennas, pp 387–434. https://doi.org/10.1002/9780470973370.ch12
7. Upadhyaya TK, Kosta SP, Jyoti R, Palandoken M (2014) Negative refractive index material-inspired 90-deg electrically tilted ultra wideband resonator. Opt Eng 53(10):107104
8. Pandya A, Upadhyaya TK, Pandya K (2021) Tri-band defected ground plane based planar monopole antenna for Wi-Fi/WiMAX/WLAN applications. Prog Electromagn Res C 108:127–136
9. Breed G (2008) An introduction to defected ground structures in microstrip circuits. High Freq Electron 7(11):50–54
10. Desai A, Upadhyaya T, Palandoken M (2018) Dual band slotted transparent resonator for wireless local area network applications. Microw Opt Technol Lett 60(12):3034–3039
11. Khandelwal MK, Kanaujia BK, Kumar S (2017) Defected ground structure: fundamentals, analysis, and applications in modern wireless trends. Int J Antennas Prop 2017:22. Article ID 2018527. https://doi.org/10.1155/2017/2018527

12. Hanae E, Touhami NA, Aghoutane M, El Amrani S, Tazon A, Boussouis M (2014) Miniaturized microstrip patch antenna with defected ground structure. Prog Electromagn Res C 55:25–33. https://doi.org/10.2528/PIERC14092302
13. Chiang KH, Tam KW (2008) Microstrip monopole antenna with enhanced bandwidth using defected ground structure. IEEE Antennas Wirel Propag Lett 7:532–535. https://doi.org/10.1109/LAWP.2008.2005592
14. Patel U, Parekh M, Desai A, Upadhyaya T Wide slot tri-band antenna for wireless local area network/world-wide interoperability for microwave access applications. Int J Commun Syst e4897

A Review on Optimization LNA Topologies for Wideband Application

Manish Kumar, Manish Gupta, Gaurav Kumar Sharma, and Aasheesh Shukla

Abstract This paper exhibits review on different low-noise amplifier (LNA) topologies for wideband frequency applications. The LNA design metric includes gain, noise figure (NF), power dissipation, bandwidth, and linearity with broadband wideband input impedance matching. These all specified parameters required a tradeoff for optimization of LNA. This paper includes all-important LNA topologies and provides good insight to optimize LNA performance. In this paper, FOM is used to compare all the specifications on single platform. In addition, a performance summary of result has been given which will help the creation of new idea. Moreover, researchers also utilize the capabilities of soft computing like particle swarm optimization (PSO) and firefly algorithms (FA) in the area of LNA optimization and added at the end of Sect. 2.

Keywords Low-noise amplifier (LNA) · Noise figure (NF) · Complimentary common gate (CCG) · Particle swarm optimization (PSO) · Firefly algorithms (FA)

1 Introduction

Low-noise amplifier (LNA) is front-end circuit of receiver chain as shown in Fig. 1. A very weak signal is received by LNA from antenna end. Therefore, focus on LNA design in such a manner that noise is added by this circuit as little bit as possible to improve the SNR. As LNA is used in a number of applications such as healthcare, agriculture, industrial settings Internet of thing (IoT), etc. The nature

M. Kumar (✉) · M. Gupta · G. K. Sharma · A. Shukla
GLA University, Mathura, India
e-mail: manish.kumar@gla.ac.in

M. Gupta
e-mail: manish.gupta@gla.ac.in

G. K. Sharma
e-mail: gauravkr.sharma@gla.ac.in

A. Shukla
e-mail: aasheesh.shukla@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_2

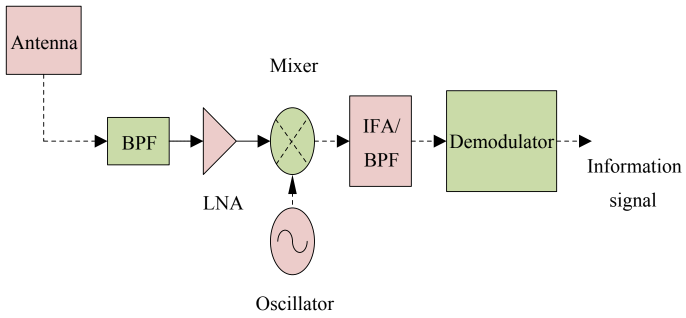


Fig. 1 Receiver end of RF circuit

of these circuits always demands a front-end circuit (LNA) which should have the metric like highly linear, low-noise figure (NF) and consumes very low power with broadband matching throughout the band. It is not possible to optimize the LNA in all criteria as mentioned above. Henceforth a tradeoff is required between parameters. Designing of LNA is applications based, as in low power application, focus on reduction in power. A high NF and nonlinearity are tolerated at the cost of low power. Scaling in Complementary metal-oxide semiconductor (CMOS) Technology makes the scholar closer to objective of low power but it also degrades the performance in terms of mobility, velocity saturation [1–4].

Differential noise cancellation techniques is used in low NF applications. Inductor-less and mutually coupled LNA techniques are used in small area-sensitive applications. This paper represents the available different design techniques to implement the LNA in Ultra-wideband (UWB) and other high-frequency bands in Sect. 2 and results and comparison with each other's in Sects. 3 and 4 represents the conclusion of paper.

2 LNA Topologies

A detailed study of different available LNA topologies has been discussed in this section for UWB band. Distributed LNA design typology is shown in Fig. 2, in this topology a broadband input impedance matching is achieved. As MOS transistors are connected in cascade. Therefore, total gain is the multiplication of all stage gain. Proposed design is good in gain, NF, and broadband matching but suffers from high power consumption and large chip area [5–7].

Common Gate (CG) topology is another technique as shown in Fig. 3a, in this topology signal is fed at source terminal and gate is common between source and drain terminal. Input impedance looking from source terminal is approximately equivalent to inverse of transcundutance ($Rs = 1/g_m$). Therefore, a broadband input impedance can be matched with antenna impedance (50 Ω) by setting the value of $g_m = 20mS$

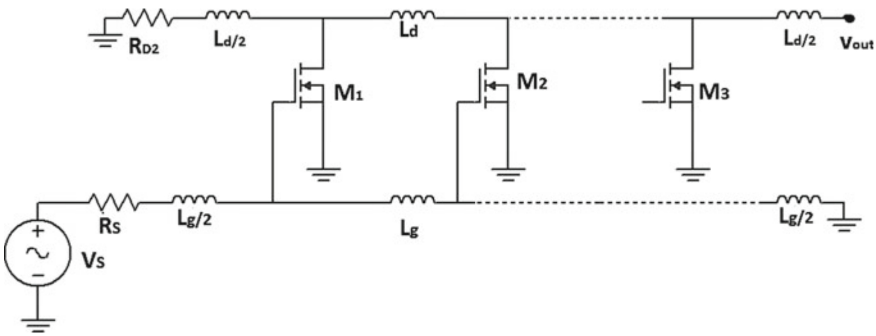


Fig. 2 Distributed amplifier topology

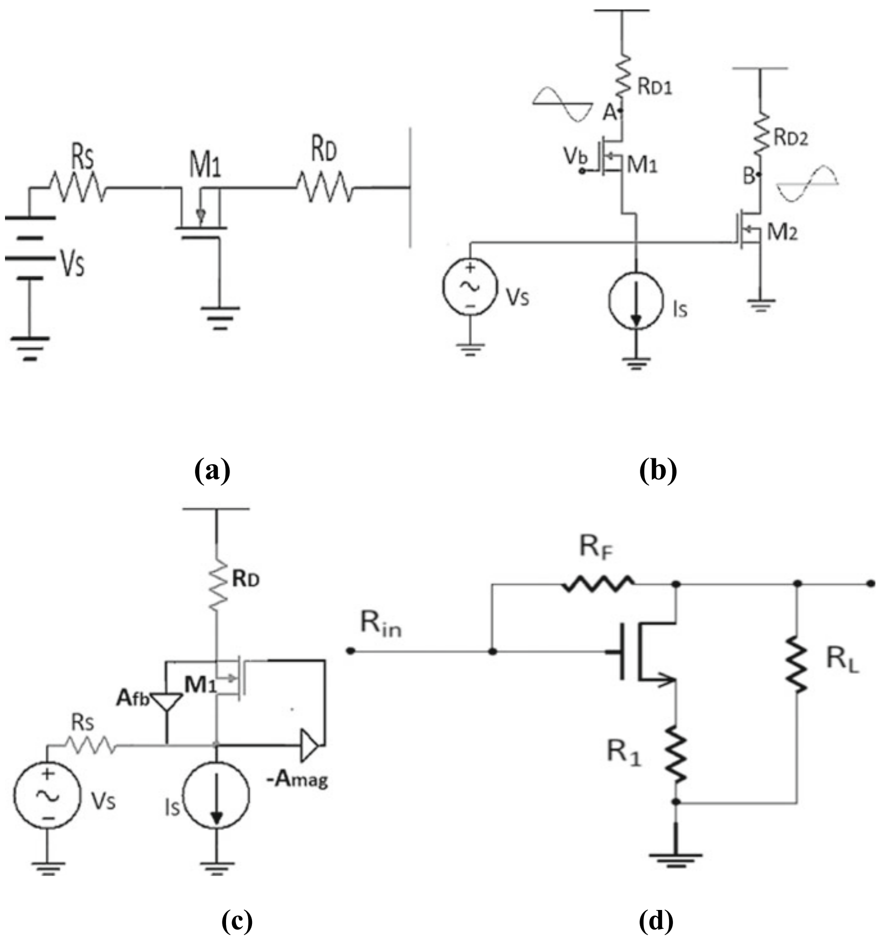


Fig. 3 **a** CG amplifier, **b** noise cancellation, **c** feedforward and feedback, **d** resistive shunt feedback

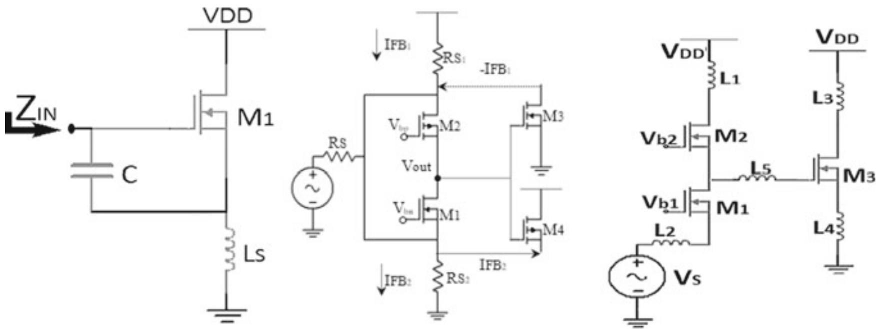
and independent from other MOS parameters. However, in this topology minimum, NF is quite high which is given by $1 + \gamma g_{d0}/g_m$, and cannot be reduced due to g_m bounded by matching criteria, where g_{d0} represents the output conductance. Due to a tradeoff between matching criteria and minimum NF, common gate offers a limited bandwidth [8]. Due to high minimum NF in CG topology, a method of noise cancellation has been adopted in differential topology as shown in Fig. 3b. Voltage gain at terminal A and B are in 180° phase difference and noise are in same phase. Therefore, resultant voltage gain ($V_A - V_B$) is the sum of individuals with reduction in noise a lot. In this topology, extra hardware is required. Henceforth consumes large power and required high biasing current [9–11].

To reduce the problem faced by differential topology a feedback with forward is used as shown in Fig. 3c in which feedback path is used to make circuit stable, while forward path helps to reduce the biasing voltage. In this topology, a significant gain minimum NF is obtained but as extra amplifiers increased the parasitic, which is nothing but the limited bandwidth [12]. To make the system stable resistive shunt feedback topology is used as shown in Fig. 3d in this gain expression is approximately equivalent to ratio of resistance R_F and R_1 but due to a number of resistance thermal noise cross the upper limit. Therefore, cannot be used in low NF-required applications [13–16].

As LNA has a lot of constrain for UWB application, a very popular technique is source degeneration as shown in Fig. 4a. In this topology, an inductor is used between source and ground. The input impedance looking at gate terminal is given by Eq. (1)

$$Z_{\mathrm{IN}} = \frac{1}{C_{\mathrm{eq}}} + sL_s + \frac{g_m L_s}{C_{\mathrm{eq}}} \tag{1}$$

Where equation one has capacitive inductive and a resistive term, first two-term can be eliminated at resonance frequency and a broadband 50 Ω impedance matching



(a) (b) (c)

Fig. 4 **a** Source degeneration topology, **b** current reused topology, **c** cascode and cascode topology can be obtained by setting the value of gm, LS, and Ceq. However, as inductors consume large area, therefore, increase in chip size of the device [17].

Since the large chip, area means high power consumption. A current reused technique is used as shown in Fig. 4b. Complementary MOS M_1 and M_2 shared the common biasing current, therefore reduction in power consumption with additional advantage of high gain. However, as both MOS connect in cascode. Therefore, current reuse topology faces the problem of high overdrive voltage [18]. As gain and noise factor (nf) of a system, in which a number of stages are connected in cascade given by

$$nf = nf_1 + \frac{nf_2 - 1}{G_1} + \frac{nf_3 - 1}{G_1 G_2} + \frac{nf_4 - 1}{G_1 G_2 G_3} + \cdots + \frac{nf_n - 1}{G_1 G_2 \ldots G_n} \quad (2)$$

where G_1, G_2, …, G_N and nf_1 nf_2 … nf_n represent the gain and noise factor of individual stage. Since the noise of first stage is the main contributor to effect of the signal to ratio (SNR) of the system. Therefore, of succeeding stage can be neglected. A cascode with cascade stage LNA is shown in Fig. 4c. This topology provides high gain and low NF but still faces the problem of high power consumption and nonlinearity [19–21].

A number of other techniques for bandwidth extension using CCG with mutually coupled have been proposed, good in terms of low power consumption, high bandwidth. But all have used a large number of inductors. Therefore, required a large area [22, 23]. In the modern era of technology, optimization plays a vital role in academia as well as in industry application. Since there is a lot of well-known optimization technique such as particle swarm optimization (PSO), firefly algorithm (FA), cuckoo search algorithm (CSA), and genetic algorithm. Due to being computationally efficient and can be easily implemented. In all these theorems, an objective function is designed to cover all the tradeoffs required to optimize their LNA metric [24, 25].

All the passive and active components as shown in Fig. 5 was simulated through PSO and FA [26]. In optimization of LNA, it becomes important to include all constrain while designing objective functions.

3 Comparison and Discussion

In this paper, a rigorous exercise has been done to compare all possible LNA topology in UWB band and upper-frequency band. LNA has a design metric and a tradeoff required in all parameters for optimization. The FOM used given by equation (2) as a tool for comparison in various parameters and technologies, Equation (33) includes power gain (S_{21}), bandwidth, linearity, power dissipation, and NF. The comparison of different results of the reported LNA_S is shown in Table 1.

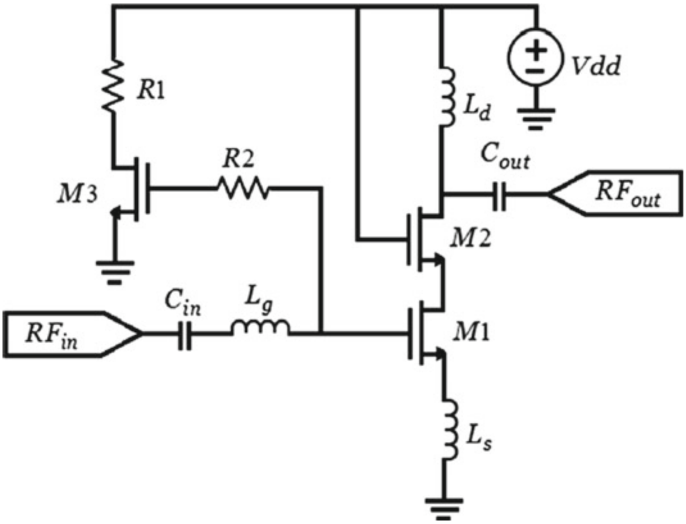


Fig. 5 The schematic of LNA optimized through PSO, CSA [24, 25]

$$\text{FOM} = 20\log_{10}\left(\frac{S21_{\text{av.[lin]}} \times \text{BW}_{[\text{GHz}]} \times \text{IIP3}_{[\text{mW}]}}{P_{\text{dc[mW]}} \times \left(F_{\text{av.[lin]}} - 1\right)}\right) \tag{3}$$

4 Conclusion

Growing research in the field of wireless communication required a front-end circuit, which consumes minimum power, high gain, and low NF. In this paper, focus was on available LNA optimization topology and tradeoff between parameters. Low power requirement makes a challenge on existence topology. Therefore, degradation in various parameters with low power is also discussed. An optimization through soft commuting algorithms (PSO, FA) is added in the last graph of Sect. 2. A rigorous and compact review had been done throughout the paper and different results are compared in Table 1.

Table 1 Performance summary of state of art with proposed work

Ref	NF (dB)	3 dB bandwidth (GHz)	S_{11} (dB)	P_{DC} (mW)	S_{21} (dB)	Supply (V)	IIP_3 (dBm)	Topology	Tech (nm)	FOM
05	2.9	0.1–11	<−12	21.6	8	1.8	−3.55	Distributed amplifier	180	–
06	3.8–6.9	2.7–9.1	<−10	7	10	0.6	–	Distributed amplifier	180	3.3
07	4.4	0.1–20	<−10	12.6	12.7	1.2	−1	Distributed amplifier	90	9.43
09	3.0	2.1	<−10	12.6	5.2	1.2	10.5	Noise cancellation	130	–
10	2.7–3.3 2.9–3.4	10 5.2	<−7.5 <−17.5	13.7 7.0	10.5 1.0.7	1 1	−3.5 6.0	Noise cancellation	65 65	2.05 −3.4
11	2	0.3–0.92	−18	3.6	21	1.8	−3.3	Noise cancellation	180	–
12	2.57	1.05–3.05	−20	12.6	16.9	1.8	−0.7	Dual negative feedback	180	–
13	2.9–4.66	3–8	<−10	3.9	16.4	1.8	−2.2–−4.3	Resistive shunt feedback	180	4.4–8.9
14	2.3	2–4.6	<−9	12.6	9.8	–	−7	Resistive shunt feedback	180	–
15	1.9–2.6	0.5–8.2	–	42	25	2.7	−4 to −16	Resistive shunt feedback	90	–
17	2.4	2.4–2.48	−10	9	19	3	–	Source degeneration	500	–
18	4.5	5	−12	0.9	9.2	0.6	−16	Current reuse	180	–
19	4	2.9–11	<−10	9.5	16	1.8	–	Cascade-cascode	180	–
20	6	>6	−10.4 to −35.5	2.15	10.4	1.1	0.99	CG with T-matching	90	–

(continued)

Table 1 (continued)

Ref	NF (dB)	3 dB bandwidth (GHz)	S_{11} (dB)	P_{DC} (mW)	S_{21} (dB)	Supply (V)	IIP_3 (dBm)	Topology	Tech (nm)	FOM
21	4.9–6	0.1–2.2	<−8	0.4	12.3	1	−11.5 −9.5	Active shunt feedback	130	17.9

References

1. Brandolini M, Sosio M, Svelto F (2007) A 750 mV fully integrated direct conversion receiver front-end for GSM in 90 nm CMOS. IEEEJ Solid-State Circ 42(6):1310–1317
2. Chen H-C, Wang T, Chiu H-W, Kao T-H, Lu S-S (2009) 0.5 V 5.6 GHz CMOS receiver subsystem. IEEE Trans Microw Theor Tech 57(2):329–335
3. Balankutty A, Yu S-A, Feng Y, Kinget P (2010) A 0.6 V zero-IF/low-IF receiver with integrated fractional-N synthesizer for 2.4 GHz ISM-band applications. IEEE J Solid-State Circ 45(3):538–553
4. Balankutty A, Kinget PR (2011) An ultra-low voltage, low-noise, high linearity 900 MHz receiver with digitally calibrated in-band feed-forward interferer cancellation in 65 nm CMOS. IEEE J Solid-State Circ 46(10):2268–2283
5. Heydari P (2007) Design and analysis of a performance-optimized CMOS UWB distributed LNA. IEEE J Solid-State Circ 42(9):1892–1905
6. Yu YH, Chen YJ, Heo D (2007) A 0.6 V low power UWB CMOS LNA. IEEE Microw Wirel Compon Lett 17(3):229–231
7. Chen M, Lin J (2009) A 0.1–20 GHz low-power self-biased resistive feedback LNA in 90 nm digital CMOS. IEEE Microw Wirel Compon Lett 19(5):323–325
8. Fu CT, Kuo CN, Taylor S (2010) Low-noise amplifier design with dual reactive feedback for broadband simultaneous noise and impedance matching. IEEE Trans Microw Theor Techn 58(4):795–806
9. Jussila J, Sivonen P (2008) A 1.2 V highly linear balanced noise-cancelling LNA in 0.13 μm CMOS. IEEE J Solid-State Circ 43(3):579–587
10. Chen KH, Liu SI (2012) Inductorless wideband CMOS low-noise amplifiers using noise-canceling technique. IEEE Trans Circ Syst I Reg Papers 59(2):305–314
11. Sanghyun W, Woonyun K, Laskar J (2009) A 3.6 mW differential common-gate CMOS LNA with positive-negative feedback. In: IEEE international solid-state circuits conference Dig. Tech. Paper (ISSCC), pp 218–219
12. Kim J, Hoyos S, Martinez JS (2010) Wideband common-gate CMOS LNA employing dual negative feedback with simultaneous noise, gain, and bandwidth optimization. IEEE Trans Microw Theor Techn 58(9):2340–2351
13. Meaamar A, Boon CC, Yeo KS, Do AV (2010) A wideband low power low-noise amplifier in CMOS technology. IEEE Trans Circ Syst I 57(4):773–782. Reg. Papers
14. Kim C-W, Kang M-S, Anh PT, Kim H-T, Lee S-G (2005) An ultra-wideband CMOS low noise amplifier for 3–5 GHz UWB system. IEEE J Solid-State Circ 40(2):544–547
15. Zhan J-H, Taylor S (2006) A 5 GHz resistive-feedback CMOS LNA for low-cost multi-standard applications. In: IEEE international solid-state circuits conference (ISSCC), February 2006, pp 721–730. Dig. Tech. Paper
16. Perumana B, Zhan J-H, Taylor S, Carlton B, Laskar J (2008) Resistive-feedback CMOS low-noise amplifiers for multiband applications. IEEE Trans Microw Theor Techn 56(5):1218–1225
17. Kim W-S, Li X, Ismail M (1999) A 2.4 GHz CMOS low noise amplifier using an inter-stage matching inductor. In: 42nd Midwest symposium on circuits and systems (Cat. No. 99CH36356), vol 2. IEEE
18. Hsieh H-H, Lu L-H (2007) Design of ultra-low-voltage RF frontends with complementary current-reused architectures. IEEE Trans Microw Theor Tech 55(7):1445–1458
19. Shim Y et al (2007) Design of full band UWB common-gate LNA. IEEE Microw Wirel Compon Lett 17(10):721–723
20. Chang J-F, Lin Y-S (2011) 0.99 mW 3–10 GHz common-gate CMOS UWB LNA using T-match input network and self-body-bias technique. Electron Lett 47(11):658–659
21. Parvizi M, Allidina K, El-Gamal MN (2016) An ultra-low-power wideband inductorless CMOS LNA with tunable active shunt-feedback. IEEE Trans Microw Theor Tech 64(6):1843–1853
22. Kumar M, Deolia VK (2018) A wideband design analysis of LNA utilizing complimentary common gate stage with mutually coupled common source stage. Analog Integrated Circ Sig Process 1–1. https://doi.org/10.1007/s10470-018-135

23. Qin P, Xue Q (2017) Design of Wideband LNA employing cascaded complimentary common gate and common source stages. IEEE Microw Wirel Lett 27(6):587–589
24. Kennedy J, Eberhart R (1995) Particle swarm optimization (PSO). In: Proceedings of IEEE international conference on neural networks, Perth, Australia
25. Schneider, J, Kirkpatrick S (2007) Stochastic optimization. Springer Science & Business Media
26. Kumar M, Deolia VK (2019) Performance analysis of low power LNA using particle swarm optimization for wide band application. AEU-Int J Electron Commun 111:152897

Speech Recognition Based Robotic Arm Writing

S. Yuvaraj, Abhishek Badholia, P. William, K. Vengatesan, and Rahul Bibave

Abstract The goal of this research is to develop a low-cost, easily controlled robotic arm system based on the Arduino Uno that is lightweight, inexpensive, and simple to use for those who are physically challenged. The arm's unique feature is that it may be completely voice-operated, which means that it will only react to commands given by a human operator to the arm. Despite the fact that there are several approaches to allow the robot to work without the need to order it manually, the voice-based strategy provides superior precision and efficiency for the robot to carry out sensitive tasks than other alternatives. The mechanical design of the robot, the selection of appropriate motors, and the use of electrical devices to direct the robotic joints are all included in the hardware element. The software component consists of algorithms that convert required words into right sequence with target places, as well as algorithms that govern the robot's movement to ensure that it moves in accordance with the requirements. The writing process is guided by the use of speech recognition software in this case. Either by implementing a system or by utilizing Android apps, this speech recognition is made available to the user.

Keywords Stepper Motor · Arduino Uno · Speech Recognition · Servo Motor

S. Yuvaraj
Department of Electronics and Communications Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India

A. Badholia
Department of Computer Science and Engineering, School of Engineering and Information Technology, MATS University, Raipur, India

P. William (✉)
Department of Information Technology, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune, India
e-mail: william160891@gmail.com

K. Vengatesan
Department of Computer Science and Engineering, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune, India

R. Bibave
Department of Electrical Engineering, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_3

1 Introduction

In recent years, the word "robot" has come to refer to an unmanly system, also known as automation, that is widely utilized in industrial applications and is becoming more popular. In general, robots are utilized and manufactured in the same way that humans are; they are referred to as machines fitted with electric systems that are capable of doing duties that are analogous to those performed by humans. Aims are being made to progress technology in the robotics industry, and efforts are being made in the areas of researching, producing, and upgrading robots for a range of practical activities, as well as in the field of artificial intelligence. Robots are developed to help people in their jobs and to reduce the amount of work required by humans. Nowadays, robots are trained to replicate human behavior and do jobs that are similar to those performed by humans. Numerous research organizations are developing robotic arms capable of performing fundamental human arm activities. Among the several activities, one of them is developing writing abilities. The suggested robotic arm may be used by physically disabled individuals to perform writing operations.

The major purpose of constructing the suggested system is to allow physically impaired folks to write anything they are thinking or feeling. Additionally, this concept may be utilized in a number of other applications, such as data accounting in industries, which can be achieved by wireless communication between different places using the same technology. Employees will save both time and effort as a result. Education is crucial in giving a variety of tactics and goals to students that are both creative and inventive. The word "robot" is now often used to refer to an unmanned system or automation, both of which are frequently seen in industrial applications. Robots are often used and built in the same way as human humans, and they are frequently referred to as machines. And these electronic devices are capable of executing similar activities to humans. With the assistance of technology advancements in the robotics field, attempts are being made to investigate, create, and enhance robots for a variety of practical uses. Robots are intended to assist people in their job and to reduce human effort.

In recent years, a large number of robots have been built to replicate human behavior and perform duties that are similar to those of humans. A large number of research organizations are working on the development of different robotic arms that will be able to execute basic functions that are comparable to those done by the human arm. While there are a variety of interests, one of them is the development of writing skills. Individuals with physical disabilities may be able to utilize the new robotic arm to do writing activities. Although the primary goal of developing this system is to enable physically challenged individuals to write what they say, the system's design can also be applied to a variety of other applications, such as data accounting in industries, which can be accomplished through wireless communication from one location to another. As a result, workers' time and effort are saved. This paper is used to operate the robot's writing activities, i.e., rapid and smooth motions of the robot. The presented robot design has both a hardware and software component. The hardware component includes the machining of the framework and the design of the robot. Additionally, in the hardware portion, we will use electrical devices and a trustworthy selection of motors for efficient guiding and control of the robot joints. Now for the software component, which includes the methods for converting the needed words to an ordered list of target points. Control algorithms are used to move the robot in accordance with the specified parameters.

The voice recognition approach provides the robot with the ability to write. The system may be given with voice recognition either via the use of a microphone or through the use of android apps. Thus, the robots are capable of simulating human writing in response to the user's input.

2 Problem Statement

The system is separated into two parts: one for the transmitter and another for the receiver. The transmitter is the first of these two parts. Three-axis accelerometers, an Arduino Uno, and a radio frequency transmitter module make up the transmitter section. There are two PMDC motors in the receiver component as well as two wheels and a motor driver integrated circuit (IC) in the receiver component. A total of two independent 5 V power supplies are needed in this instance, one for the reception components and another for the transmitter components. Each time the angle of the palm changes, the robot advances in one of four directions on the x–y plane: forward, backward, right, or left. Robots that respond to gestures may be controlled with the hand or by joystick and buttons. Here, one must use a gesture hand to control the robot. The transmitting device is held in the hand and comprises a radio frequency transmitter and an accelerometer. This enables the robot to communicate with it and tell it to do appropriate activities such as traveling in the desired direction and stopping the lift. All of these actions are controllable using hand gestures. The accelerometer is the primary component in this case. Disadvantages of the currently used approach:

- The mechanism is a bit challenging to apply
- Time consumption.

3 Related Works

A robotic arm may also be a kind of mechanical arm that can be programmed and that acts in a manner comparable to human arms in terms of range and positioning, operational capabilities, and reach, among other things. A robotic arm is used as a manipulator, and it has the capacity of doing duties that are comparable to those performed by a human arm, for example. Numerous methods of controlling a robotic arm are possible. Various persons have attempted to control robotic arms in the past via computer terminals/devices, joysticks, or by linking them with simulations, allowing them to be controlled from a remote position anywhere in the globe. Typically, the robotic arms are monitored and controlled by a central controller located at the terminal, which makes use of data read and written to and from the terminal in order to maneuver the arm to certain coordinates in space as specified by the user. It responds to what the user says and is based on writing, which serves as the human–machine interface.

How to replicate human writing using universal robot employment will be the first article we'll look at. It expresses a technique for fluid and quick motions of a dynamic robot and writing mimicry of the robot for both human kinematics and trajectory while signing, as well as a technique for fluid and quick motions of a dynamic robot. It makes use of a Wacom tablet as well as a universal robot 5 (UR5) for writing down ideas. During the production process, this technique makes use of natural human kinematics and directions, as well as a critique tablet, a professional tablet, and a Wacom tablet. The two-dimensional coordinates (x and y) are utilized as signals, and p is the pressure signal that is used to compute the amount of pressure that is required to establish contact with the writing surface (see Fig. 1). The hardware is managed via the use of customized drivers created by the CARO group at the University of Southern Denmark. URScript [1] is used to analyze the robot writing task, which is accomplished using the movep, movel, and movej functions.

Because the small tip of a calligraphy brush needs no power input and makes it impossible to examine the drawing system minutely using a camera, robotic calligraphy demanded retaining the direction for making a perfect brush stroke up in the air ahead of time. Chinese letters are drawn utilizing methods that are based on the major grouping of characters [2–5]. The second group strategy use trials to develop a brush model and its parameters in order to discover how to draw the Chinese character [6, 7]. Numerous Algorithms from the third category do comprehensive analysis of

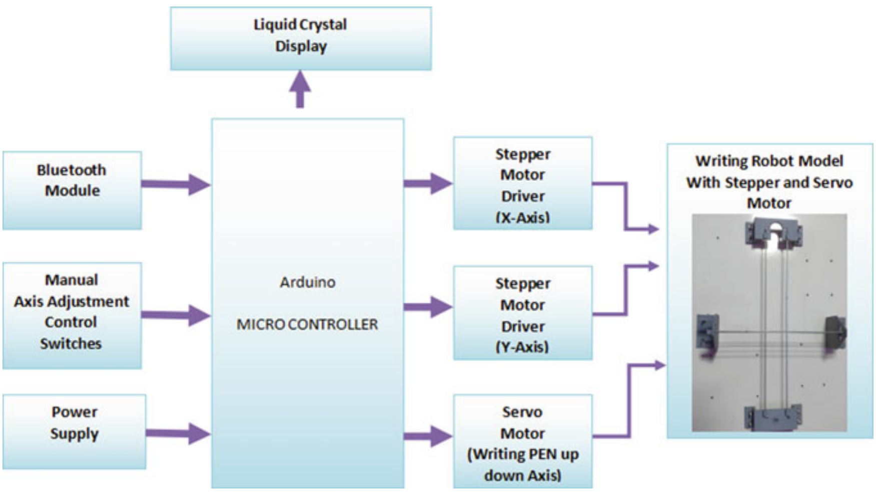


Fig. 1 Block diagram of robotic arm

brush attributes in order to parametrize strokes and manually alter parameters [9]. The researchers used visual feedback to modify the *xy*-coordinates of the strokes in a notable example is [10], where the connecting positions of strokes are specified precisely. Through a basic Research Interface [11], the controller talks with a remote computer. The algorithms described in this section are executed using the remote computer and the Robot OS [12]. The Orocos framework may be used to construct algorithms that continuously check the robot's contact and hence trajectory execution [13]. Convergence is not guaranteed in this procedure, just as it is not guaranteed in the method of placing or forming the initial spline [14]. In the papers Static signature synthesis [15–17], the authors explore a variety of unique ways for producing synthetic handwritten signature pictures for biometric applications. Motor equivalency is similar in that it is separated into two stages: the development of an effector-independent action plan and the execution of that plan via the appropriate neuromuscular channel. Using perceptual assessment, it was discovered that the synthetic ones had a median degree of perplexity of 44.06%, confirming their authenticity. Those fresh samples are created by the synthesis algorithm from the samples that have been submitted by the present user.

Robotic arms are characterized as programmable robot manipulators that do activities that are comparable to those performed by a human arm, and they are becoming more popular. A wide range of engineering prostheses is available to perform the main functions of the human arm. This article [18] is concerned with the employment of a robotic arm to assist a physically handicapped individual in the act of writing. The robotic arm and the patient's amputee hand may be connected together, allowing the patient's spoken utterances to be recorded and analyzed. The robotic arm is equipped with a pen for writing, which separates it from other similar robotic arms on the market.

The solution is designed for use with a speech recognition system that allows the user to write down whatever he or she says out loud. The robotic arm is programmed to record the words that are uttered into a microphone by anybody at any time. The robotic arm will be outfitted with a pen, which will allow it to do writing-related duties. A low-cost device that can be constructed to allow physically disabled persons to write [19] will be used in this study.

An arrangement is created for the purpose of creating a letter or a number [20, 21] in design work. The concept makes use of a timing belt and a servomotor to glide the tool over the surface in an *x*–*y* manner. The proximity sensors are used to prevent the end effector from colliding with the supporting beam during its operation. The microcontroller is in charge of the movement of the robot [22, 23].

4 Proposed Method

Through wireless connectivity, the writing robot enables physically unable individuals to write. The Inkspace program is used to code the G-Code file for movement, and the processing software is used to transfer the G-Code file to the microcontroller.

Once this is completed, the CNCashieldadrive sends instruction signals to the servo and stepper motors. The *X*–*Y* axis now acts in accordance with the settings given by the user in the controller unit, as shown in the diagram. When the required code is sent to the controller block, the controller block interacts with the motor driver and the DAC, which provides the pulse width signal for the motor unit to process, and the final output is written down and shown on the paper by means of the output unit.

The logic algorithm for the controller is written using the Arduino IDE software.

Control logic for Arduino programs: The Arduino IDE software is used to program the microcontroller [24]. The code includes a collection of libraries, setup, and loop functions. The user's vocal commands provide the input.

The recommended methodology is as follows:

- Receiver and controller module
- Writing module
- Transmitter module.

4.1 Transmitter Section

The android application is utilized in the transmitter part to deliver voice instructions to the main section. The mic is activated through voice instructions. The user's input instructions will be delivered through an android application or microphone [25]. When the user speaks the word or instructions, the programmer recognizes them and delivers the appropriate signal.

Speech recognition is an interdisciplinary branch and subfield of computational linguistics that develops methods and technology to aid in the detection and translation of spoken language into written language using a system.

The voice recognition system requires training by individual speakers who read words or terminology into the computer. Precision and accuracy are boosted when the system analyses the individual's distinctive voice and uses it to recognize the individual's speech. Speaker-independent systems are those that do not need training. Whereas systems that are speaker-dependent need training.

The term "voice recognition" refers to the process of recognizing the speaker rather than the content of the speaker's speech.

Speech recognition technology refers to the ability of a program or computer to recognize words or phrases spoken in a language and convert them to a machine-readable format. It is used in a variety of applications. The microphone captures the user's voice as it is spoken into the device's microphone. The Google voice API is then utilized to convert the audio signal to text, and the process is repeated. This modification is accomplished via the use of the Hidden Markov model approach. Following that, the Bluetooth module is assigned a value to use.

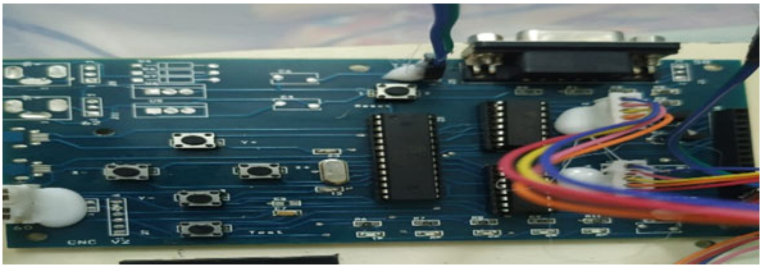


Fig. 2 Microcontroller

4.2 Receiver Section

Bluetooth gets the relevant signal from the android application in the receiver area. Then it sends orders to the Arduino board to finish the procedure.

Bluetooth allows us to transmit and receive serial data wirelessly. Bluetooth-enabled devices will provide wireless point-to-point connectivity, as well as wireless access to mobile phones.

Microcontroller

The Arduino module processes the data received through Bluetooth. The Arduino module interfaces and connects the (x, y) axis stepper motors and the z-axis servo motor. The Bluetooth module transmits data to the controller.

The analog signals are transformed into digital signals with the help of the Inkscape and Arduino IDE software, and the Arduino Uno is then interfaced with the results. The Arduino Uno is responsible for transmitting all of these signals to the stepper motors on the x and y-axes, as well as to the servo motor (Fig. 2).

4.3 Writing Module

In response to Arduino inputs, the motor moves and spins the belts along the desired axis. In the writing module, the stepper and servo motors are set up and configured. The X- and Y-axes are moved with the help of stepper motors. A servo motor, on the other hand, is connected to the pen and controls the Z-axis, which allows the pen to move up and down. Using the data obtained from the Arduino module as a basis. Finally, the writing action is carried out based on the input provided by the user.

As seen in the figure, the step motor aids in the movement of the belts, allowing the servo motor to reach the precise location where the text is to be written when required. The servo motor is used to assist in the writing process. It allows the pen to go up and down the page, allowing it to write in the appropriate places on the page. The outcome is that it becomes capable of noting down the words and concluding the words that have been entered (Figs. 3, 4 and 5).

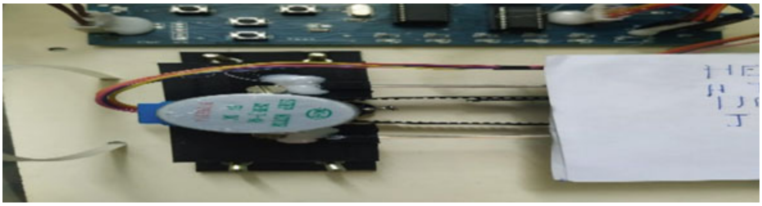


Fig. 3 Step motor

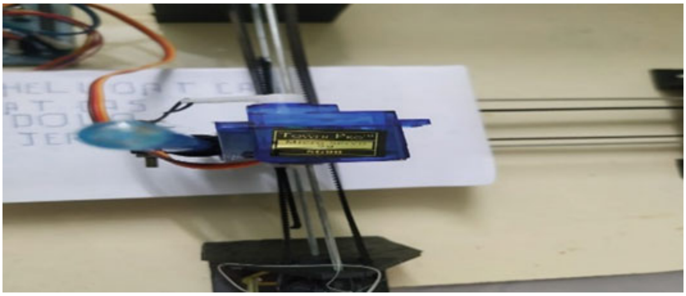


Fig. 4 Servo motor

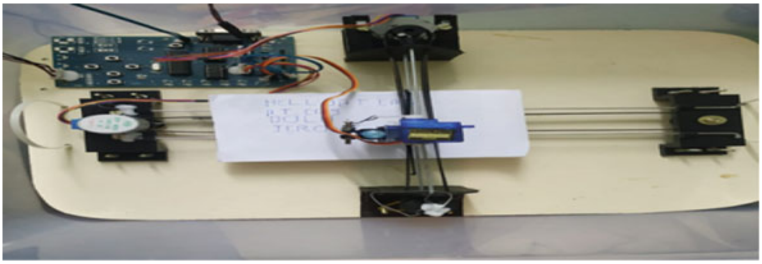


Fig. 5 Complete system module

5 Result and Analysis

The motor moves and rotates along the specified axis in response to Arduino signals. Finally, the user's input is used to perform the writing action. As seen in the graphs below, we tested the whole system using a variety of input instructions and examined the results based on each input and output (Fig. 6 and Table 1).

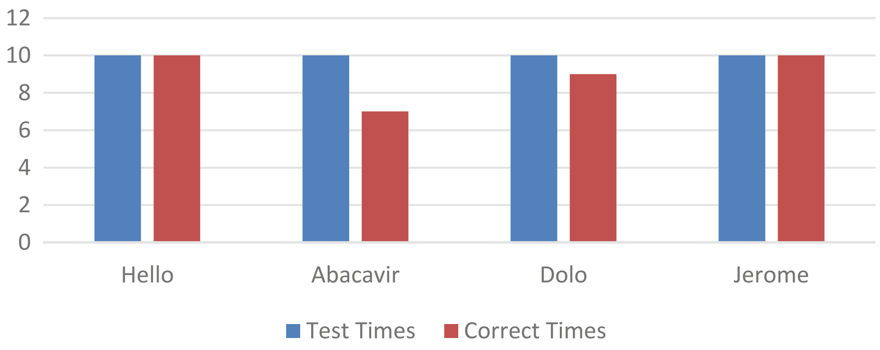


Fig. 6 Bar graph for the analysis

Table 1 Test and correct times table

Command	Correct times	Test times
Jerome	10	10
Dolo	9	10
Abacavir	7	10
Hello	10	10

6 Conclusion

Physically disabled persons may benefit from the use of the suggested robotic arm, which can execute writing activities. The major goal of the proposed method is to allow physically impaired persons to express themselves verbally via writing. Besides that, this architecture may be used in a number of other applications, such as data accounting in the industrial sector, which can be achieved by wireless connections between different places. As a result, workers' time and effort are saved.

References

1. Universal Robot Manual, URScript Programming Language. http://www.sysaxes.com/manuels/
2. Yao F, Shao G, Yi J (2004) Extracting the trajectory of writing brush in Chinese character calligraphy. Eng Appl Artif Intell 17(6):631–644
3. Lo KW, Kwok KW, Wong SM, Yam Y (2006) Brush footprint acquisition and preliminary analysis for Chinese calligraphy using a robot drawing platform. In: 2006 IEEE/RSJ international conference on intelligent robots and systems, pp 5183–5188
4. Lam J, Yam Y (2007) A skeletonization technique based on Delaunay triangulation and piecewise Bezier interpolation. In: 2007 6th international conference on information, communications signal processing, pp 1–5
5. Yao F, Shao G Modeling of ancient-style Chinese character and its application to CCC robot. In: Networking, sensing and control

6. Lam J, Yam Y (2009) Stroke trajectory generation experiment for a robotic Chinese calligrapher using a geometric brush footprint model. In: IEEE/RSJ international conference on intelligent robots and systems, 2009. IROS 2009, pp 2315–2320
7. Leung H, Wong S, Ip H-S (2008) In the name of art. Sig Process Mag 25(4):49–54
8. Kwok KW, Lo KW, Wong SM, Yam Y (2006) Evolutionary replication of calligraphic characters by a robot drawing platform using experimentally acquired brush footprint. In: IEEE international conference on automation science and engineering, 2006. CASE'06, 2006, pp 466–471
9. Yao F, Shao G, Yi J (2004) Trajectory generation of the writingbrush for a robot arm to inherit blockstyle Chinese character calligraphy techniques. Adv Robot 18(3):331–356
10. Kwok KW, Yam Y, Lo KW (2005) Ga-based homography transformation for vision rectification in robot drawing system. In: 44th IEEE conference on decision and control, 2005 and 2005 European control conference. CDC-ECC'05, pp 2047–2052
11. Schreiber G, Stemmer A, Bischoff R (2010) The fast research interface for the KUKA lightweight robot. In: IEEE workshop on innovative robot control architectures for demanding (Research) applications—how to modify and enhance commercial controllers (ICRA 2010)
12. Quigley M, Gerkey B, Conley K, Faust J, Foote T, Leibs J, Berger E, Wheeler R, Ng A (2009) ROS: an open-source robot operating system. In: ICRA workshop on open source software, vol 3
13. Bruyninckx H, Soetens P, Koninckx B (2003) The real-time motion control core of the Orocos project. In: Proceedings. ICRA'03. IEEE international conference on robotics and automation, vol 2, pp 2766–2771
14. Galbally J, Fierrez J, Martinez-Diaz M, Ortega-Garcia J (2009) Improving the enrollment in dynamic signature verification with synthetic samples. In: Proceedings of the IAPR international conference on document analysis and recognition (ICDAR), pp 1295–1299, July 2009
15. Miatliuk K, Wolniakowski A, Diaz M, Ferrer MA, Quintana JJ (2019) Universal robot employment to mimic human writing. In: 20th international Carpathian control conference (ICCC)
16. Mueller S, Huebel N, Waibel M, D'Andrea R (2013) Robotic calligraphy—learning how to write single strokes of Chinese and Japanese characters. In: IEEE/RSJ international conference on intelligent robots and systems
17. Ferrer MA, Diaz-Cabrera M, Morales A (2015) Static signature synthesis: a neuromotor inspired approach for biometrics. IEEE Trans Pattern Anal Mach Intell 37(3)
18. Balaganesh M, Aadhitya CS, Logashanmugam E, Manikandan R (2010) Robotic arm showing writing skills by speech recognition. In: INTERACT-2010, Chennai, pp 12–15
19. Chao F, Huang Y, Zhang X, Shang C, Yang L, Zhou C, Hu H, Lin C-M (2017) A robot calligraphy system: from simple to complex writing by human gestures. Eng Appl Artif Intell 59
20. Deussen O, Lindemeier T, Pirk S, Tautzenberger M (2012) Feedbackguided stroke placement for a painting machine. In: Proceedings of 8th symposium on computational aesthetics in graphics, visualization, and imaging, pp 25–33
21. Ferrer MA, Diaz-Cabrera M, Morales A (2015) Static signature synthesis: a neuromotor inspired approach for biometrics. IEEE Trans Pattern Anal Mach Intell 37(3):667–680
22. Kwok KW, Lo KW, Wong SM, Yam Y (2006) Evolutionary replication of calligraphic characters by a robot drawing platform. In: IEEE international conference on automation science and engineering, pp 466–471
23. Mueller S, Huebel N, Waibel M, D'Andrea R (2013) Robotic calligraphy—learning how to write single strokes of Chinese and Japanese characters. In: IEEE proceedings of IROS, pp 1734–1739

24. Girhe P, Yenkar S, Chirde A (2018) Arduino based cost effective CNC plotter machine. Int J Emerg Technol Eng Res (IJETER) 6(2)
25. William P, Badholia A (2021) Analysis of personality traits from text based answers using HEXACO model. In: 2021 international conference on innovative computing, intelligent communication and smart electrical systems (ICSES), pp 1–10. https://doi.org/10.1109/ICSES52305.2021.9633794

A Comprehensive Investigation on Role of Machine Learning in 6G Technology

S. Bharathi and P. Durgadevi

Abstract Fifth generation network stands for broadband cellular network. 5G wireless networks face considerable problems due to the fast expansion of data-centric intelligent systems. It cannot completely fulfill the high compute, low latency, and other requirements of 6G applications. 6G is the next generation of wireless technology, and it is similar to 5G cellular technologies in that respect. In 6G networks, higher frequencies may be used, allowing for more capacity and decreased latency. The 6G sector is expected to make substantial advances in imaging, presence technologies, and location awareness. There will be numerous smart usage scenarios utilizing Machine Learning (ML) technology in the 6G perception that creates complicated reciprocities and a large data warehouse as well as comprehensive services. New opportunities will arise, along with a variety of privacy challenges, when ML and 6G come together. ML structures that are hidden, or the exact source of ML, can preserve privacy in 6G. On the other hand, ML can be taken by surprise or misapplied, leading to privacy violations. It is worth noting that 6G and ML collaboration might be a double-edged sword in some scenarios, rather than intrude or preserve privacy. Due to several ongoing initiatives, this study focuses to provide a comprehensive analysis of role of machine learning (ML) in 6G and its trends and technologies.

Keywords 6G · Machine learning · Technology

1 Introduction

The term 6G refers to the latest wireless technology, which is also known as 5G. For the development of 6G wireless technology, which will replace 5G wireless,

S. Bharathi (✉) · P. Durgadevi
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai 26, India
e-mail: bs3778@srmist.edu.in

P. Durgadevi
e-mail: durgadep@srmist.edu.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_4

work has already begun in 2019. In 6G networks, "connected things" will become "networked intelligence," as opposed to earlier cellular network generations that were only entrusted with a restricted set of AI responsibilities. After the commercialization of 5G wireless communication networks in the United States a few years ago, researchers and businesses have started investigating 6G systems. 5G wireless networks face considerable problems due to the fast expansion of data-centric intelligent systems. For a fact, haptic Internet-based telemedicine necessitates an air interface latency of less than 0.1 ms(ms) [1]. Even so, the current latency is only 1 ms, which is not acceptable. The ubiquitous mobile ultra-broadband, ultra-high data density, and ultra-high-speed-with-low-latency communications of 5G cannot completely fulfill the high compute, low latency, and other requirements of 6G applications [2]. There are more stringent criteria for power consumption, latency, and reliability in 6G than there were in 5G. Furthermore, it offers superior quality service (QoS) as well as greater coverage than previous generations of wireless communication. Because of the increased roles of intelligence and autonomy, 6G is being hailed as a transformational wireless communication generation. Furthermore, it offers superior quality service (QoS) as well as greater coverage than previous generations of wireless communication. Because of the increased roles of intelligence and autonomy, 6G is being hailed as a transformational wireless communication generation (Fig. 1).

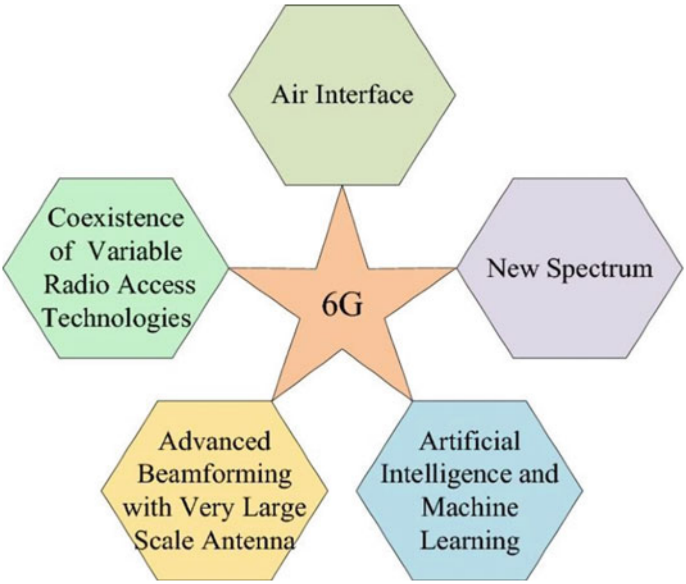


Fig. 1 The vision of 6G

2 Transformation of Network Standards

2.1 First Generation Network

The first generation (1G) wireless network was established in the 1980s and was primarily designed for voice communications. Transmits data via analog signals and does not have a standard for wireless communication. Hard handovers are one consequence, but others include a lack of security and privacy assurances and poor transmission efficiency. Due to the lack of encryption in phone services, Data transfer and telephone calls are neither private nor secure. It is no surprise, however, that the network's security is a major priority for everyone on it, including the users.

2.2 Second Generation Network

Mobile telephony and basic data services like fax and text messaging were made possible thanks to 2G's high-quality, secure voice and data networks. The 2G network relied heavily on GSM technology. The "Groupe Speciale Mobile"—a team of technical experts assembled by the Postal and Telecommunications Administration (CEPT) Conference to create digital mobile communication technology—was known as such [1]. To complement the landline Integrated Services Digital Network (ISDN) technology, this wireless version was created. Global Systems for Mobile Communication (GSM) became the acronym's full name when it was first coined. In 1991, the 900 MHz frequency was used to implement the GSM standard. Base Station Subsystem (BSS) or the radio network makes up the GSM architecture, which is made up of the Mobile Station and Network and Switching Subsystems (MSNSS). However, a smart network subsystem is incorporated, enabling smart functionality including prepaid services and short messaging services (SMS).

2.3 Third Generation Network

3G was created in the early twenty-first century as an improvement on 2G's characteristics. It allowed for higher data transfer speeds between 300 kbps and 30 Mbps, as well as services like video conferencing and remote monitoring systems. Wideband Code Division Multiple Access (WCDMA) and Universal Mobile Telephone Service (UMTS) were critical in the development of 3G. The 3GPP was established in 1998 to oversee the implementation of UMTS and other 3G-enabling technologies. The US-based 3GPP2 organization was also established to produce worldwide 3G system specifications [3]. It was found that [56, 57] provide key indications of 4G LTE's performance, whereas [58] presented the achievements of their route loss prediction model using a radial-basis-function neural network. With a data throughput of 20–100 ms faster and reduced latency, 4G LTE outperformed 3G.

2.4 Fourth Generation Network

4G Long-Term Evolution (LTE) networks were deployed in 2009, with internet service of up to 1 Gb/s downstream and up to 500 Mbits upstream [4]. There is a significant improvement in terms of spectrum efficiency as well as reduced latency in these systems, which makes them ideal for high-end applications such as DVB, HDTV, or even video chat. They also give higher spectrum efficiency and lower latency. With the LTE, old and new technologies are combined, including CoMR, MIMO, and Orthogonal Frequency Division Multiplexing (OFDM), as well as Multiple Input Multiple Output (MIMO).

2.5 Fifth Generation Network

5G networks have a function-based core network [5]. Networks are becoming more dynamic than ever before because of the usage of NFV, SDN, and cloud technologies, posing a broader spectrum of security **vulnerabilities**. The signaling burden can be overwhelmed by a large number of devices and services, increasing the risk of a denial-of-service (DoS) or resource assault. Currently, two approaches exist for dealing with signal overloads [6]. For starters, a large number of devices can communicate with each other thanks to lightweight authentication and key agreement methods. The second method includes the use of protocols that enable devices to be grouped together utilizing different group-based AKA protocols.

2.6 Sixth Generation Network

Data rates in the 100 Gbps–1 Tbps range are predicted with 6G, as is latency of less than 1 ms [7]. This paves the way for future use in holographic communication, tactile internet, virtual reality, and many other fields. Table 1 highlights the advancements in wireless technology, data speeds, and supporting applications from 1 to 6G. Table 1 compares 5G, Beyond 5G (B5G), and 6G technologies, as well.

Table 1 Comparison of 5G, B5G, and 6G

Description	5G	Beyond 5G	6G
Bandwidth of RF	6 GHz	6 GHz	• The absence of radio frequency (RF) (optical, VLC)
	mm-Wave for wireless communication	mm-Wave for wireless communication	• 6 GHz
Need for data rate	20 Gigabits per second	100 Gigabyte per second	1 Tera bits per second
Radio only delay	100 ns	60 ns	10 ns
End to end delay	5milli second	One millisecond	Less than one millisecond
Processing delay	100 ns	50 ns	10 ns
Type of devices	Sensors, smartphones, drones	Sensors, smartphones, drones, XR equipment	Sensors and distributed ledger technology, CRAS, brain-computer interface, XR, and smart implants

3 Technologies Supported by 6G

Some essential technologies, like those described above, are already proven to be beneficial in the 6G network in critical regions. They provide 6G networks with excellent dependability, low latency, and safe and efficient transmission services. To be sure, many of these new technologies come with a price: increased security and privacy issues. This is what we'll be talking about in this part.

3.1 AI

All other 6G network technologies put together, artificial intelligence (AI) is commonly regarded as one of the most important components of the network's future architecture. Even to claim that artificial intelligence (AI) has received a lot of attention on the network would be a gross underestimation. Because of all the attention, new security and privacy issues have emerged [8]. Although artificial intelligence (AI) is presumably operated in isolated regions where large volumes of training data and powerful yet private computer centers are accessible in the 5G network, the 6G network will see increased emphasis on AI. Once again, artificial intelligence (AI) technologies may be broken down into the many architectural levels they serve inside an organization. The physical layers include devices like data connections and network infrastructure while the computer layers include SDN, NFV, cloud/edge/fog computing, etc. Each of these points will be addressed in turn in the paragraphs that follow.

The desired range for NO_2 to be considered as "Good" should be between 0–40 μg/m^3, 41–80 μg/m^3 to be considered "Satisfactory", 81–180 μg/m^3 to be considered "Moderate", will be considered as "Poor" if the range is between 181–280 μg/m^3, "Very poor" if the range is between 281–400 μg/m^3 and "Severe" if the range exceeds 400 $^{\mu g}$/m^3.

3.2 Molecular Communication

All other 6G network technologies put together, artificial intelligence (AI) is commonly regarded as one of the most important components of the network's future architecture. Even to claim that artificial intelligence (AI) has received a lot of attention on the network would be a gross underestimation. Because of all the attention, new security and privacy issues have emerged [8]. Although artificial intelligence (AI) is presumably operated in isolated regions where large volumes of training data and powerful yet private computer centers are accessible in the 5G network, the 6G network will see increased emphasis on AI. Once again, artificial intelligence (AI) technologies may be broken down into the many architectural levels they serve inside an organization. The physical layers include devices like data connections and network infrastructure while the computer layers include SDN, NFV, cloud/edge/fog computing, etc. Each of these points will be addressed in turn in the paragraphs that follow.

3.3 Quantum Communication

Another message technique with tremendous application potential in 6G networks is quantum communication. Another of the primary advantages is that it may greatly improve the security and reliability of data transfer. Quantum communication is vulnerable to eavesdropping, measurement, and replication by an adversary [9]. This means that the receiver will always be aware of the obfuscation. Theoretically, quantum communication may provide complete security, making it ideal for long-distance communication. It provides several new options and raises communication to a level that is unattainable by conventional communication technologies (Fig. 2).

Quantum communication, on the other hand, is not yet a cure-all for all security and privacy problems. Even though quantum encryption for quantum communication has come a long way, fiber attenuation and operation mistakes make long-distance quantum communication a difficult proposition.

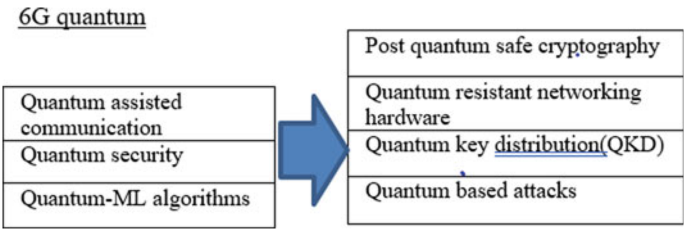


Fig. 2 Quantum communication with 6G

3.4 Blockchain

In a 6G network, Blockchain technology offers a wide range of applications. Decentralized networks, distributed ledgers, and spectrum sharing are just a few examples. It is possible that blockchain-based network decentralization will ease network administration while also improving in spectrum resources than the mm-wave region, which is used for terahertz communication. It also makes use of electromagnetic and light wave phenomena. THz, like other technology, presents privacy and security concerns [10]. These are mostly concerned with spoofed identities and harmful activity. THz frequencies have an electromagnetic signature that might be utilized for physical layer authentication operations. Despite the widespread belief that THz communication makes eavesdropping harder, an eavesdropper might nevertheless intercept a signal if it is sent in narrow beams. This type of eavesdropping assault is discussed, though.

3.5 Terahertz Technology (THz)

The high transmission rates required for 6G make mm-wave bands ineffective, notwithstanding their widespread use in 5G networks. There is no way to employ more of the Radio Frequency (RF) spectrum for future technology. Because of these considerations, terahertz technology has been developed. The 0.1–10 THz range, which is more rich use it more often than most people realize. Siri and Alexa, for example, rely on machine learning to understand concepts or words in a discussion. Keep in mind that Alexa will never ask you to fix the same mistake more than once.

Detection of dangerous software and online customer service are both automated decision-making uses (chatbots). Your selections are painstakingly categorized and calculated by an algorithm when you input your preferences and no person is on the other end of the line to appreciate or find useful.

Because the dynamic contexts of daily life are too complex to write from the ground up, machine learning is critical to AI. Using machine learning, an AI may learn to identify hints and meanings in the same manner as people do, but at a far lower cost.

3.6 Visible Light Communication (VLC)

The utilization of visible light communication technology is a potential strategy for meeting the rising need for wireless connection. Although it is still in the early phases of development, VLC has already been used in a variety of applications, including indoor positioning systems and the Vehicular Ad Hoc Network (VANET) network. For instance, Luo et al. [11] have published hundreds of articles on VLC-based locating technology.

VLC offers greater bandwidths and is more resistant to electromagnetic interference than RF, which suffers from both interference and a high latency.

4 Machine Learning Algorithm with 6G

Machine learning, a subset of AI, learns implicitly using algorithms, pattern recognition, and inference [8]. Instead of having to enter every piece of information a software has to provide, users may just type it in themselves.

Machine learning has a wide range of applications, use more often than most people realize. Siri and Alexa, for example, rely on machine learning to understand concepts or words in a discussion [12]. Keep in mind that Alexa will never ask you to fix the same mistake more than once.

Detection of dangerous software and online customer service are both automated decision-making uses (chatbots). Your selections are painstakingly categorized and calculated by an algorithm when you input your preferences and no person is on the other end of the line to appreciate or find useful (Table 2).

Because the dynamic contexts of daily life are too complex to write from the ground up, machine learning is critical to AI. Using machine learning, an AI may learn to identify hints and meanings in the same manner as people do, but at a far lower cost [13].

5 Machine Learning's Potential in the 6G Era

It is inconceivable to think about 6G's future without also considering the role that machine learning plays in it. Machine learning will benefit greatly from the advent of 6G, but only because it requires machine learning to function [6].

One of the major problems with 6G, as we have discussed before, is that the technology to make it a reality simply does not exist. Faster speeds, reduced latency, and greater data transfer capabilities are all advantages of 5G. This is mainly being accomplished using resources that we currently have on hand. (This is also why 5G infrastructure issues are a thing.)

Table 2 Comprehensive analysis of ML algorithm with 6G and its applications

#	Method	Pros	Cons	Algorithms	Applications
1	Supervised algorithm	Task-specific effective	Require large labeled training data	Linear regression, SVM, random forest	Wireless resource allocation, encoder, and decoder design
2	Unsupervised algorithm	Unlabeled data utilization	Implicit, less accuracy	PCA, K-means, latent variable models	Detect malicious attack, NOMA user grouping, User association
3	Reinforcement algorithm	Indirect supervision, learn toward interactive	Require many resources, interoperability	A3C, Q-learning, SARSA, PPO, DDPG	Autonomous driving, wireless caching, UAV
4	Batch processing	Easy deployment, one-off training	Less representative	Gradient descent (mini batch)	Wireless caching, offloading
5	Online	Updating done regularly, Efficient memory	Less defenseless to noise data	Stochastic gradient descent, online convex optimization	CR, UAV movement
6	Model based method	High computational accuracy	Coarse function approximation	Decision tree, Neural network, Probabilistic graphical model	mm-Wave modeling estimation, Trajectory prediction
7	Instance-based method	Excessive consumption of memory	Learning stage is not required	KNN algorithm	Power spectrum density estimation

To get to 6G, we need to create the technologies that will overcome these problems. Currently, the most anticipated uses for machine learning with 6G revolve around the potential benefits it would bring to infrastructure—benefits that will appear more like requirements. Let's take a deeper look at each of these three possibilities.

5.1 Communication-Efficient Distributed Training and Inference

Like the human brain, artificial intelligence requires a significant amount of energy to function properly. To use wireless connections correctly and efficiently, you need a lot of data transfer capacity. Using AI at terabyte-level frequencies would put too much strain on the system, making it unusable [14].

Some academics offer "distributed training and inference" architectures to get around the shortage of resources. This resource constraint has been circumvented by some researchers using “distributed training and inference” architectures (DTI) [15]. Therefore, resources for “smart” tasks will be distributed across clouds, networks, and devices. Even though it is not optimal, the Hong Kong University of Science and Technology thinks it is a significant step forward.

5.2 *Smart Network Re-Configurability*

Researchers expect to dramatically minimize the likelihood of bandwidth constraints by establishing a distributed resource ecosystem. Machine learning, on the other hand, will be required to deal with resource reallocation and reconfiguration more effectively when network circumstances change [16]. Quantum machine learning, according to an IEEE article, is the term used to describe the capability.

5.3 *Big Data Analytics*

5G will help big data become even more popular by facilitating the Internet of Things. With 6G, these tendencies will continue, and we’ll be able to make new use of data analytics [17]. Machine learning that is self-sustaining and proactive is expected to be employed in healthcare, the military, and certain commercial applications in the near future.

Analysis of AQI Average for all the states during the four phases BFL, DFL, AFL, and DSL.

6 Future Research Directions for 6G

AI can improve network performance by leveraging its tremendous learning and reasoning capabilities. In this part, we will discuss some potential future research problems and solutions for AI-enabled intelligent 6G networks.

6.1 *Efficiency and Precision in Computation*

Massive volumes of data gathered in 6G networks, as well as complicated network topologies, provide problems for the AI-enabled learning and training process. Furthermore, limited computer resources may be insufficient to handle huge amounts of high-dimensional data in order to satisfy the training accuracy rate. Recently, residual channels, types of media, pattern recognition, and offline training have been identified as potential approaches for boosting convergence speed, minimizing complicated calculations, and improving accuracy rate.

6.2 Learning Framework Robustness, Scalability, and Flexibility

As previously said, 6G networks (e.g., vehicle networks and UAV-enabled networks) display significant dynamics in several areas, such as BSs affiliations, wireless channels, network topologies, and mobility dynamics [18]. Devices or terminals that join or depart networks, for example, may have varying QoS and QoE needs. All of the aforementioned risks in dynamic networks necessitate continual changes to AI learning algorithm settings. Learning frameworks with high resilience, scalability, and flexibility are critical for supporting an infinite number of interacting entities and offering high-quality services in real-world dynamic networks [19]. As a result, the question of how to develop robust, scalable, and flexible learning frameworks for 6G networks remains unresolved.

6.3 Hardware Development

When developing 6G networks, the accompanying hardware development is quite difficult. On the one hand, while working in the mm-Wave and THz bands, hardware components use a lot of energy and are expensive. On either hand, some devices/terminals have limited storage and computational power. Despite the benefits of AI learning algorithms in terms of learning and recognition ability, they often need significant computational complexity, power consumption, and enough computer resources. As a result, collaboration between hardware components and AI evolutionary computation should be encouraged, which will necessitate intensive research efforts.

6.4 Distributed Energy

Some devices on the ground in 6G networks can be charged by power stations, while underwater, air, and space infrastructures, as well as some sensors, cannot be charged by power stations [20]. Furthermore, 6G networks must be able to link a large number of smart low-power devices in a flexible manner. In this situation, improved energy management systems in 6G networks become important [21]. AI methods have the potential to assist these infrastructures and devices in optimizing their energy management strategies by intelligently regulating their energy consumption or harvesting energy (e.g., energy harvesting and wireless power transfer), therefore increasing use time.

7 Conclusion

In this paper, assessed the essential technologies of 6G with machine learning to make it a reality. According to investigations on these technologies, research into each technology is still in its early stages, despite the fact that certain works have demonstrated their potential possibilities for future wireless communications. As a result, more basic effort, such as theoretical analysis and hardware design, is required to expedite the development of each technology. There's no doubt that the 6G network will raise the bar on network assistance. Our research on the comprehensive investigation on role of machine learning in 6 g technology is presented in this paper. To begin, we provided a review of technological advances from 1 to 5G, laying the groundwork for the development of the 6G network. Afterward, we looked at the technologies that are supported by the future 6G and then we compared ML algorithms with the applications on wireless communication of the 6G network to discover the trends of tomorrow's innovations and its future directions. This study concludes with a prediction of how much the 6G network will increase use. We anticipate that our research will pique people's interest and lead to more research on the technologies that used by 6G along with the machine learning algorithms.

References

1. Dogra RKJ, Jain S (2021) A survey on beyond 5G network with the advent of 6G: architecture and emerging technologies. IEEE Access 9:67512–67547. https://doi.org/10.1109/ACCESS.2020.3031234
2. Jiang W, Han B, Habibi MA, Schotten HD (2021) The road towards 6G: a comprehensive survey. IEEE Open J Commun Soc 2:334–366. https://doi.org/10.1109/OJCOMS.2021.3057679
3. Tariq F, Khandaker MRA, Wong K-K, Imran MA, Bennis M, Debbah M (2020) A speculative study on 6G. IEEE Wireless Commun Mag 27(4):118–125
4. Gui G, Liu M, Tang F, Kato N, Adachi F (2020) 6G: opening new horizons for integration of comfort, security, and intelligence. IEEE Wireless Commun Mag 27(5):126–132
5. Letaief KB, Chen W, Shi Y, Zhang J, Zhang YA (2019) The roadmap to 6G: AI empowered wireless networks. IEEE Commun Mag 57(8):84–90
6. Liu G et al (2020) Vision, requirements and network architecture of 6G mobile network beyond 2030. China Commun 17(9):92–104
7. Rikkinen K, Kyosti P, Leinonen ME, Berg M, Parssinen A (2020) THz radio communication: link budget analysis toward 6G. IEEE Commun Mag 58(11):22–27
8. Huang C et al (2020) Holographic MIMO surfaces for 6G wireless networks: opportunities, challenges, and trends. IEEE Wireless Commun Mag 27(5):118–125
9. Chi N, Zhou Y, Wei Y, Hu F (2020) Visible light communication in 6G: advances, challenges, and prospects. IEEE Veh Technol Mag 15(4):93–102

10. Kishk M, Bader A, Alouini M-S, Aerial base station deployment in 6G cellular networks using tethered drones
11. Wang C-X, Huang J, Wang H, Gao X, You X, Hao Y (2020) 6G wireless channel measurements and models: trends and challenges. IEEE Veh Technol Mag 15(4):22–32
12. Huang T, Yang W, Wu J, Ma J, Zhang X, Zhang D (2019) A survey on green 6G network: architecture and technologies. IEEE Access 7:175758–175768. https://doi.org/10.1109/ACCESS.2019.2957648
13. Bariah L et al (2020) A prospective look: key enabling technologies, applications and open research topics in 6G networks. IEEE Access 8:174792–174820
14. Polese M, Jornet JM, Melodia T, Zorzi M (2020) Toward end-to-end, full-stack 6G terahertz networks. IEEE Commun Mag 58(11):48–54
15. Chowdhury MMU, Saad W, Guvenc I (2020) Mobility management for cellular-connected UAVs: a learning-based approach. 2020 IEEE international conference on communications workshops (ICC Workshops), pp 1–6. IEEE
16. https://www.digitaltrends.com/mobile/what-is-6g/
17. https://5g-ppp.eu/wp-content/uploads/2021/06/WhitePaper-6G-Europe.pdf
18. Yang H, Alphones A, Xiong Z, Niyato D, Zhao J, Wu K (2020) Artificial-intelligence-enabled intelligent 6G networks. IEEE Netw 34(6):272–280. https://doi.org/10.1109/MNET.011.2000195
19. Xiao Y, Shi G, Li Y, Saad W, Poor HV (2020) Toward self-learning edge intelligence in 6G. IEEE Commun Mag 58(12):34–40. https://doi.org/10.1109/MCOM.001.2000388
20. Chen M, Challita U, Saad W, Yin C, Debbah M (2019) Artificial neural networks-based machine learning for wireless networks: a tutorial. IEEE Commun Surv Tutorials 21(4):3039–3071. https://doi.org/10.1109/COMST.2019.2926625
21. Yang P, Xiao Y, Xiao M, Li S (2019) 6G wireless communications: vision and potential techniques. IEEE Netw 33(4):70–75

Handoffs in Next-Generation Wireless Networks

Payal Mahajan and Zaheeruddin

Abstract Over the years, a great advancement has been seen in the field of the wireless communication. Many new access technologies have been developed along with the existing wireless access technologies. For an efficient wireless communication system, a seamless handoff from one type of access technology to another type of access technology is needed. So that a user can enjoy all the wireless access technologies' services at the same time without an interruption in the ongoing call or data session. In this paper, we have discussed about the various types of handoffs in wireless communication. A review of work done on the Handoffs is given and then a detailed LTE-Handoff procedure is explained along with the parameters of handoffs in LTE and the phases of handoffs in LTE are also explained.

Keywords LTE · Handoff · Wireless access technologies · Wireless communication · Handoff parameters

1 Introduction

Over the last 15 years, mobile data usage has increased 400-fold, and it is predicted to increase roughly 7-times by 2024, accomplishing 80 Exabytes per month by 2022. Ultra-cell-densification is a major 5G strategy for meeting the enormous data needs. The spectral efficacy of the fifth-Generation network might be greatly enhanced by decreasing the base stations (BSs) coverage area, hence lowering the no. of customers assisted by every Base Station and increasing reuse of frequency [1]. Though, increased handoff (HO) rate, i.e., repeated change of handling Base Station for a mobile user, is a visible influence of densifying design. The capacity boost is obtained with this method, albeit at the costs of higher HO frequency and signaling overhead due to the HO operation. The signaling overhead disrupts flow of data, lowering the consumers' throughput. Future cellular networks will need to handle data-hungry services with higher data rates, which could be accomplished through

P. Mahajan (✉) · Zaheeruddin
Department of Electrical Engineering, Jamia Millia Islamia, New Delhi 110025, India
e-mail: payal27mahajan@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_5

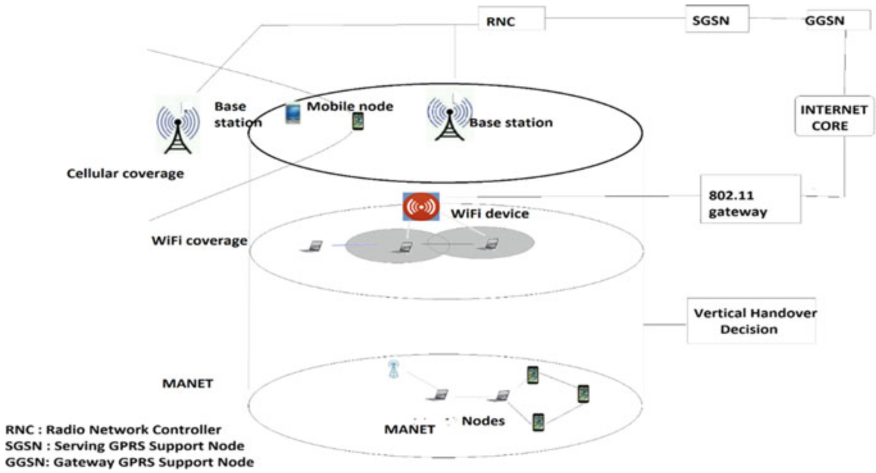


Fig. 1 Heterogeneous networks

cell increased density (small cells). It is equally vital to have consistent Handoff methods owing to increased data rates, as this has a direct effect on the end user's quality of experience (QoE) [2]. When it comes to tiny cell research, the large number of studies focus entirely on capacities and productivity assessment, with only a few help this process to HO management. Cellular networks' mobility features allow users to go anywhere within the coverage region while still being serviced (Fig. 1).

Various characteristics must be maintained and kept in mind when developing a communication network in order for it to be effective. The handover method is one of the primary regulating factors [3]. A handoff occurs in any network when an ongoing call or data session is relocated to a station other than the one it is currently on. When a user switches from one cell to the other, a handover may be needed. Handovers happen after numerous criteria such as signal quality, signal power, interference, and so on are taken into account. When a mobile user is on the go, the signal from the base station will eventually fall below a certain threshold, preventing the call from continuing [4]. The call will be disconnected at this time. However, because we use soft handovers, handoff demands are made to multiple neighboring base stations before the call is dropped. Handover Instigation, Handover Decision, and Handover Implementation are the three steps that make up the overall handover procedure. For time-division, multiple access (TDMA), code-division multiple access (CDMA), frequency-division multiple access (FDMA), or an amalgam system, the channels alteration because of handoff may be via a frequency band, time slot, code-division multiple access (CDMA) code words. There should be an uninterrupted handoff among 2G/3G/CDMA/LTE access technologies [5].

1.1 Author's Contribution

i. Detailed study of heterogeneous networks is done.
ii. Work of many authors in the similar field is highlighted.
iii. Types of Handoffs are given.
iv. Working of Handoff Algorithm in LTE is given and is explained in detail so that the researchers could get the better knowledge about the handoff working in LTE networks.

2 Literature Review

Various papers have been studied to do this review. Some of these are given as:

Wasan et al. [6], The goal of this research is to see how distinct HCP parameters influence the operation of a fifth-generation network. Based on various HCP configurations and mobile speed concepts, various system situations are offered and explored. The impact of numerous suggested framework scenarios on 5G network execution is anticipated to be demonstrated by the different phone speeds. In terms of OP, lower HCP levels offer substantial improvements when compared to higher HCP values. Similarly, lesser HCP values have notable downsides in view of maximum PPHP for all mobile speed situations as indicated by high HCP levels.

Tayyab et al. [7], presented the ideas of radio access flexibility in cellular systems in this study, along with potential problems and existing research priorities. To emphasize the fundamental variations in basic HO situations, they presented an outline of Handover handling in long-term evolution (LTE) and fifth-generation new radio (NR) in this paper. In addition, their paper discusses Handoff administration issues and techniques, as well as the essential factors that must be considered while developing an effective Handoff system.

Padmapriya et al. [8], in their work, constructed a VHO model based on the utility model, in which the handoff happens only to the appropriate cells, avoiding any issues with network access.

Mandour et al. [9], the HO choice factors in this work are focused on the Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and User Equipment (UE) factors such as position and moving direction inside the femtocell. The proposed algorithm's goal is to select the best suitable target femtocell from a large number of participants and to minimize unnecessary HO in femtocell-based cellular networks. The findings show that the suggested methodology will decrease/increase the likelihood of Handoff failure/success, accordingly.

Preeti et al. [10], investigated and analyzed the handoff process in an LTE network in this paper. Distance and speed have been deemed critical elements in this handoff process. In order to boost signal power, a comparison of path loss models is conducted. Unwanted handovers likelihood and handovers probability of failure has been investigated thoroughly. Their simulation results suggested that the efficiency of the above-mentioned characteristics in the defined network has improved.

Narvez et al. [11], provided an overview of LTE handover algorithms, with a focus on those designed for large-scale transportation. They demonstrated how the new algorithms improve handoff success while also enhancing network data flow. This suggested that characteristics like position, speed, position, and direction must be factored into procedures to enhance transportation handoff. They also discussed algorithms for mobile relays, which they believe will be a significant area of future research.

Mei et al. [12], have investigated the viability and accuracy of mobility-aware bandwidth and handoff forecasts in 4G/LTE and 5G networks in this research. They collected long continuous samples with significant bandwidth, channels, and contextual data from public transport networks in order to reach this aim. In fixed-route movement settings, they proposed Recurrent Neural Network models to exploit the temporal variations of bandwidth evolution. They proposed a novel challenge of handoff forecasting among 4G and 5G for co-existing 4G and 5G networks, which really is critical for low-latency operations as self-driving strategies in actual 5G situations.

Ahmad et al. [13], proposed an enhanced handoff technique that combines the current UE's traveling path with its information and history. To estimate the orientation of the UE, the suggested method tracks its locations. On the basis of handoff numbers, signaling measurements, packet delay ratio, packet delay ratio, packet loss ratio, and throughput, the suggested algorithm is analyzed. The suggested algorithm greatly decreased the incidence of handoffs, signaling measures, packet delay ratio, and packet loss ratio, and boosted throughput, according to the modeling with LTE-Sim.

Saeed et al. [14], proposed Fuzzy Logic-based LTE Handover optimization approach for long-term evaluation (LTE) networks in this research (FLLH). It involved employing fuzzy logic to determine optimal handoff margin (HOM) for the handoff procedures as well as the suitable time-to-trigger (TTT) for a successful handoff. The FLLH handover optimization method is tested against four well-known handoff algorithms. In comparison to the self-optimization technique, the suggested handoff optimization technique provided the lowest average of handovers per user while also having the highest throughput.

Yuansheng et al. [15], in their paper, they applied deep reinforced skills to develop long-term performance in terms of user quality of services and network capacity by modeling the handoff in RAN slicing as a Markov decision process. Simulation experiments have been used to demonstrate the accuracy of their suggested handoff strategy.

3 Types of Handoffs

Handoffs can be classified in various ways. There are:

i. Depending upon the number of connections involved.

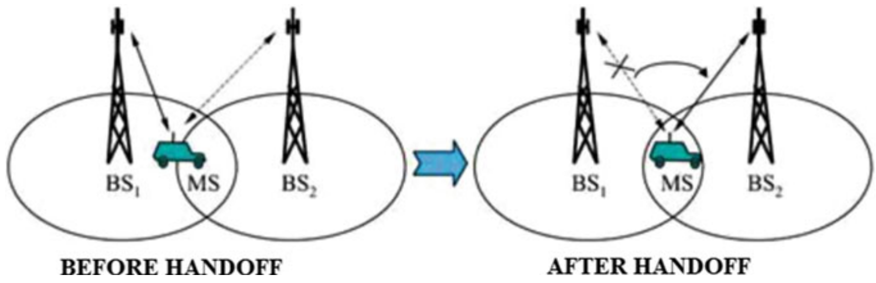


Fig. 2 Hard handoff

ii. Depending upon number of cells involved.
i. Depending upon the number of connections involved:
   On this basis, we have two types of handoffs. These are:

   a. Hard handoffs.
   b. Soft Handoffs/ Softer Handoffs.

a. **Hard Handoffs**: The serving station is withdrawn before fresh assets can be allocated in this form of handoff. As a result, there is a communication gap always because the mobile node cannot talk with both access points at the same time (old and new) [16] (Fig. 2).

b. **Soft Handoffs**: The mobile link with neighboring BS new is created first, followed by the release of old base station. Because a connection is created before breaking away from the old, this approach minimizes handoff failures. It also has a Quick Base Station Selection feature (FBSS).
c. **Softer Handoffs**: This is similar to soft handoff. The only variation is that it happens between the two sectors of same cell [17] (Fig. 3).

ii. Depending upon the number of access technologies involved:

   a. **Horizontal Handoffs:** This is also known as an intra-technology handoff, and it is a handoff between Base Stations that share the same local network, such as 3G to 3G. This is frequent in homogenous networks such as GSM and CDMA [18].
   b. **Vertical Handoffs:** This is also known as inter-technology handoff, and it happens when networks of different technologies, such as 3G and 4G, switch over. Vertical handoffs necessitate the effective completion of both layer 2 (Data Link layer) and layer 3 of the handoff procedures [19] (Fig. 4).

3.1 Handoffs in LTE

Handoff process in LTE may be categorized into three phases: preparation for handover, execution of handover, and completion of handover. The procedure begins

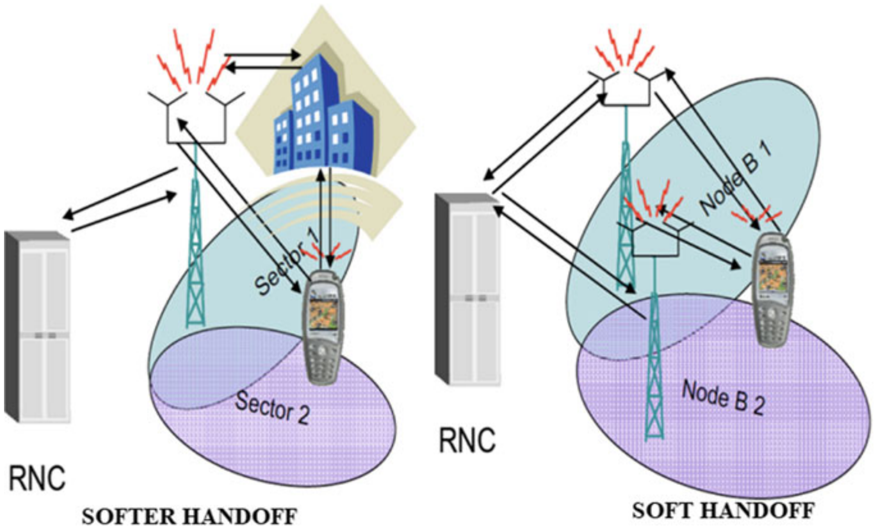


Fig. 3 Soft and softer handoffs

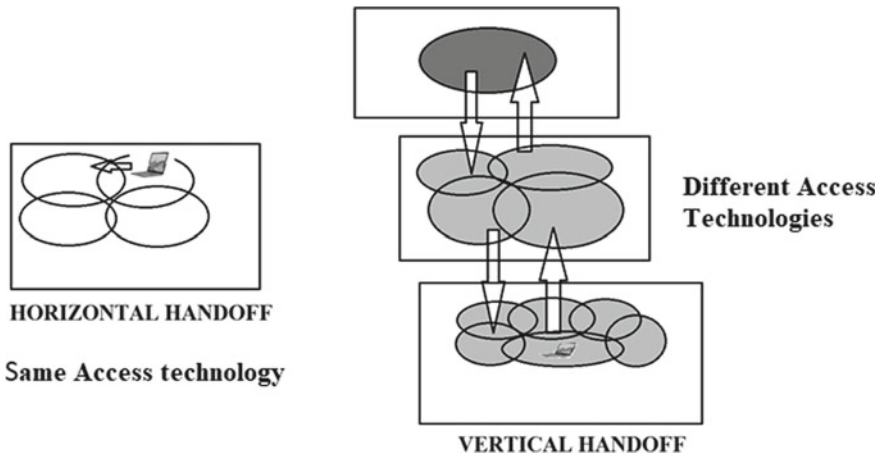


Fig. 4 Horizontal and vertical handoffs

with the User Equipment (UE) recording measurement of a handover event to the advanced serving node B (eNB). The Evolved Packet Core (EPC) does not engage in the control plane handling handover process, i.e., planning messages are shared directly between the eNB's [20]. This is the case when implementing X2 interface, otherwise, MME would be used for HO signaling. The LTE handover process begins on the basis of the measurement reports from a UE to the SeNB serving which takes the handover decisions.

The handover method has various parameters which are used to enhance its efficiency and it is a very important job to set these parameters to the optimum values. In LTE handover triggering is typically based on measuring the quality of the connection and some other metrics to improve the output [21]. Various Parameters considered for taking handover decisions in LTE are:

i. Handoff initiation threshold levels RSRQ and RSRP.
ii. Hysteresis margin.
iii. Time-to-Trigger (TTT).
iv. The dimensions and outline of averaging window.

The LTE-Handoff procedure begins, based on the reports of measurement received from the mobile node to the serving SeNB which allows the choice of handoff. The mobile node regularly tests the reference signal of all the exposed SeNBs obtained powers (RSRPs) and records a calculation showing that a handoff will be activated if a definite circumstance exists. Event A3 is caused when a neighboring cell is stronger on the basis of an assured compensation than the serving cell. The time for the mobile node to record measurements is just when the TTT timer expires. The situation is defined in this way:

$$M_n + O_{\mathrm{cn}} + O_{\mathrm{fn}} > M_p + O_{\mathrm{cp}} + O_{\mathrm{fp}} + O_{\mathrm{ff}} + \mathrm{hys}$$

where M_p and M_n are serving SeNB and neighboring SeNB's RSRP values calculated, respectively. hys is the constraint for hysteresis to avoid condition oscillation because of vanishing, and O_{ff} is the offset for this event. Here the offset for the A3 event as A3 Offset, in order to distinguish the A3 event offset from other offsets. Offsets O_{cn} and O_c are CIOs, respectively, for adjacent SeNB and the allocated SeNB [22]. For good results, the handover parameters need to be configured. Under fading conditions, too low transfer offset and TTT values result in back-and-forth Ping-Pong handoffs among the cells. Very high values may then be the reason for calls failures during handoff when the radio situations in the serving cell are too poor for transmission. Figure 5 shows the working of Handoff in LTE. The Handoff in LTE can be explained as follows:

In LTE, all handoffs are hard handoffs, indicating the source eNB's air interface link is terminated prior to re-establishing a communication link with the destination eNB. Both for X2-based and S1-based handoffs, the technique for re-establishing the access to the target node is the same, and it needs the usage of the randomized access procedures to synchronize and start upward broadcasts. The UE is usually informed of the precise random access sign to utilize in needed to execute an assertion operation. Handoff can lead to alterations to the supporting MME and S-GW in the network infrastructure, which can be related to motion and the geographic location of nodes in the network, as well as load and traffic management systems. The handoff mechanism among eNBs inside the same MME/S-GW is depicted in Fig. 5. Downlink packets on the forward path and on the new direct route may appear swapped at the target eNB after the downstream path is changed at the S-GW. Before delivering any packets obtained on the current direct route, the target eNB shall deliver all forwarded packets

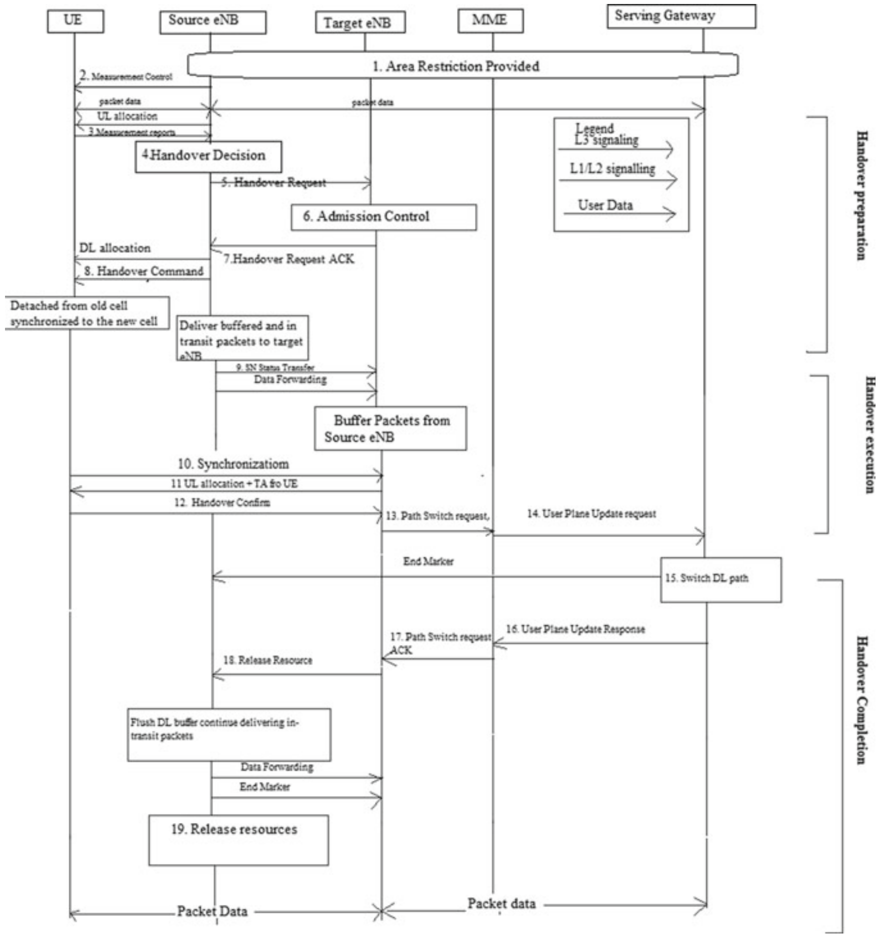


Fig. 5 Handover procedure in LTE

to the UE. Upon handoff, the serving eNB sends any downstream PDCP SDUs that have not yet been recognized by the UE to the target eNB in a logical order. Any residual downlink RLC PDUs are discarded by the serving eNB. Consequently, the downlink RLC context is not forwarded by the source eNB to the destination eNB.

The X2-based handoff procedure can be divided into four stages [23]:

1. Decision: On the basis of statistical data obtained from the UE, the E-UTRAN determines to perform a handoff.
2. Preparation: The serving eNB sends a handoff request to the target eNB. It contains a list of payloads that will be sent as well as information on whether downlink data forwarding is required. The destination eNB confirms the handoff request and replies with a list of admitted bearers as well as downlink and uplink GTP tunnel endpoints for forwarding of data.

3. Execution: The UE is alerted and dissociates from the source eNB after the assets have been set up on the target side. The source eNB forwards downlink packets acquired from the S-GW to the target eNB. If lossless handoff is necessary, PDCP status and hyper frame number data are being sent. The MME is alerted of the modifications at this point, allowing the user plane to specify immediate uplink and downlink pathways with the Serving Gateway.
4. Handoff completion: After the handoff is completed successfully, resources in the source eNB are freed.

The following are the primary stages of an S1-based handoff [11]:

1. Decision: Based on measurement reports, the E-UTRAN chose to begin the handoff and alert the serving MME.
2. Preparation: The serving MME detects the target MME and tells it, which then starts preparing the target assets. Since a shift of S-GW is also needed, the target MME begins resources planning in the target S-GW as well. The target MME notifies the source MME of the different addresses and tunnel destinations to allow data forwarding as the final prior step.
3. Execution of Handoff: The source MME starts the handoff execution phase by informing the source eNB, which then informs the User Equipment. The UE disconnects from the source eNB and uses random access techniques to synchronize with the destination eNB. The target MME starts the procedure of establishing direct uplink and downlink paths for the P-GW to the target S-GW once it has been created.
4. Handoff completion: After the handoff is completed successfully, resources in the source eNB and source MME are freed.

4 Conclusion

The handoff plays an important role in this world of heterogeneous networks so that all the wireless access technologies could work together and all the users may enjoy the uninterrupted services. In this paper, we have tried to explain various parameters that are considered for taking Handoff decisions. Further, it is mentioned that handoff plays an important role in the efficient wireless communication networks' working. In this paper, a detailed study of handoff types is given and the Handoff Procedure in LTE is explained in detail.

5 Future Work

In the present work, we have talked about the handoff types in different types of networks present in today's world. Further, explanation can be given for the handoffs in various heterogeneous networks that can help researchers to learn about the handoff procedures in Heterogeneous Networks. Also, in future we will work for the development of various self-optimizing networks so that the handoff algorithms are self-healing in nature which in return will lead to the development of the efficient handoffs in the wireless heterogeneous networks.

References

1. Madaan J, Kashyap I (2015)"An overview of vertical handoff decision algorithm. Int J Comput Appl 111(3)
2. Mamman M (2021) Call admission control algorithm with efficient handoff for both 4G and 5G networks. Int J Wireless Mobile Netw (IJWMN) 13
3. Neeraja S, Sasibhushana Rao G (2017) A comparative study on handoff algorithms for GSM and CDMA cellular networks. Int J Electric Comput Eng (2088–8708) 7(3)
4. Abolade JO, Fakolujo OA, Orimogunje A (2017) Handover in mobile wireless communication network-a review. Int J Adv Eng Manag Sci 3(9):239916
5. Abraham J, Kannampallil T, Patel VL (2014) A systematic review of the literature on the evaluation of handoff tools: implications for research and practice. J Am Med Inform Assoc 21(1):154–162
6. Saad, Kadhim W, Shayea I, Hamza BJ, Mohamad H, Daradkeh YI, Jabbar WA (2021) Handover parameters optimisation techniques in 5G Networks. Sensors 21(5):5202
7. Tayyab M, Gelabert X, Jäntti R (2019) A survey on handover management: from LTE to NR. IEEE Access 7:118907–118930. https://doi.org/10.1109/ACCESS.2019.2937405
8. Padmapriya T, Saminadan V (2016) Handoff decision for multi-user multi-class traffic in MIMO-LTE-A networks. Procedia Comput Sci 92:410–417
9. Mandour M, Gebali F, Elbayoumy AD, Abdel Hamid GM, Abdelaziz A (2019) Handover optimization and user mobility prediction in LTE femtocells network. In: 2019 IEEE international conference on consumer electronics (ICCE), pp 1–6. https://doi.org/10.1109/ICCE.2019.8662064
10. Lal P, Yamini V, Noor Mohammed V (2017) Handoff mechanisms in LTE networks. In: IOP conference series: materials science and engineering, vol 263, no 5. IOP Publishing, p 052033
11. Narváez, Trejo EA, Bonilla CMH (2018) Handover algorithms in LTE networks for massive means of transport. Sistemas y Telemática 16(46):21–36
12. Mei L, Gou J, Cai Y, Cao H, Liu Y (2021) Realtime mobile bandwidth and handoff predictions in 4G/5G Networks. arXiv preprint arXiv:2104.12959
13. Ahmad R, Sundararajan EA, Othman NE, Ismail M (2018) Efficient handover in LTE-A by using mobility pattern history and user trajectory prediction. Arabian J Sci Eng (Springer Science & Business Media BV) 43(6)
14. Saeed M, Kamal H, El-Ghoneimy M (2016) Solving LTE handover problem using a new fuzzy logic optimization technique. Int J Adv Sci Tech Res (1JAST) 6(6):119–133
15. Wu Y, Zhao G, Ni D, Junyi D (2021) Dynamic handoff policy for RAN slicing by exploiting deep reinforcement learning. EURASIP J Wirel Commun Netw 2021(1):1–17
16. Tripathi ND, Reed JH, VanLandinoham HF (1998) Handoff in cellular systems. IEEE Pers Commun 5(6):26–37
17. Abolade JO, Fakolujo OA, Orimogunje A (2017) Handover in mobile wireless communication network-a review. Int J Adv Eng Manag Sci 3(9):239916
18. Vetrivelan P, Narayanasamy P (2013) SMIRT with call admission control (CAC) based vertical handover decision for seamless mobility in multi-access 4G heterogeneous wireless overlay networks. In: Proceeding of the international multiconference of engineers & computer scientists 2012, vol 1, IMECS 2012, Hongkong, 14–16 Mar 2012
19. Wang HJ, Katz RH, Giese J (1999) Policy-enabled handovers across heterogeneous wireless networks. In: proceeding of ACM WMCSA

20. Yang J, Ai B, Salous S, Guan K, He D, Shi G, Zhong Z (2019) An efficient MIMO channel model for LTE-R network in highspeed train environment. IEEE Trans Veh Technol 68(4):3189–3200
21. Alhabo M, Zhang L, Nawaz N (2017) A trade-off between unnecessary handover and handover failure for heterogeneous networks. In: Proceeding 23rd Eur. Wireless Conference, pp 1–6
22. Zhang H, Huang W, Liu Y (2019) Handover probability analysis of anchor-based multi-connectivity in 5G user-centric network. IEEE Wireless Commun Lett 8(2):396–399
23. Ulvan A, Bestak R, Ulvan M (2013) Handover procedure and decision strategy in LTE-based femtocell network. Telecommun Syst 52(4):2733–2748

A Review: Recent Progress in Multiple Access Techniques (RSMA and IDMA) for 5G and Beyond Networks

Devendra Kumar

Abstract Nowadays, wireless communication is played prominent role in modern scenarios. To make an effective communication, it is important to bring new changes in technology. In this paper different types of multiple access technologies have been studied and find that Performance-wise and easy to access are two parameters that have received much attention. In today's era, NOMA (Non-orthogonal multiple access) technologies are giving better output as compared to other technologies for 5G and Beyond Communication network. Due to this, many types of multiple access technologies have been surveyed for the applications of NOMA. In which RSMA (Rate splitting multiple access) and IDMA (Interleave division multiple access) are the more suitable for NOMA applications. It has been noted IDMA that being a property like interleaving, IDMA distinguishes people very well. In every system there is always a trade-off exists, here the receiver-side complexity is increasing. To reduce this effect, RSMA is the better choice to minimize the receiver complexity. It has been found that RSMA increases the effective execution with easier architecture which assists with limiting execution delay. In this paper, it has been analyzed that RSMA gives better execution with low-complexity structure.

Keywords IoT · 5G · IDMA · RSMA · NOMA · Multiple access schemes

1 Introduction

Since 1980, there has been a development of one network generation in almost every decade, which covers a total of four decades (1G, 2G, 3G, 4G) and now 5G network is expected to be launched soon in 2021. In 1980, this journey was started with 1G (first generation), while over time there were many changes in it like 2G, 2.5G, 3G, and 4G generations but all of them got improved along with the change and now continuing till 5G. At present time, as you know, all of us are using the 4th network generation of wireless communication technology in our mobiles. Very few people

---

D. Kumar (✉)
Department of Electronics and Communication, GLA University, Mathura 281406, India
e-mail: devendrakumar.ec@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_6

would know about 1G generation, it is the oldest cellular network system, which was the first mobile network generation started with 1G (first generation). The 1st generation is considered the father of wireless communication technology; it was launched in the year 1980. 1G network is the first generation of wireless mobile communication, it used to work on analog signals to transmit data. The first generation mobile network had many limitations, yet it started because it made it possible to 'communicate' with each other for the first time. Could only make calls from this mobile network, there were no other features. In this, only one person could call from one person to another for a limited distance. Its internet speed capability was 2.4 kbps, it was a new technology of that generation, so these cellular phones were also expensive at that time. The full-form of 2G is called 'Second Generation' and in Hindi 'Second Generation'. 2G technology is the first digital cellular network, which was more secure than 1G with better quality voice. 2G is the second generation of mobile networks, which used digital signals for the first time. It was launched in Finland in 1991, in which GSM technology was used. The transfer data speed of 2G network was up to 64 kbps [1–3]. 2G network telephone technology introduced calls as well as internet data services such as text messages, picture messages, and picture messages, but 2G was not yet successful in terms of video calls, video conferencing, mobile television, and multimedia services. A few years after 2G technology, the new version of mobile network 2.5G and 2.75G network technology came, which brought some improvement in the speed of internet data services (GPRS and EDGE). 2.5 g began to be used for internet communication services such as Wireless Application Protocol, Multimedia Messaging Service (MMS), e-mail, and World Wide Web access. With 2.5G or GPRS (General Packet Radio Service) the data transfer speed was approximately 48–120 kbps and 2.75G which is based on EDGE (Enhanced Data GSM Environment) data transfer speed according to its theory is 384 kbps (actually 100 kbps). kbps). 2.75G or EDGE data connectivity internet speeds are faster than GPRS (2G and 2.5G) connections, but generally slower than 3G networks. 3G technology is called "3rd Generation" of wireless mobile networks. It was developed in the year 2001 [4–6]. 3G technology is specially designed for multimedia cellphones which are commonly called 'smartphones' in Hindi. High-speed bandwidth and data transmission speeds in 3G networks increased from 144 kbps to 2.05 mbps over the previous 2G generation, allowing simultaneous telephone calls, internet access, e-mail exchanges, and easy messaging. In the 3rd generation, new features were introduced in the audio and video codecs, in which better voice calling quality was given. Video calls, TV streaming, 3D games, large files transfer, multimedia applications, YouTube, high data speed, and many more features were launched in the 3G generation. Only after the arrival of 3G, companies also started new data plans. Remember this, the camera with which we take Selfie from the mobile is the 3G generation itself due to the front-facing camera. 3G network is also known as IMT-2000. 3.5G and 3.75G are HSPA and HSPA+, respectively. Whose data speeds provide 14 Mbps and 168 Mbps, respectively [7–9]. Using 3.75G MIMO (Multiple Input and Multiple Output) enabled data access at high speed and low latency. 4G network is the 4th generation wireless communication technology of mobile. 4G wireless technology is called "fourth generation" of mobile networks in Hindi, which started in the year 2011. 4G network data speeds are several times higher than 3G, with a maximum network speed of 100 mbps/sec on high-mobility communications (such as by cars and trains) and 1 Gbps/sec on low-mobility communications (such as pedestrian or stationary). Huh. The 4th generation has many features similar to the 3G generation, as it allows you to use higher data speeds to browse the internet, play online games, download and stream videos, but all of this is much faster than 3G networks. The 4G bandwidth speed is 200 mbps more than the 3G network, which takes less than about 5–6 minutes to download a high-quality movie. In the beginning, 4G internet data plans were more expensive, but after the arrival of jio SIM, people started using 4G network very fast. 4G technology is used in two ways, one is LTE and the other is VoLTE—LTE (Long Term Evolution). 5G technology is the upcoming 5th generation of wireless mobile networks, which is going to run faster than the current 4G LTE (Long Term Evolution) standard. Its connectivity, speed, voice quality, security and other features will be better than all the previous generations [10–14]. 5G network is a complete wireless communication technology without any limitation, it supports wwww (wireless world wide web) the most. 5G technology is expected to be launched after the beginning of 2020 but there are still many areas in the existing 4G wireless network system in which 4G network still needs to improve. Now in the coming times, 5G networks are going to affect every industry, due to which safe transportation, remote healthcare, agricultural service, digitized logistics, and more areas will also prove to be a successful one. The development of 5G will happen through the route of 4G, most of the telecom companies will not move directly from 4G system to 5G but from 4G to 4.5G to 4.5G Pro and then finally to 5G. As is well noted, 5G is set to begin, and 3GPP, URLLC, and other activities are required, and that is why NOMA is utilized in 5G. So, let's think regarding NOMA [11–16]. Non-Orthogonal Multiple schemes are the abbreviation for Non-Orthogonal Multiple Access. Subscribers are still not obligated to maintain symmetry throughout this approach. The various Multiple schemes approach for NOMA scenarios, RSMA and IDMA, will be studied and compared in this research.

The remaining of the manuscript are organized as follows: The second segment includes an introduction to IDMA, while section three examines the novel approach RSMA. A comparable examination of RSMA and IDMA is also included in the fourth part. Lastly, at the article's conclusion.

2 Interleave Division Multiple Access (IDMA)

TDMA or FDMA was highly widespread and convenient to use in previous systems, so after some time, advancements in technology and methods were seen. As a result, experts devise a unique Multiple Access system known as CDMA. It has numerous advantageous and outstanding properties, such as (a) frequencies spectral efficiency (b) robustness [17].

After subscribers were increasingly enraged, it became a crucial required choice to discontinue CDMA [18]. With this in mind, IDMA, a unique multiple access protocol evolved on CDMA, was developed to overcome CDMA's limitations. The most appealing feature of IDMA is how it can identify numerous subscribers by employing distinct interleaves, allowing each subscriber to obtain or recognize its respective knowledge with their own time. As per the findings, the key reason for utilizing IDMA for NOMA is its low code rate. The IDMA theory is distinguished by the fact that it will not perceive observable interaction as noise signal. IDMA, on the other hand, failed to generate significant enthusiasm for installation and acceptance [18]. IDMA's structure is depicted in the diagram under. If we look at the transmission end, the encoding will be the first section, then variation (depending on user-specified parameters), and last network (where use interleaving or scrambling to differentiate the users). The overall detecting strength is calculated as follows.

$$R(k) = G_x D_x(k) + \beta_x(k) \tag{1}$$

where

$$\beta_x(k) = \sum_{x' \neq x} G_{x'} D_{x'}(k) + n(k) \tag{2}$$

showing the phrase "blockage" for x subscriber $D_x(k)$ represents the bits of transmissions, $G_x(k)$ the medium, and $n(k)$ the disturbance at the kth interval. The elementary signal estimator (ESE) is indeed the initial module on the receiver station, giving information on the a priori log-likelihood ratio (LLR) [19], followed by decoder or de-interleaving. According to just the identification of transferred bitstream, the turbo-like loop of upgrades is repeated hundreds of times throughout the subscribers. Furthermore, it must be said that IDMA's framework is overly complicated. There are many various mechanisms that require an excessive amount of durations. As a result, the delay time has increased. i.e., RSMA is likewise recognized in this work as having better performance with little delay when opposed to IDMA (Fig. 1).

3 Rate Splitting Multiple Access

A rising Wireless Communication framework is expected for employing 5G and well beyond communications infrastructure in today's modern technologies. RSMA is a multiple schemes protocol that is both powerful and successful. Linear Precoder (LP) on the transmission end and Successive Interference Cancellation (SIC) on the receiver end was employed in RSMA. Interestingly, there are two main types of RSMA messages: secret and common, with the common information being accessed by x subscribers and the secret information being accessed by just the principal subscriber [20]. Each subscriber will have partially decrypted the obstacle. This feature of RSMA makes it safer and much more convenient versus IDMA. Although SIC is employed on the received signal, there is really no possibility of loss of performance, but it is also less sophisticated than IDMA. In the same way, IDMA

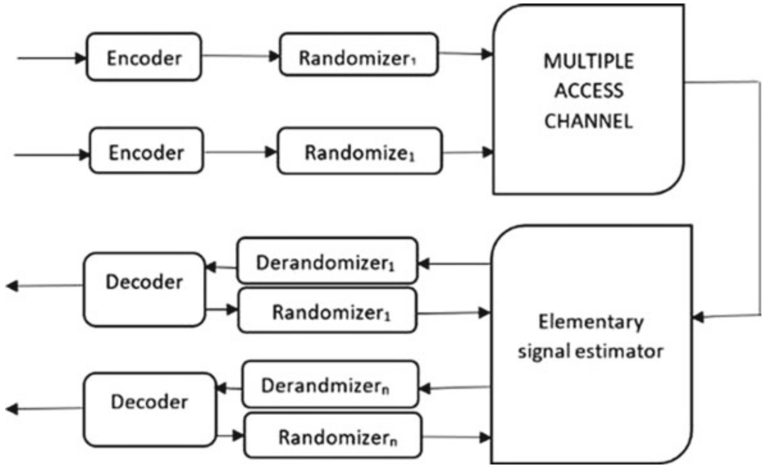


Fig. 1 Architecture of IDMA approach

requires knowledge of RSMA's framework. Figure 2 shows that RSMA framework is less complicated than IDMA framework since ESE is not required [21]. (it means that matched of subscribers). The transmission end of something like the RSMA design is identical to the transmission end of IDMA. Finally, at the recipient's end,

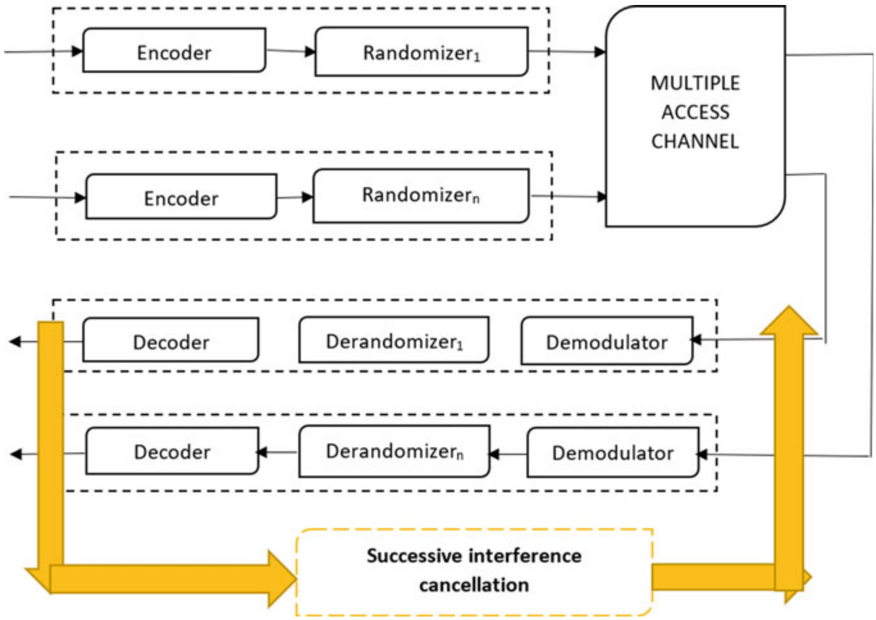


Fig. 2 Architecture of RSMA approach with SIC

decode the entire data, which is then decoded developed a good understanding of a decoding.

As a result, it is a much easier procedure than the IDMA recipient procedure. The main difference between these two topologies is in terms of SIC. The strategy's quality will improve only with support of SIC. In method [22], the latency will likewise be low due to the little complication.

Hence, the WSR (weighted sum rate) for two users if given by:

$$\mathrm{R}_{\mathrm{RS(u)}} = \max u_1 R_1 + u_2 R_2 \tag{3}$$

$$C^{12} + C^{12} \leq R_{12} \tag{4}$$

$$R_K \geq R_K^{\mathrm{th}}, k \in \{1, 2\} \tag{5}$$

$$C \geq 0 \tag{6}$$

where $C = \left[c_1^{12}, c_2^{12}\right]$ is the common rate.

It has already been demonstrated that both strategies have benefits and downsides. As a result, comparing various multiple access methods is becoming more important.

4 Comparison of IDMA and RSMA

Almost all MA approaches are necessary for high techniques, as it could be illustrated [18]. To begin, a bar graph was used to demonstrate that NOMA produces superior results for 5G than OMA. Unless a result of the bar graph, this can be concluded that as the percentage of subscribers grows, the average sum rates of NOMA exceed that of OMA [23] (Fig. 3).

Otherwise, we'll compare the two more influential strategies for NOMA solutions in this part. However, applications in present era require great efficiency and a low latency technique. As a result, it will be demonstrated that RSMA MA performs better than IDMA MA. The RSMA is faster than the IDMA in terms of difficulty, as previously stated. Furthermore, using the figure, it would be demonstrated that WSR vs SNR of RSMA obtains superior improvement over IDMA [19] (Fig. 4).

The comparing pattern of RSMA, RSMA-SIC, and IDMA for AWGN performances is shown in Figure 8. Obviously, this figure has been seen following comparing the results of Figs. 5, 6, and 7. This pattern shows because when BER grows, so does SNR with both types of multiple access. Every MA is producing amazing results, but when compared, the RSMA MA is outperforming the others. Their goal BER for a ten-user network is 10-3, and it achieves 5 dB for RSMA-SIC and 7.8 dB for IDMA, which is also an amazing achievement for NOMA workloads [18]. In the use of the preceding outcomes, it is apparent that RSMA efficiency is

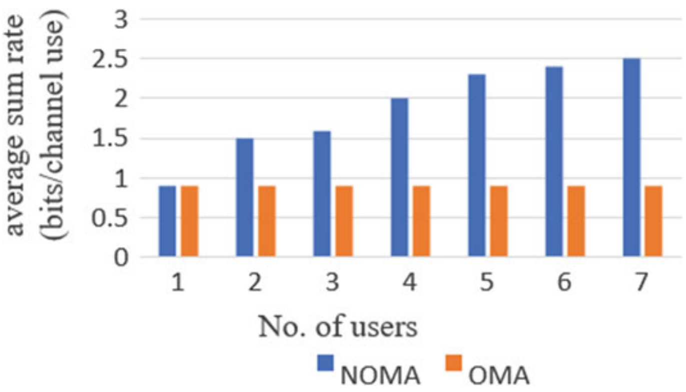


Fig. 3 Average sum rate versus number of users for NOMA and OMA

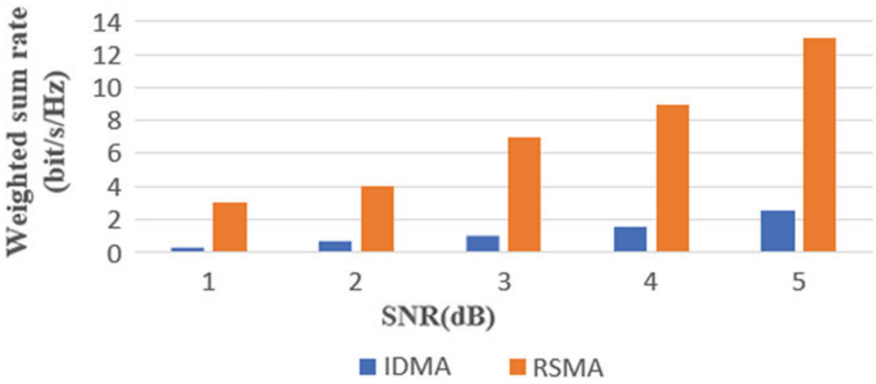


Fig. 4 Weighted sum rate versus SNR Comparison of different schemes (IDMA and RSMA)

superior to IDMA efficiency. The description of how IDMA differs from RSMA MA [24] is presented in Table 1. It captures full infrastructure details for RSMA and IDMA. Throughout this table, it has been shown that in IDMA, the shared feature engineering is employed, that will not accommodate suspect individuals.

RSMA, from the other side, provides assistance for selected individuals, which really is useful when selecting treatment.

Let's really discuss the receivers including both MA, considering we realize that perhaps the transmitters structure from both MA is the same, but the receiving structure of IDMA is dissimilar. So, as we have seen, SIC has been used in RSMA though not in IDMA. Among the most significant differences is that IDMA is entirely directly proportional to the number of subscribers present, and subscribers are assigned at randomness. However, in RSMA, subscribers must retain their lines, or number of subscribers. The third component is that in IDMA and RSMA, the demodulator and ESE blocks are utilized to rebuild the subscribers' cycles. If there are

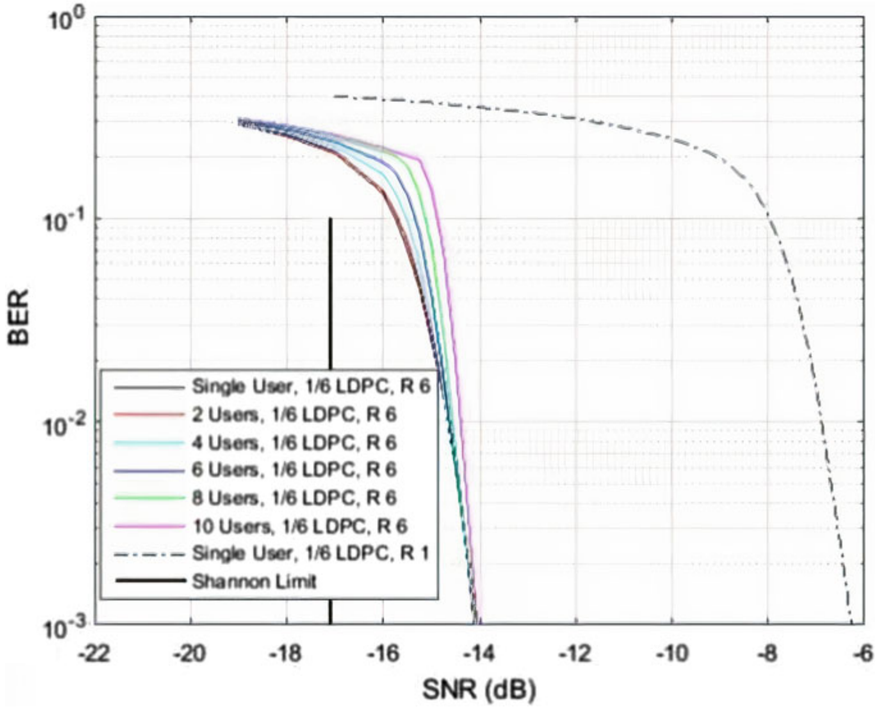


Fig. 5 AWGN performance of IDMA MAS

x subscribers and *n* LDPCs (low-density integrity checks), IDMA will be utilized for *x*n* decode, whereas RSMA would only use single LDPC. This is yet another advantage of RSMA, demonstrating that this was the more effective MA (Table 2).

5 Conclusion

The two alternative MA protocols, IDMA and RSMA, were studied in this article for NOMA scenarios. It is concluded that RSMA would be a more straightforward and reliable MA strategy than IDMA. Moreover, while IDMA delivers satisfactory accuracy, RSMA is simpler in terms of capability, with the same quality and reduced latency. In comparison to IDMA, the SNR of RSMA produces comparatively better results. Eventually, only with help of this research, we can say whether RSMA is a preferable MA technique for NOMA scenarios than every other popular MA technique.

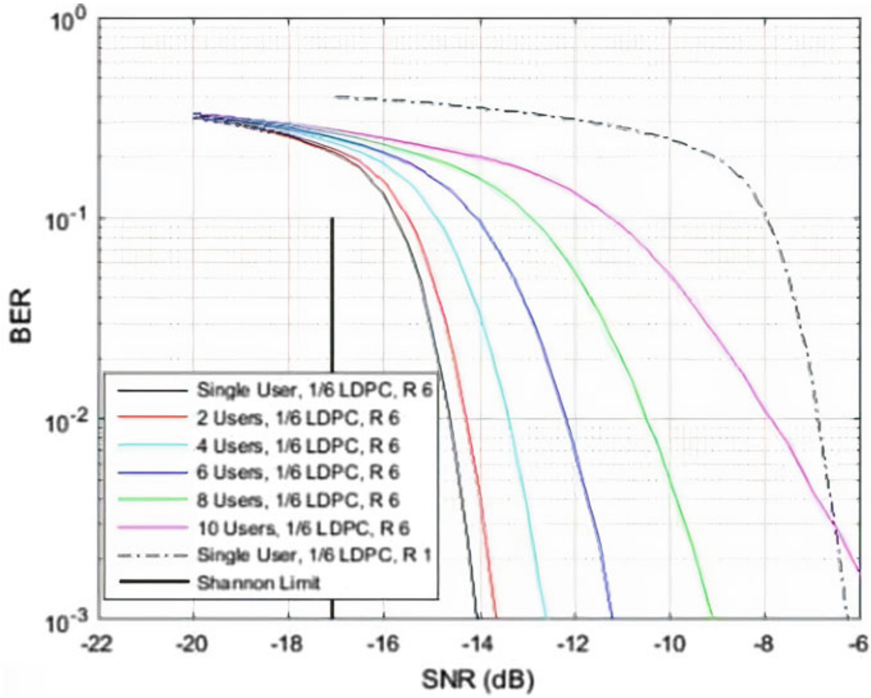


Fig. 6 AWGN performance of RSMA MAS

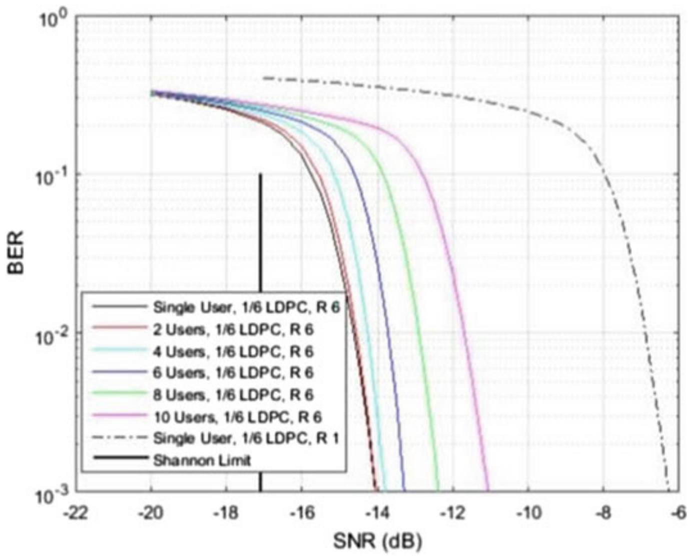


Fig. 7 AWGN performance of proposed RSMA-SIC MAS

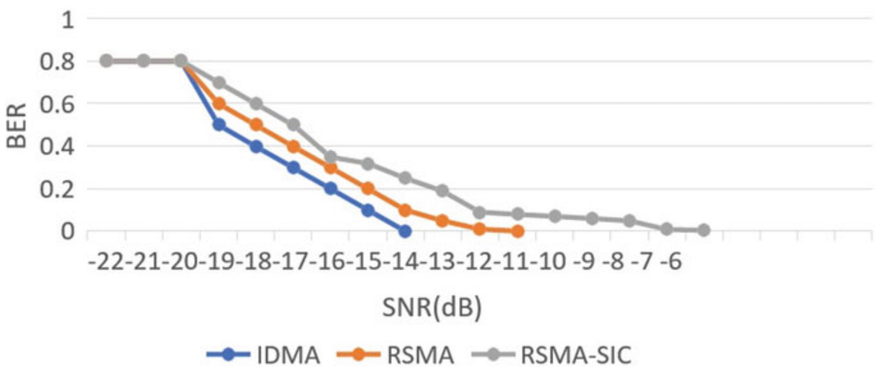


Fig. 8 AWGN performance comparison of RSMA and IDMA

Table 1 Differentiation of IDMA and RSMA for NOMA applications

Parameters	IDMA	RSMA
Complexity	Higher	Lower
Processing delay	Higher	Lower
Architecture	Complex	Simpler
Use of Mechanism	Joint processing mechanism	Single arbitrary mechanism

Table 2 Comparison of RSMA receiver and IDMA receiver

Parameters	IDMA receiver	RSMA receiver
SIC used	No	Yes
Processing dependency	Depend upon number of users and turbo processing	Only depend upon number of users in order manner
Use of demodulator	No	Yes
LDPC decoding used	It depends upon the number of users	Only one LDPC decoding used
Suitable environment	Suitable for both asynchronous and synchronous	Suitable for synchronous only

References

1. Mao Y, Clerckx B, Li VOK (2018) Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming SDMA and NOMA. EURASIP J Wireless Commun Netw 1:133
2. Mao Y, Clerckx B, Li VOK (2018) Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission. In: 2018 IEEE 19th international workshop on signal processing advances in wireless communications (SPAWC). IEEE, pp 1–5
3. Moltafet M, Mokari N, Javan MR, Saeedi H, Pishro-Nik H (2018) A new multiple access technique for 5G: power domain sparse code multiple access (PSMA). IEEE Access 6:747–759. https://doi.org/10.1109/ACCESS.2017.2775338

4. Shukla A, Deolia VK (2017) Performance analysis of modified tent map interleaver in IDMA systems. J Electric Eng 68(4):318–321
5. Akbil B, Aboutajdine D (2015) Improved IDMA for multiple access of 5G. Int J Commun Netw Inf Secur (IJCNIS) 7(3):138–146
6. Hao C, Clerckx B (2017) MISO networks with imperfect CSIT: a topological rate-splitting approach. IEEE Trans Commun 65(5):2164–2179
7. Zhu J, Wang J, Huang Y, He S, You X, Yang L (2017) On optimal power allocation for downlink non-orthogonal multiple access systems. IEEE J On Select Areas Commun 35(12)
8. Su X, Yu HF, Kim W, Choi C, Choi D (2016) Interference cancellation for non- orthogonal multiple access used in future wireless mobile networks; Su et al. (2016) EURASIP J Wireless Commun Netw 2016:231. https://doi.org/10.1186/s13638-016-0732-z
9. Zhang Y, Wang HM, Zheng TX, Yang Q (2016) Energy-efficient transmission design in non-orthogonal multiple access. IEEE Trans Veh Technol 66(3):2852–2857
10. Joudeh H, Clerckx B (2016) Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: a rate-splitting approach. IEEE Trans Commun 64(11):4847–4861
11. Ding Z, Adachi F, Vincent Poor H (2015) The application of MIMO to non-orthogonal multiple access. IEEE Trans Wireless Commun 15(1):537–552
12. Zhao M, Ma M, Yang Y (2010) Efficient data gathering with mobile collectors and space-division multiple access technique in wireless sensor networks. IEEE Trans Comput 60(3):400–417
13. Ping L et al (2003) Interleave-division multiple-access (IDMA) communications. In: Proceeding 3rd international symposium on turbo codes and related topics
14. Hao C, Clerckx B (2017) MISO networks with imperfect CSIT: a topological rate-splitting approach. IEEE Trans Commun 65(5):2164–2179
15. Dai M, Clerckx B, Gesbert D, Caire G (2016) A rate split-ting strategy for massive MIMO with imperfect CSIT. IEEE Trans Wireless Commun 15(7):4611–4624
16. Joudeh H, Clerckx B (2016) Sum-rate maximization for linearly pre-coded downlink multiuser MISO systems with partial CSIT: a rate-splitting approach. zzzIEEE Trans Commun 64(11):4847–4861
17. Dai M, Clerckx B, Gesbert D, Caire G (2016) A rate splitting strategy for massive MIMO with imperfect CSIT. IEEE Trans Wireless Commun 15(7):4611–4624
18. Kumar H, Tomar VK, A review on performance evaluation of different low power SRAM cells in nano-scale Era. Wireless Personal Commun. https://doi.org/10.1007/s11277-020-07953-4
19. Kumar H, Tomar VK (2020) Design of low power with expanded noise margin subthreshold 12T SRAM cell for ultra-low power devices. J Circ Syst Comput. https://doi.org/10.1142/S0218126621501061
20. Mao Y, Clerckx B, Li VOK (2018) Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA. In: 2018 15th international symposium on wireless communication systems (ISWCS). IEEE, pp 1–5
21. Dai L, Wang B, Yuan Y, Han S, Chih-Lin I, Wang Z (2015) Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. IEEE Commun Mag 53(9):74–81
22. Kumar HK, Tomar VK (2019) Stability analysis of sub-threshold 6T SRAM cell at 45 nm for IoT application. Int J Recent Technol Eng (IJRTE) 8(2):2432–2438
23. Shukla A (2020) Optimal multiple access scheme for 5G and beyond communication network. In: International conference on information and communication technology for intelligent systems. Springer, Singapore
24. Kumar H, Tomar VK (2019) Single bit 7T subthreshold SRAM cell for ultra- low power applications. Int J Adv Sci Technol 28(16):345–351

A New Model for Real-Time Intrusion Prevention Systems for DDoS Attacks

Mohammed Nadir Bin Ali, Mohamed Emran Hossain, Touhid Bhuiyan, Mohammed Shamsul Hoque, and J. Karthikeyan

Abstract Nowadays the internet has made a momentous impact on our daily life but we are not safe enough in the internet world. Last two decades, network security scholars have shown several innovative and practical solutions to save us from network and internet attacks. Among all the internet threats denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks are considered the most notorious and devastating ones. These attacks are one of the main threats that are a serious security problem for today's internet. To exhaust the resources of target networks, these attacks are launched by generating a huge amount of network traffic. This study proposes a new model for real-time intrusion prevention systems for DDoS attacks. It is true that creative attackers are continuously developing effective attacking tools and techniques to impose maximum damage due to the rapid technological advancement. The proposed Efficient Detection System of Network Intrusion (EDSONI) model makes use of both the detection and prevention of this malicious activity properly. CICIDS2017 dataset has been applied to this proposed system to experiment with the detection and prevention performance.

Keywords DDoS · EDSONI · Attack · Detection · And prevention

M. N. B. Ali (✉) · M. E. Hossain · T. Bhuiyan · M. S. Hoque
Daffodil International University, Dhaka, Bangladesh
e-mail: it@daffodilvarsity.edu.bd

M. E. Hossain
e-mail: emran@daffodilvarsity.edu.bd

T. Bhuiyan
e-mail: tauhidt.bhuiyan@daffodilvarsity.edu.bd

M. S. Hoque
e-mail: hoque.eng@daffodilvarsity.edu.bd

J. Karthikeyan
Vellore Institute of Technology, Vellore, India
e-mail: sigashanmu@svcet.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_7

1 Introduction

Nowadays, networks and internet have had a noteworthy effect on our day-to-day life. We are fully dependent on internet for sharing confidential and valuable information. Then again, on account of high dependence on the internet, hackers always investigate the shortcomings of the internet to paralyze the targeted servers or devices. Using this weakness, attackers always want to gain illegal access to damage targeted resources. In the world of computing, different security domains exist, and each one addresses various aspects of security [1]. Quite a few firewalls and cyber security systems are in action these days but they are not safe enough from cyber-attacks. Different cyber security systems are continuously implementing their defense mechanisms but dangerous attacks like zero-day are on the attack more dangerously on a daily basis. DDoS (Distributed Denial-of-Service) attacks are known to be the strongest and most destructive ones.

1.1 DoS and DDoS Attacks Overview

There is plenty of evidence on the DoS and DDoS attacks on the internet nowadays. DoS attacks typically flood servers and networks and practically perturb their victim's resources. On the other hand, a DDoS security attack uses a lot of compromised computers to attack a targeted traffic, slowing the computer and its network connection to a halt. This is a very powerful many-to-one technique, which is difficult to prevent. When a DDoS attack strikes, all users of a website or other online resources are completely denied access, halting the operations of the victim organization websites in its tracks. A typical attack scenario involves a good number of login attempts or calls to a website or server. The huge volume of requests floods the targeted resource, which loses the ability to tend to its legitimate users. There are three basic categories of attack (Table 1).

Table 1 Different categories of attack [2]

S No.	Categories	Description
1	Volume-based attacks	This attack uses high traffic to flood the network bandwidth. Example-ICMP flood
2	Protocol attacks	Protocol attacks are directed to servers, Routers, and Firewall, e.g., SYN flood and Smurf attacks
3	Application attacks	These attacks target web applications and are regarded as the most modern and fatal attacks. This is GET floor for http, DNS, SMTP, and SSL

1.2 Some Common Types of DDoS Attacks

DDoS attack has two types [3]: one, the bandwidth exhaustion create abstraction leading to the breakdown of network; two, resource exhaustion that exhausts key resources, memories, etc. Leading to the breakdown of the server [4]. There are different vectors of DDoS attacks, which aim to overwhelm the servers, firewalls, and different devices that vary destructively (Tables 2 and 3).

Table 2 Well-known DDoS attacks

Attacks	Description
SYN flood	SYN flood is a type of denial-of-service attack in networks where attackers continuously send SYN requests to the target network server to make it unresponsive and over consumed. This type of attack sends spoofed messages and is able to shut down the server services
UDP flood	In UDP flood attack, the target server receives a huge number of fake UDP packets sent by the attacker with the aim of occupying the whole resources
HTTP flood	HTTP flood is an HTTP Distributed Denial-of-Service attack method where the attackers exploit target web server and application by sending seemingly-legitimate session-based sets of an HTTP POST or GET request
Ping of death	By sending malicious pings to a system ping of death influences IP protocols. It was a well-known DDoS attack and nowadays it is less effective in network
Smurf attack	Smurf is a malware program that exploits by spoofing IP and pings using ICMP in a network
Slowloris	Slowloris is a type of denial-of-service attack distinctly different from other types. It uses legitimate HTTP traffics perfectly by opening and maintaining several simultaneous HTTP connections between the target server and the attacker. Thus, it makes overwhelms the target server
Application-level attacks	Application-level attack refers to comprise seemingly legitimate and innocent requests to the target server application by exploiting application vulnerabilities where the goal of this attack is to overwhelm the target application with requests
Botnet	Botnet refers to a group of malware-infected computer devices controlled by attackers with the aim to perform DDoS attacks, command and control, steal data, and so on
Zero-day DoS	Zero-day refers to a type of attack that is totally unknown to the vendors. This attack uses previously unknown flaws to exploit target network or system. Zero-day DoS attacks amplify malicious network traffic and prevent legitimate users to use network resources

Table 3 Last four years some well-known DDoS attacks

Year	Target	Brief description
2018	GitHub engineering	Due to DDoS attacks, GitHub.com was unavailable on February 28 from 17:21 to 17:26 UTC [5]
2017	Melbourne IT	On April 13 domain name registrar Melbourne IT, two of its subsidiaries TPP Wholesale and Netregistry suffered a DDoS attack. [6]
	UK national lottery	UK National Lottery www.national-lottery.co.uk and its mobile app were unavailable due to DDoS attacks on local time 23:00 to 03:00, September 30 [7]
2016	HSBC internet banking	HSBC's internet banking service in United Kingdom's was unavailable for several hours in January 2016 due to DDoS attack [8]
	Bank of Greece website	Website server of Bank of Greece was remain offline for more than 6 h under a series of DDoS attacks [9]
2015	Canadian government websites	DDoS attack affects Government of Canada website which impacted internet access, email, and information technology assets on June 17, 2015 [10]
	BBC websites	All the BBC's websites were unavailable early on Thursday (31st December) morning because of a large web attack [11]

1.3 DDoS Prevention Techniques

It is difficult to stop DDoS attacks completely. We can mitigate in tolerant level by developing and using different techniques. There are many DDoS defense mechanisms that try to prevent systems from DDoS attacks:

- Should disable unused port and services.
- Should install latest security patches.
- Should disable IP broadcast.
- Should use Firewalls and Routers.
- Ingress/Egress filtering [12].
- Router-based packet filtering.
- Load balancing.

1.4 Challenges of DDoS Attacks

The dangerous history of DDoS attacks shows the prevention of DDoS is nearly impossible. DDoS Protection is a challenge; the reasons are:

Table 4 Comparison of different parameters

Authors	Year	Attacks	Defense	Challenges	Recommendation
Shruthi [13]	2017	Yes	Yes	No	No
Vinko et al. [14]	2017	Yes	Yes	Yes	No
This paper	2018	Yes	Yes	Yes	Yes

- Difficulty in Managing Legitimate and Attack Traffic.
- Lack of large-scale testing approaches.
- Lack of infrastructure and expertise.
- Lack of effective traffic analysis and defense system.
- There are no common characteristics of DDoS.

2 Related Works

A deep study of the existing literature on DDoS architecture, detection, and prevention issues was assumed earlier to the introduction of the EDSONI model on DDoS attacks. In order to find out where further improvement could be made, relevant literature was reviewed. The study revealed that the threat implications should be the initial consideration in producing improved better detection and prevention techniques on DDoS attacks. Table 4 shows the comparison of the few existing relevant papers with our work.

2.1 DDoS Attack Tools and Their Comparison

There are several tools that are able to produce legal traffic and attack traffic [15]. Researchers hardly paid attention to the fact that DDoS use botnets to launch any attacks (Table 5).

3 EDSONI Research Methodology

EDSONI research methodology including detailed explanations of lab architecture, Dataset, hardware, and software are used for finding expected results are presented below.

Table 5 Comparison of DDoS attack tools

Year	Tools	Target impact	Type of attack	OS supported	Number of Zombies	IP spoofing	Attack model
2011	Aldi Botnet [16]	Resources	http, TCP	Widows	Multiple	Yes	web base
	SSL DoS [17]	Resources	TCP	Widows, Linux	Single	–	–
2012	GoldenEye [18]	Resources	TCP	Windows, Linux	Single	No	–
	HOLC [19]	Resources	TCP	Windows	Multiple	No	–
	HULK [20]	Resources	TCP	Windows, Linux	Single	No	–
2015	Advanced UDP attack tool	Bandwidth	UDP	Windows	–	Yes	IRC based
2016	LOIC-IFC	Bandwidth	HTTP flooding	Linux	Multiple	Yes	Agent base

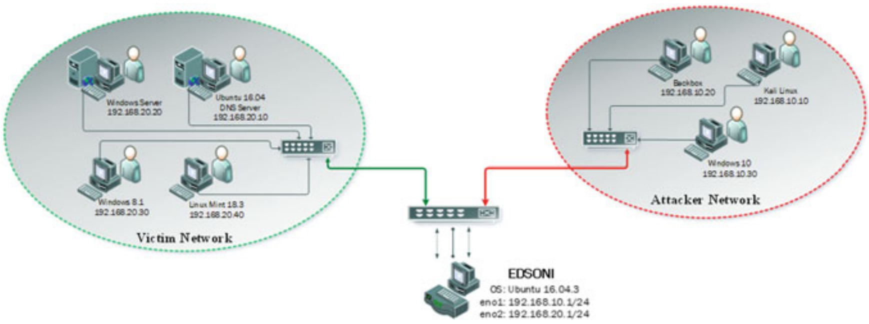


Fig. 1 Lab architecture

3.1 *Lab Architecture*

See (Fig. 1; Table 6).

3.2 *Dataset*

For dataset, we have used CICIDS2017 for network security and intrusion detection as it has diverse set of attacks. In this dataset, seven attack profiles based on the last updated list of common DDoS attack families have been created. The dataset is executed by using related tools and codes (Table 7).

Table 6 Hardware and software are used in lab architecture

	Devices	OS	IP
Victim network	Servers	Ubuntu16.04(DNS server)	192.168.20.10
		Web server (Windows)	192.168.20.20
	PCs	Linux Mint 18.3(64-bit)	192.168.20.30
		Windows 8.1(64-bit)	192.168.20.40
EDSONI	PC (Router)	Ubuntu16.04.3(64-bit)	192.168.10.1 192.168.20.1
		Scapy, Yara v3.7.0, Scapy-python3 v0.25, Yara-python v3.7.0	
Attacker network	PCs	Kali Linux(64-bit)	192.168.10.10
		Backbox 5.1(64-bit)	192.168.10.20
		Windows10(64-bit)	192.168.10.30

4 Findings

The details of the proposed enhanced Efficient Detection System of Network Intrusion (EDSONI) model are discussed below (Fig. 2).

i. Sniffer

The very first step of EDSONI is Sniffer. In this system, the "raw socket" has used for sniffing the packets from specific interface. Here all packets are saved in a file with PCAP extension that is the reason it is easy to analyze traffic for further analysis.

ii. Extractor

Generally, when a raw packet is sniffed, it contains information that is padded with both signed and unsigned bits and characters. In extractor step from every raw packet, EDSONI extractor unpack and collect all possible information from the packet. After extracting, it is easy to get version, Internet Header Length, Type of Service, length of the packet, identification tag, flag, fragment offset, TTL, protocol, Header Checksum, source IP address, destination IP address, source MAC address, destination MAC address, and Payload, etc. of the packet.

iii. Normalizer

In this step, normalizer normalizes all the data that were collected from extractor step for post-processing. So that the module selector engine and other post-processing steps can process them easily.

iv. *Module Selector*

In Module Selection Engine, it selects the correct module such as TCP, UDP, ICMP, and other modules based on the analysis of PDU (Protocol Data Unit) and Packet Header.

v. Pre-processor

Table 7 Similarities and differences between CICIDS2017 dataset and public datasets based on last IDS dataset evaluation framework [21]

	Network	Traffic	Label	Interact.	Captu.	Protocols					Attack diversity					
						http	https	SSH	FTP	Email	Browser	Bforce	DoS	Scan	Bdoor	DNS
DARPA	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No
KDD99	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No
DEFCON	No	No	No	Yes	Yes	Yes	No	Yes	No	No	No	No	No	Yes	Yes	No
CAIDAs	Yes	Yes	No	No	No	–	–	–	–	–	No	No	Yes	Yes	No	Yes
LBNL	Yes	Yes	No	No	No	Yes	No	Yes	No	No	–	–	–	Yes	–	–
CDX	No	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes
KYOTO	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TWENTE	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes	No	Yes	No	No
UNASS	Yes	No	Yes	No	Yes	Yes	No	No	No	No	No	No	No	No	No	No
ISCX 2012	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
ADFA 2013	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No
CICIDS 2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

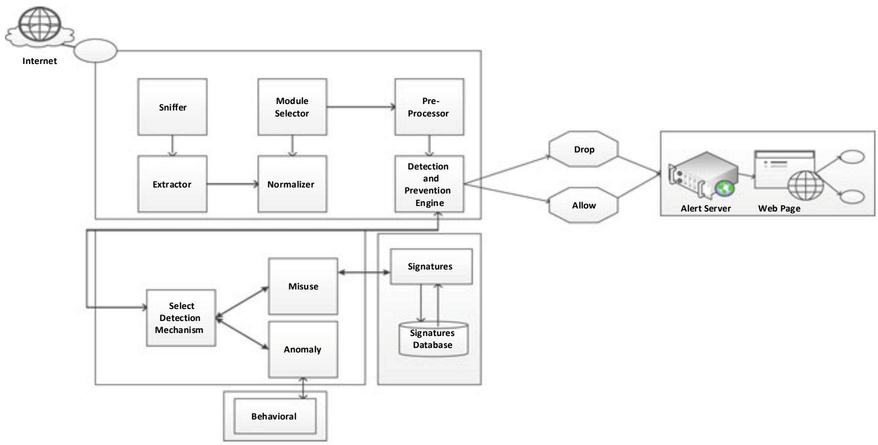


Fig. 2 Proposed EDSONI model

Pre-processor is used for examination of packets and detection of suspicious activity, modification of packets so that the detection engine can properly interpret them. It analyzes sessions, payload, fragments, segments under particular module which saves the processing time. After that traffic will go to detection engine.

vi. Detection and Prevention Engine

After pre-processor detection engine triggers a detection process in real-time with the help of selection of the detection mechanism. It is stated that EDSONI works in two modes. These are detection and prevention modes. In detection mode it only detects the intrusion, logs the information, and forwards the packet to the prevention mode.

vii. Logs and visualization

Finally, visualize the results and logs of the detection and prevention.

4.1 Implementation and Experimental Result

Table 8 shows the experimental results of evaluation metrics in terms of weighted average for the six selected machine-learning algorithms derived from generated dataset. Execution time for the testing process is also calculated and shown in the table. Among six applied machine-learning algorithms Adaboost, J48, K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Naïve Bayes, and Random Forest (RF), we have observed that based on the execution, MLP the slowest one requires 2836 s with 97.689% classification accuracy rate and on the contrary, the fastest one KNN requires only 1.2 s with 97.397% classification accuracy rate. Additionally, based on the weighted average of three evaluation metrics (Pr, Rc, and F-Measure), the highest accuracy rate refers to Naïve Bayes, Adaboost, and J48 algorithms.

Table 8 Experimental performance result of 6 machine learning algorithms

Algorithm	Testing dulii size	Accuracy	TPR	FPR	Pr	Re	F-measure	Exec. time (sec)
Adaboost	3523 1	97.779	0.978	0.016	0.983	0.978	0.979	128.4
J48	3523 1	97.745	0.977	0.018	0.982	0.977	0.979	23.4
KNN	3523 1	97.397	0.974	0.027	0.977	0.974	0.975	1.2
MLP	35231	97.689	0.977	0.034	0.977	0.977	0.975	2836
Naïve Bayes	3523 1	97.782	0.978	0.001	0.987	0.978	0.981	15
RF	3523 1	97.407	0.974	0.023	0.979	0.974	0.976	333.6

Considering the evaluation metrics and the execution time, Naïve Bayes is the best algorithm with the highest accuracy rate and short execution time among six applied machine-learning algorithms (Fig. 3; Table 9).

As an evaluation metrics, we have used above table features which are given below:

$$\text{Accuracy} : (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}), \tag{1}$$

$$\text{TPR} : \text{True Positive Rate} = \text{TP}/\text{TP} + \text{FN}, \tag{2}$$

$$\text{FPR} : \text{False Positive Rate} = \text{FP}/\text{FP} + \text{TN}, \tag{3}$$

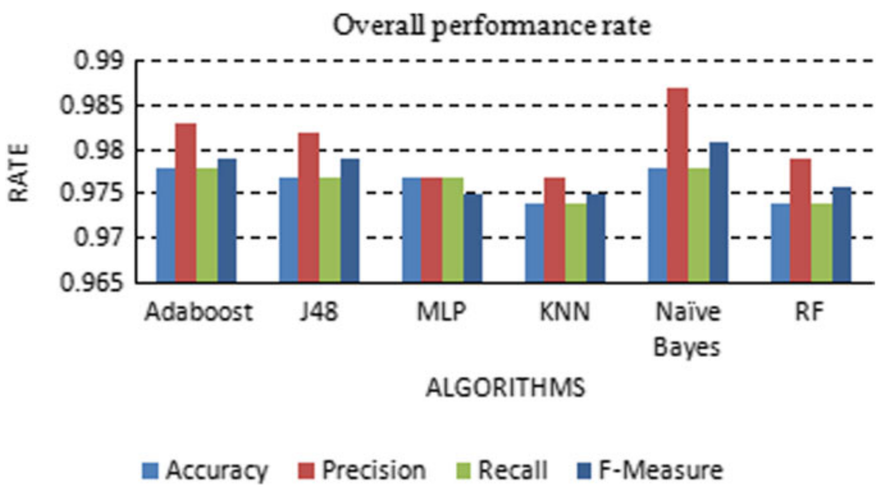


Fig. 3 Overall performance rate of different algorithms

Table 9 Results of Naïve Bayes

Attacks	Testing data size	Accuracy	TPR	FPR	Precision	Recall	F-measure	MCC
BENIGN	27165	97.76	0.97	0.01	1.00	0.97	0.98	0.94
DDoS	299	99.11	0.97	0.09	0.49	0.97	0.65	0.69
DoS GoldenEye	6202	99.97	1.00	0.00	0.99	1.00	0.99	0.99
DoS Hulk	181	99.98	0.99	0.00	0.99	0.99	0.99	0.99
DoS Slowlons	700	99.97	1.00	0.00	0.99	1.00	0.99	0.99
DoS Slowhttptest	628	98.69	0.99	0.13	0.58	0.99	0.73	0.76
Heartbleed	56	99.99	1.00	0.00	0.97	1.00	0.98	0.98

Precision ($\boldsymbol{P_r}$): It is the ratio of correctly classified attack flows (TP), in front of all classified flows (TP + FP).

Recall ($\boldsymbol{R_c}$): It is the ratio of correctly classified attack flows (TP), in front of all generated flows (TP + FN).

F-Measure: It is a harmonic combination of the precision and recalls into a single measure.

$$\boldsymbol{P_r} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},\ \boldsymbol{R_c} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},\ \boldsymbol{F}_{\text{measure}} = \frac{2}{\frac{1}{\mathrm{Pr}} + \frac{1}{\mathrm{Rc}}} \tag{4}$$

Matthews's correlation coefficient (MCC): MCC is defined in terms of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

5 Conclusion and Future Work

It appears to be really difficult to completely defend the network from denial-of-service (DDoS) attacks on the internet today. In this study, we propose a system that evaluated a model called EDSONI for DoS and DDoS attacks. This article concludes that it is possible to train a Naïve Bayes that would successfully classify unseen data in different scenarios with a high accuracy in computer-generated simulation. Challenges and different DDoS prevention approaches are discussed in this paper. The proposed system can be used for the commercial purpose and it can implement as a part of firewall system to combine the working of detection and prevention systems. The enhancement can be made to check the same approach for different other attacks over the network in future.

References

1. Aldibotnet, DDoS Attack Tools, 2012. https://asert.arbornetworks.com/ddos-tools
2. Behal S, Kumar K (2017) Characterization and comparison of DDoS attack tools and traffic generations: a review. IJ Netw Secur 19(3):383–393
3. DDoS Attack on Bank of Greece https://www.hackread.com/anonymous-ddos-attack-bank-greece-website-down/. Accessed 8 June 2018
4. eSecurity Planet, 4 May 2017, www.esecurityplanet.com/network-security/types-of-ddos-attacks.html. Accessed 11 June 2018
5. GitHub engineering, DDoS incident Report. https://githubengineering.com/ddos-incident-report/. Accessed 9 June 2018
6. Github (2013), DDoS Attack Tools. https://github.com/
7. Hacker group 'Anonymous' claims credit for federal cyber-attacks. http://ottawacitizen.com/news/politics/federal-computer-servers-cyber-attacked-clement. Accessed 9 June 2018
8. HSBC internet banking services down after DDoS attack. https://www.theguardian.com/money/2016/jan/29/hsbc-online-banking-cyber-attack. Accessed 9 June 2018
9. Kali Linux Tutorials (2011) THC-SSL-DoS—a denial of service tool against secure web-servers and for Testing SSL-Renegotiation. https://www.thc.org/thc-ssl-dos
10. Kritikos K, Kirkham T, Kryza B, Massonet P (2017) Towards a security-enhanced PaaS platform for multi-cloud applications. Futur Gener Comput Syst 67:206–226
11. Hu L, Bi X (2011) Research of DDoS attack mechanism and its defense frame. In: Computer research and development (ICCRD), 3rd international conference, pp 440–442
12. Melbourne IT, DDoS Attacks of 2017 https://www.tripwire.com/state-of-security/featured/5-notable-ddos-attacks-2017/. Accessed 9 June 2018
13. Mittal A, et al (2011) A review of DDOS attack and its countermeasures in TCP based networks. Int J Comput Sci Eng Surv (IJCSES) 2. https://doi.org/10.5121/ijcses.2011.2413
14. Ferguson P, Senie D (1998) Network ingress filtering: defeating denial of service attacks which employ IP source address spoofing. In: RFC 2267, the internet engineering task force (IETF)
15. Packet Storm (2015) DDoS Attack Tools. http://packetstormsecurity.org
16. Sharafaldin I, et al (2018) Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Proceedings of the 4th international conference on information systems security and privacy (ICISSP)
17. Shruthi P (2017) Network security in digitalization: attacks and defence. Int J Res Comput Appl Robot
18. Sourceforge (2012) DDoS Attack Tools. http://sourceforge.net/projects
19. UK National Lottery (2017) DDoS Attacks of 2017 https://www.tripwire.com/state-of-security/featured/5-notable-ddos-attacks-2017/. Accessed 9 June 2018
20. Vinko Z, KF, VS (2017) Denial of service attacks, defenses and research challenges. Springer Science Business Media New York. https://doi.org/10.1007/s10586-017-0730-x
21. Web attack knocks BBC websites offline https://www.bbc.com/news/technology-35204915. Accessed 9 June 2018

Implementation of Four-Pass Protocol Scheme Using Mathematical Series Cipher Encryption and Decryption in a Communication Network

V. Joseph Emmanuvel and E. J. Thomson Fredik

Abstract Security features come into play when it is compulsory or enviable to defend the information communication from an opponent who attempts to break the confidential information. Security of the information is the encryption of the message, which messes up the message so that the message is indecipherable by the opponent and the addition of the code based on the contents of the message which can be used to verify the uniqueness of the sender. Some information shared by two persons should be unidentified to the malicious intruders. An example is an encryption key used in concurrence with the transformation to mess up the message before transmission and unscramble it on reception. In this connect, a new algorithm has to be designed for performing and ensuring the security-related information transmission between two parties. The algorithm should be designed in such a way that an opponent cannot overcome its purpose. The secret message needs to be generated for implementing security in the algorithm. In designing the consent security service, the method has to be developed for the sharing and allocation of secret information. This research paper proposes a new technique for implementing four-pass protocol schemes which will use mathematical series perceptions for the encryption and decryption of information in a communication network.

Keywords Cryptography · Encryption · Decryption · Mathematical series cipher · Four-pass protocol

1 Introduction

The invention of individual computers has been a revolving point in the history of humankind. Following that, internetwork came into existence, which has made some

V. J. Emmanuvel · E. J. Thomson Fredik (✉)
Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
e-mail: thomson500@gmail.com

V. J. Emmanuvel
e-mail: josejan1978@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_8

surprising variances in the day to day human life. Since such types of inventions are active, threats have become a potential for violation of security. Since the capability of computer storage is so high, security of information is more often required to preserve the information while it is being transmitted in a communication network. Malicious people cannot be stopped from their nosy activities like meddlesome into others' communication. The only option to protect our message is to process it in a disparate format, where the sender and receiver can only know the inimitable message. In [1], William Stallings, 2006 explained that Cryptography is a technique used for deciphering a message without any knowledge of the enciphering detail. The original intelligible message or data that is fed into the algorithm as input is called the plain text and the scrambled message developed is called the ciphertext. The algorithm that performs various substitutions and transformations on the plain text is called data encryption. On the other hand, the encryption algorithm which runs in reverse and takes the ciphertext, the secret key and produces the plain text is called data decryption. The secret key is the input to the encryption algorithm and it is value-dependent of the plain text. An efficient algorithm has to be developed such that an opponent who knows the same and has access to one or more ciphertexts would be unable to decipher the text or find out the key. There are numerous three-pass algorithms that make use of the same keys for all the three steps of encryption and decryption. In the aspect of overcoming such boundaries in present system, a new technique has been projected by implementing four-pass protocol schemes which will use mathematical series perceptions for the encryption and decryption of information in a communication network that utilizes various keys in every pass. The traditional cryptographic general process is depicted in Fig. 1

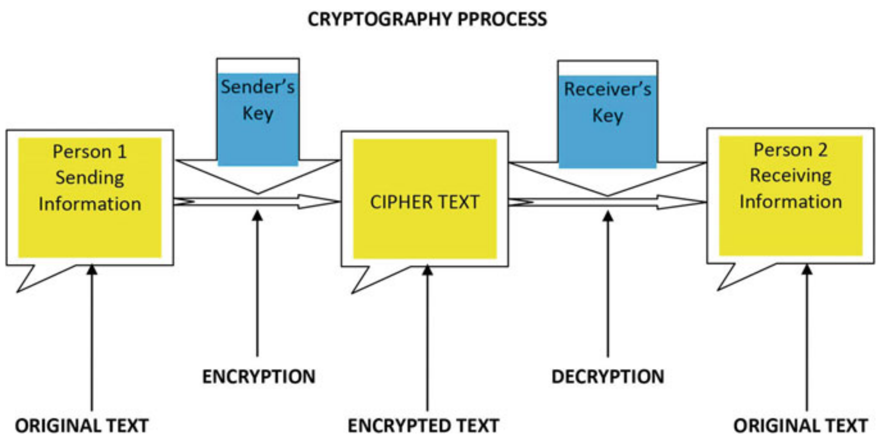


Fig. 1 Cryptography process

2 Literature Survey

In today's cryptography, there are three dissimilar mechanisms involved: Symmetric Key Encipherment, Asymmetric Key Encipherment, and Hashing.

Symmetric Key Encipherment

In [2], Behrouz A Forouzan, 2015 stated that Symmetric Key Encipherment is otherwise called Secret Key Encipherment or Secret Key Cryptography. Here an individual will be sending a message to another individual over an insecure channel so that the third party cannot understand the contents of the message being transmitted through the channel. The person one and two encrypts and decrypts the message using an encryption and decryption algorithm, respectively. In [3], Ayush, 2010 mentioned that the SSL, DH, RSA, and SSH algorithms have been used in such types of symmetric processes. In [4], Amin Subandi, et al. 2017 recommended one more Protocol Implementation using the same three passes in Vigenere Cipher Classic Cryptography Algorithm where the investigators initiated a larger matrix representation as to the key. In [5], Boni Oktavianab, et al. 2016 introduced a cryptographic technique that uses the Three-Pass Protocol Implementation which gets the help of Caesar Cipher Cryptography where the key used has been 3 with shift values of 6 and 5, respectively where all these methods utilized the same keys for all the three consecutive passes. In [6], Joseph Emmanuvel and Thomson Fredrik, 2020 gave a new proposal of implementing three-pass protocol in a communication network where number cipher series encryption and decryption were done by using three different keys at all the three various passes. In [7], Duc Manh Nguyen et al. 2019 suggested a three-pass protocol procedure for many bits transfer, where the same keys were utilized in all the passes. In [8], Robbi Rahim and Ali Ikhwan 2016 clearly gave a picture that the three-pass protocol is a structure that permits the sender to send an encrypted message to a beneficiary without distributing keys to the receiver, where the sender and receiver never swap over the keys and communication is performed in three directions where the sender and receiver use the same keys. In [9], Blackledge et al. 2019 proposed an application using phase-only digital encryption to the three-pass protocol leading to a new 'no- key-exchange algorithm'. After providing a study on the theoretical background to the method, they presented an algorithm on a step-by-step basis together with three examples of cryptanalysis. In [10], Robbi Rahim et al. 2021 proposed the three-pass protocol using Pohlig Hellman Algorithm. The proposed algorithm solved a key distribution problem in symmetric and asymmetric cryptography by allowing senders and receivers to process encryption and decryption due to lack of key distribution processes.

3 Proposed Work

In the new work proposed, a novel four-pass protocol scheme has been suggested which uses the even numbers, square of numbers, factorial series, and cubic series for the first, second, third, and fourth passes,

respectively to encrypt and decrypt the message.

Pass—I: 2, 4, 6, 8, 10,……..First Key (KE1)
Pass—II: 1, 4, 9, 16, 25, 36,…..Second Key (KE2)
Pass—III: 1, 1, 2, 6, 24, 120,…….Third Key (KE3)
Pass—IV: 1, 8, 27, 64, 125, …….Fourth Key (KE4)

The Encryption algorithm for all the four passes is as follows

$$\text{ET1} = (\text{OT1} + \text{KE1})\ \text{Mod}\ 26 \quad - - - \text{Pass} - \text{I} \tag{1}$$

where RT1 = Encrypted Text1, OT1 = Original Text1, KE1 = Key1 and 26 (Total number of English alphabets).

In the same way,

$$\text{ET2} = (\text{OT2} + \text{KE2})\ \text{Mod}\ 26 - - - \text{Pass} - \text{II} \tag{2}$$

$$\text{ET3} = (\text{OT3} + \text{KE3})\ \text{Mod}\ 26 - - - \text{Pass} - \text{III} \tag{3}$$

$$\text{ET4} = (\text{OT4} + \text{KE4})\ \text{Mod}\ 26 - - - \text{Pass} - \text{IV} \tag{4}$$

The four-pass protocol process is explained in the below flow chart (Fig. 2).

4 Testing and Implementation

Implementation of four-pass protocol scheme mathematical series has been tested and verified in this section. Consider the illustration specified below. Let us acquire into description the following text as the original text.

4.1 Original Text: Think Twice Do Once

The first key taken into act in the Pass—1 is an even number series. Subsequently, the square, factorial, and cubic numbers series are taken into consideration. Now the

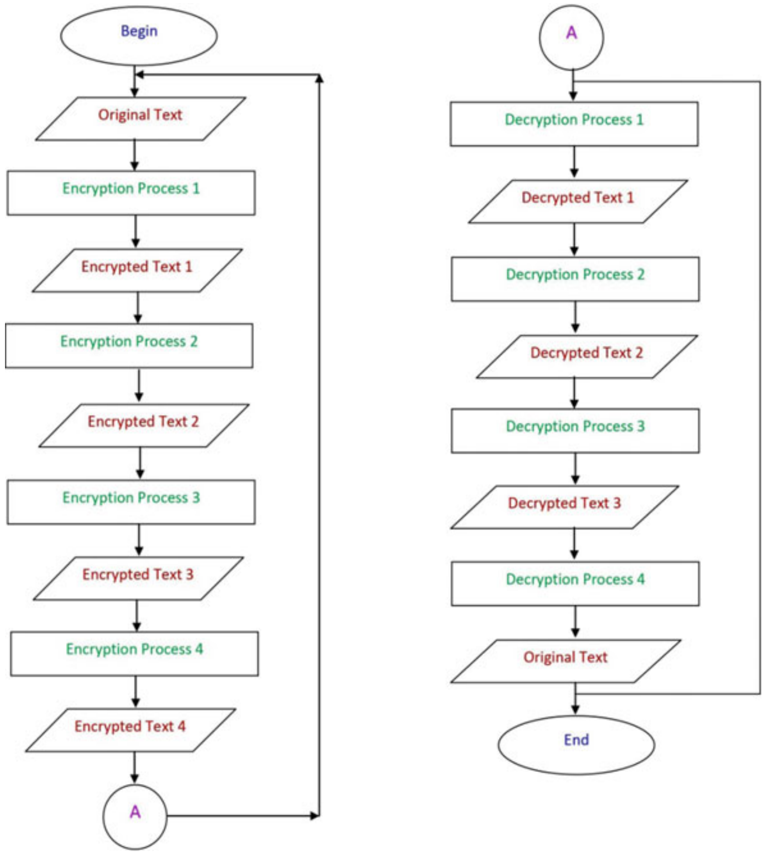


Fig. 2 Four-pass protocol scheme

encryption occurs four times as per the keys used and the encryption algorithm is as follows.

Algorithm

Step 1: The first key taken in Pass—I to encrypt the original text is an even number series. Hence the first even number 2 is taken as the key in first pass.

$$\text{ET1} = (\text{OT1} + \text{KE1}) \text{ Mod } 26 \quad - - - \text{Pass} - \text{I}$$

Original Text = T (First Letter)
ET1 = (T + 2) Mod 26, i.e., ET1 = 2. Hence we get the encrypted text V. ET1 = V.

Step 2: The second key taken in Pass—I to encrypt the original text is the second number of the even number series. Hence the second even number 4 is taken as the key in first pass.

$$\text{ET1} = (\text{OT1} \ + \ \text{KE1}) \text{ Mod } 26 \quad - - - \text{Pass} - \text{I}$$

Original Text = H (Second Letter)
ET1 = (H + 4) Mod 26, i.e., ET1 = 4. Hence we get the encrypted text L. ET1 = L.

Step 3: The third key taken in Pass—I to encrypt the original text is the third number of the even number series. Hence the third even number 6 is taken as the key in first pass.

$$\text{ET1} = (\text{OT1} \ + \ \text{KE1}) \text{ Mod } 26 \quad - - - \text{Pass} - \text{I}$$

Original Text = I (Third Letter)
ET1 = (T + 6) Mod 2,6 i.e., ET1 = 6. Hence we get the encrypted text O. ET1 = O.

Step 4: The fourth key taken in Pass—I to encrypt the original text is the fourth number of the even number series. Hence the fourth even number 8 is taken as the key in first pass.

$$\text{ET1} = (\text{OT1} \ + \ \text{KE1}) \text{ Mod } 26 \quad - - - \text{Pass} - \text{I}$$

Original Text = M (Fourth Letter)
ET1 = (T + 8) Mod 26, i.e., ET1 = 8. Hence we get the encrypted text V. ET1 = V.

Step 5: The fifth key taken in Pass—I to encrypt the original text is the fifth number of the even number series. Hence the fifth even number 10 is taken as the key in first pass.

$$\text{ET1} = (\text{OT1} \ + \ \text{KE1}) \text{ Mod } 26 \quad - - - \text{Pass} - \text{I}$$

Original Text = K (Fourth Letter)
ET1 = (K + 8) Mod 26, i.e., ET1 = 8.

Hence we get the encrypted text U. ET1 = U.
Hence the final encrypted text obtained in the first pass is depicted in Table 1.

Table 1 Encrypted text in Pass—I

ENCRYPTION PROCESS - I					
OT	T	H	I	N	K
ET	V	L	O	V	U
OT	T	W	I	C	E
ET	X	K	Y	U	Y
OT	D	O			
ET	Z	M			
OT	O	N	C	E	
ET	O	P	G	K	

In the above Table 1, the encrypted text obtained in Pass—I using the Even Numbers series is VLOVU XKYUY ZM OPGK. Similarly, the second key taken into act in the pass—II is a square number series.

In Table 2, the encrypted text obtained in Pass—II using the Square of Numbers series is VPXLT HHKXU QA BDXG. The third key taken into act in the pass—3 is a factorial number series.

In Table 3, the encrypted text obtained in Pass—III using the Factorial Numbers series is VPZRR HZGRS WO NDXG. Now in the last pass, Table 4 gives the Encrypted text using the cubic numbers series. The fourth key is taken into act in the pass—IV is a cubic number series.

In Table 4, the cubic numbers series has been implemented to convert the original text to encrypted text in the fourth pass. The final encrypted text in the final pass is VXADM PEYSI BA ARSU. Now on the alternate hand, the receiver can be able to decrypt the encrypted text in the reverse order for times. The cubic, factorial, square

Table 2 Encrypted text in Pass—II

ENCRYPTION PROCESS - II					
OT	V	L	O	V	U
ET	V	P	X	L	T
OT	X	K	Y	U	Y
ET	H	H	K	X	U
OT	Z	M			
ET	Q	A			
OT	O	P	G	K	
ET	B	D	X	G	

Table 3 Encrypted text in Pass—III

ENCRYPTION PROCESS - III					
OT	V	P	X	L	T
ET	V	P	Z	R	R
OT	H	H	K	X	U
ET	X	Z	G	R	S
OT	Q	A			
ET	W	O			
OT	B	D	X	G	
ET	N	D	X	G	

Table 4 Encrypted text in Pass—IV

ENCRYPTION PROCESS - IV					
OT	V	P	Z	R	R
ET	V	X	A	D	M
OT	X	Z	G	R	S
ET	F	E	Y	S	I
OT	W	O			
ET	B	A			
OT	N	D	X	G	
ET	A	R	S	U	

and even numbers series can be utilized for the decryption process on all the four passes, respectively in the reverse order to obtain the original text. The decipherment process is as follows (Tables 5, 6, 7 and 8).

Table 5 Decrypted text in Pass—IV

DECRYPTION PROCESS - IV					
OT	V	X	A	D	M
DT	V	P	Z	R	R
OT	F	E	Y	S	I
DT	X	Z	G	R	S
OT	B	A			
DT	W	O			
OT	A	R	S	U	
DT	N	D	X	G	

Table 6 Decrypted text in Pass—III

DECRYPTION PROCESS - III					
OT	V	P	Z	R	R
DT	V	P	X	L	T
OT	X	Z	G	R	S
DT	H	H	K	X	U
OT	W	O			
DT	Q	A			
OT	N	D	X	G	
DT	B	D	X	G	

Table 7 Decrypted text in Pass—II

DECRYPTION PROCESS - II					
OT	V	P	X	L	T
DT	V	L	O	V	U
OT	H	H	K	X	U
DT	X	K	Y	U	Y
OT	Q	A			
DT	Z	M			
OT	B	D	X	G	
DT	O	P	G	K	

Table 8 Decrypted text in Pass—I

DECRYPTION PROCESS - I					
OT	V	L	O	V	U
DT	T	H	I	N	K
OT	X	K	Y	U	Y
DT	T	W	I	C	E
OT	Z	M			
DT	D	O			
OT	O	P	G	K	
DT	O	N	C	E	

Since other three passes are to be implemented, the decrypted text is still unreadable.

Now after implementing all the four passes, the final original text obtained is THINK TWICE DO ONCE. The same process of encryption and decryption has been developed into a JAVA program and the same has been executed and the results have been verified. The screenshot of the sample input (Original Text) and the subsequent output (Encrypted text) has been given below (Fig. 3).

Similarly, the same process occurs in the reverse direction and final original text will be obtained after completing all the four passes. In processing the iteration in every step in the proposed four-pass protocol scheme, it has been found that, the time taken for the encryption and decryption process using mathematical series cipher encryption and decryption is much faster than the existing three-pass protocol process. The comparison of the same in both the passes has been displayed in Tables 9 and 10, respectively.

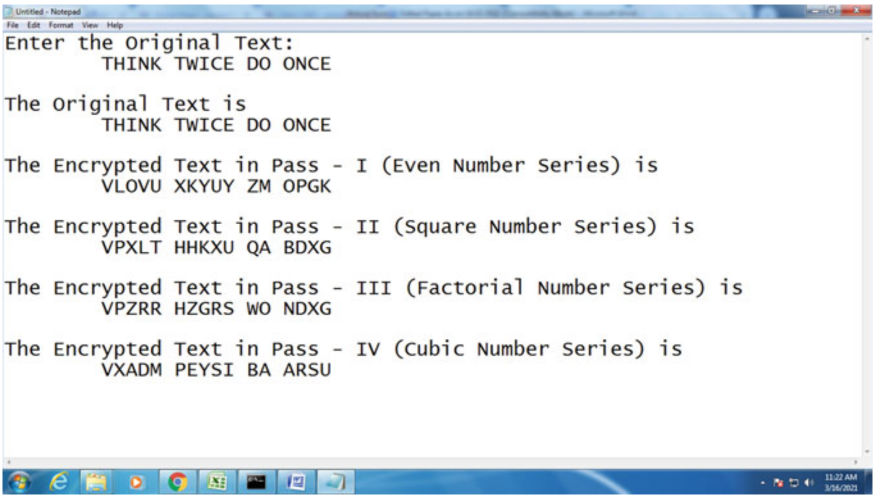


Fig. 3 Sample input and output

Table 9 Comparison of encryption process of three-pass protocol with four-pass protocol

Size of the file in kilo bytes	Three-pass encryption time in milliseconds	Four-pass encryption time in milliseconds
1	14.07	12.06
2	23.635	21.41
3	16.2	14
4	19.5	17.02
5	13.8	11
10	25.2	23.01
20	39.6	36.03
50	86.4	83
100	167.2	160.02
200	325.5	317.17
500	817.5	789.78
1000	1736.04	1635
2000	3894.7	3467.10

Table 10 Comparison of decryption process of three-pass protocol with four-pass protocol

Size of the file in kilo bytes	Three-pass decryption time in milliseconds	Four-pass decryption time in milliseconds
1	11.19	09.08
2	10.24	08.03
3	22.89	20.10
4	12.3	10.07
5	13.6	11.06
10	46.3	44
20	40.5	38.18
50	68	67
100	121.37	119.41
200	261.1	243.90
500	620.2	600.035
1000	1250.1	1213.45
2000	2931.6	2765.32

The iteration processes in both the three-pass and four-pass protocol schemes have been clearly depicted in the below charts in Figs. 4 and 5, respectively.

X-Axis represents the size of the file in kilobytes at each iteration. Y-Axis represents the time taken for encryption in Milliseconds at each iteration.

X-Axis represents the size of the file in kilobytes at each iteration. Y-Axis represents the time taken for decryption in Milliseconds at each iteration. It has been

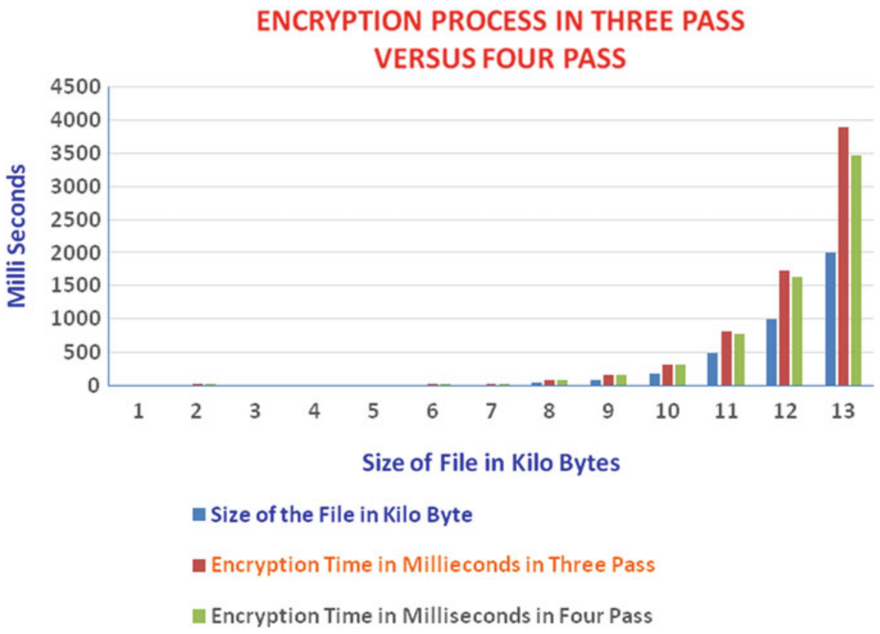


Fig. 4 Encryption in three pass versus four pass

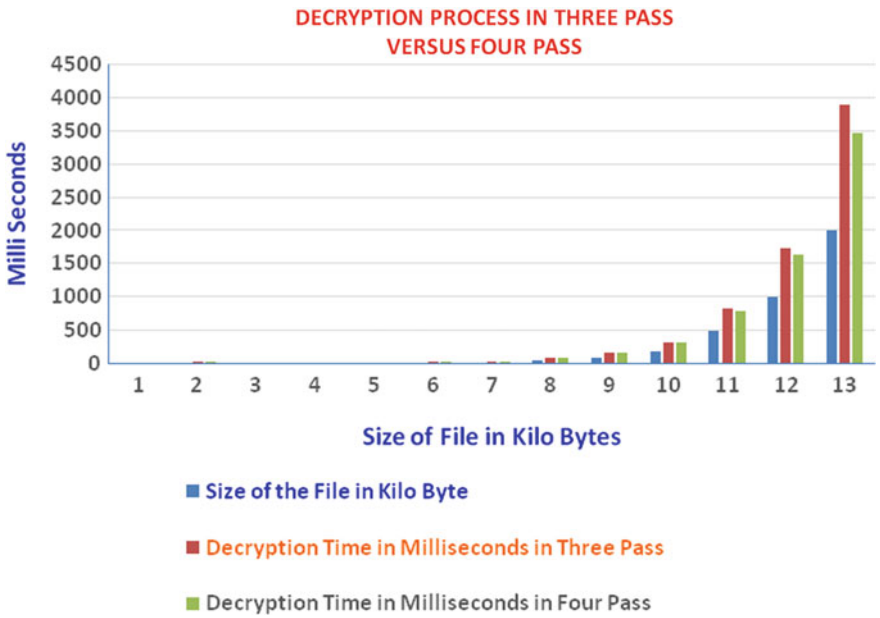


Fig. 5 Decryption in three pass versus four pass

found that the time taken by the proposed four-pass protocol scheme is much better than the existing three-pass protocol scheme in terms of security and time. Since the proposed algorithm uses four levels of encryption process, it is very difficult for intruders or hackers to break the security of the proposed methodology.

5 Conclusion

The new proposed four-pass protocol implementation using mathematical cipher encipherment ensures that the messages transmitted over a communication network are purely secured and the intruder or the attackers may find it very much difficult to capture the actual message. Utilization of the mathematical number series in four various traditions guarantees the security of the message. From the research undergone, it has been notified that by using the keys differently in all the four passes the currently proposed mechanism confirms more security when compared to the already existing schemes.

References

1. Stallings W (2003) In: Cryptography and network security—principles and practices. 3rd edn. Pearson Education, Prentice Hall. ISBN 81-7808-902-5
2. Forouzan BA (2015) Cryptography and network security. 3rd edn. The McGraw Hill Companies, Inc, ISBN 978-93-392-2094-5
3. Ayushi (2010) A symmetric key cryptographic algorithm. Int J Comput Appl (0975–8887):15
4. Subandi A, Meiyanti R, Sandy CLM, Sembiring RW (2017) Three-pass protocol implementation in vigenere cipher classic cryptography algorithm with key stream generator modification. Adv Sci Technol Eng Syst J 2(5)
5. Oktavianab B, Utama Siahaan AP (2016) Three-pass protocol implementation in caesar cipher classic cryptography. IOSR J Comput Eng 18(4)
6. Joseph Emmanuvel V, Thomson Fredrik EJ (2020) Three pass protocol implementation using number cipher encryption in a communication network. J Xi'an Univers Architect Technol X11(V111)
7. Manh Nguyen D, Kim S (2019) A quantum three pass protocol with phase estimation for many bits transfer. In: International conference on advanced technologies for communications (ATC), IEEE, Accession No: 19239948
8. Rahim R, Ikhwan A (2016) Study of three pass protocol on data security. Int J Sci Res 5(11)
9. Blackledge JM, Tobin P, Adolfo CM, Govere W, Sibanda S (2019) Phase only digital encryption using a three pass protocol. In: Proceedings of IEEE international conference on irish signals and systems
10. Rahim R, Zahri NAH, Mohd Warip MNB (2021) Additional security using three-pass protocol and Pohlig-Hellman. In: IOP conference series: materials science and engineering

A Novel Spectrum Sensing Technique for Multiple Network Scenario

Sitadevi Bharatula and B. S. Murthy

Abstract Cognitive Radio Networks (CRN) use spectrum sensing techniques to provide connectivity to secondary users (SU) while ensuring nil or minimal interference is caused to the licensed user. Noisy conditions severely impact the finding of the spectrum white spaces as SUs generally operate at low signal power conditions and spectrum sensing may get affected by shadowing or fading. Estimating the channel state information estimation under cooperative method will provide a solution to tackle this issue wherein the primary channel state is estimated, for taking a decision on whether the primary channel is occupied or not. Multistage threshold techniques provide better accuracy in finding out the vacant primary channels, but it increases the complexity of the system. Review of the research works shows that the existing techniques are developed for a single RF signal scenario and hence not useful in a multiple network environment. Hence, in this scenario, there is a need to develop an efficient and accurate technique to detect spectrum white spaces to work effectively for multiple networks. In order to implement the spectrum sensing effectively in heterogeneous wireless networks, a novel energy detection scheme is proposed in this work for accurate spectrum sensing and efficient dynamic spectrum access to the secondary users. Probability of detection of vacant primary channel is improved in this method by combination of transmitter detection and channel state information. The combination of sensing decision and channel state estimation has been used for better spectrum sensing. Under this technique, secondary user information is processed with the help of a specially designed channel filter using universal bandpass sampling algorithm. Some important features of the spectrum sensing technique proposed in this research work are that it can sense the spectrum in heterogeneous network scenarios and under low Signal-to-Noise Ratio (SNR) conditions. The results obtained with this technique are compared with the other

S. Bharatula (✉)
Department of ECE, School of Engineering, Amrita Vishwa Vidyapeetham, Chennai Campus, Chennai, India
e-mail: sitabharatula@gmail.com

B. S. Murthy
Department of Telecommunications, Ministry of Communications, Government of India, New Delhi, India
e-mail: budaraju.sm69@nic.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_9

traditional methods which showed that it gives better performance and the same has been verified for various SNR values.

Keywords Cognitive radio · Energy detection · Primary user · Secondary user · Spectrum sensing

1 Introduction

In some of the spectrum sensing approaches, channel gain, Signal-to-Noise Ratio (SNR) and other parameters related to primary user transmission are made available to the secondary users. The Channel State Information (CSI) would decrease the computational complexity in spectrum sensing. Further, detection of white spaces with multiple antenna elements provides better results under noisy conditions. But, this method increases computational elements and complexity. The technique proposed in this research article reduces the computational elements and thus makes the design simple with the help of multi-band channel filter and the CSI. For detection of multiple spectrum bands, the channel filter design incorporates guard band between the multiple bandpass signals. The insertion of guard band between signals reduces the aliasing effects and interference from co-channels.

2 Literature Survey

In the research article [1], the authors presented an algorithm to make use of unused spectrum spaces and the primary channels without interfering with the privileges of primary users. The authors discussed basic functions and general architecture of Cognitive Radio Networks (CRN). Reference [2] has presented an algorithm for interoperability among various wireless communication signals by evaluating sampling frequency and intermediate frequency values. In their work, the authors have considered RF signals of various mobile networks and down-converted these Radio Frequency (RF) signals to their baseband. Authors in [3] have presented an algorithm that allows RF frontend receivers to process different RF signals. In their article, the authors presented a method to reduce co-channel interference among the signals.

In the research article [4], the authors considered multiple antennas and analyzed the signal correlation. Generalized Likelihood Ratio approach is adopted and tests were conducted for diversity combining techniques. In their work, the authors in [5] presented energy-efficient sensing techniques to identify spectrum holes. The authors in [6] presented architecture for adaptive use of spectrum for multi-hop networks. In their article, the authors proposed a scheme for efficient spectrum management. In this method, the unlicensed user observes the primary channel and senses the signal power. The authors in [7] presented a spectrum sensing algorithm for adaptive double threshold technique based on history energy detection.

3 Energy Detection Technique Combined with Channel State Estimation

To improve the efficiency and accuracy of detection of vacant primary channels using energy detection method, channel state information has been used in this work. The hypotheses H_0 (null hypothesis) and H_1(alternate hypothesis)used by the energy detection scheme (as given in [8]) are given in (1) as;

$$\begin{aligned} x_r^n(k) &= w(k); \quad hyp.H_0 \\ x_r^n(k) &= |s_i[k]|^2.h_i^k + w_i(k); \quad hyp.H_1 \end{aligned} \tag{1}$$

Equation (1) indicates that, when the primary channel is vacant, the signal is not available and only the noise signal *w(k)* is sensed. Further, when the primary channel is occupied, this signal along with noise is sensed. The signal energy sensed in energy detection technique is given in (2);

$$T(x) = \frac{1}{K}\sum_{k=o}^{K-1} |x_r[k]|^2 \tag{2}$$

Using the probabilities, the missed detection probability is given in (3);

$$P_d = 1 - P_{md} \tag{3}$$

where P_d is detection probability. Now, the total error probability (P_e) is given in (4) as;

$$P_e = P_{fa} + (1 - P_d) \tag{4}$$

Using "Q" function, for "K" number of samples, threshold value "λ" and SNR value of "γ", the probability terms are calculated as in (5) and (6) below;

$$P_{fa} = \Pr(T(x) < \lambda) = Q\left(\frac{\lambda - \sigma_w^2}{\sigma_w^2/\sqrt{K}}\right) \tag{5}$$

$$P_D = \Pr(T(x) > \lambda) = Q\left(\frac{\lambda - \sigma_w^2(1+\gamma)}{\sigma_w^2(1+\gamma)/\sqrt{K}}\right) \tag{6}$$

4 Components of Proposed Spectrum Sensing Technique

In this method, six wireless technologies are considered, viz., Code Division Multiple Access (CDMA), Global System for Mobile Communication (GSM), Wideband CDMA (WCDMA), 802.11a, 802.11b and Long Term Evolution (LTE). The schematic blocks in the proposed technique are depicted in Fig. 1.

The RF signal is received through the wideband antenna. The received RF signal has signals of all six wireless technologies, viz., Coded Division Multiple Access (CDMA), Global System for Mobile communications (GSM), Wideband CDMA (WCDMA), IEE standard 802.11a, IEE standard 802.11b and Long Term Evolution (LTE) and is given as input to the bandpass filter. The bandpass filter has f_L and f_H as the lower and higher cutoff frequencies for the entire segment. The bandpass filter output is given to the universal bandpass sampling block, which is designed as in [3]. Using the universal bandpass sampling algorithm, bandpass sampling frequency

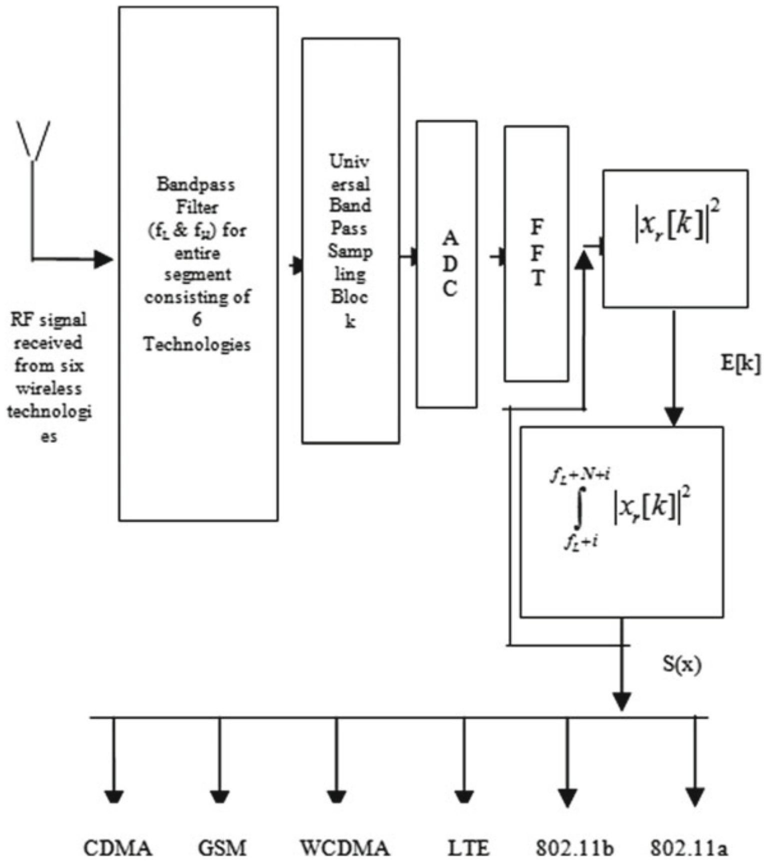


Fig. 1 Proposed spectrum sensing technique

and respective intermediate frequency values are evaluated for different wireless technology signals.

The universal bandpass sampling block will give the output as baseband signal of all six technologies. The baseband signal from the six technologies is then applied to Analog-to-Digital Convertor (ADC) and Fast Fourier Transform (FFT) blocks. The FFT output is squared to get energy of the transmitted signal, E[k]. The energy of the each of the six technologies is then applied to integration block, to extract the sensed energy, *S(x)*. The signal sensed energy for N samples with N_i as the initial value is given in (7) as;

$$S(x) = \frac{1}{N} \int_{f_L+i}^{f_L+N+i} |x_r[k]|^2 \tag{7}$$

For example, let us consider the output of universal bandpass sampling block as CDMA signal. The CDMA signal output from the ADC block is given to the FFT block, which is then squared to determine the energy of CDMA signal which is then passed through integration block to determine the total sensed energy of CDMA signal. For calculation purposes, the values considered for N_i and N for CDMA technology are 0 and 20, respectively. The integration block calculates the summation energy of the first 20 samples of CDMA signal. Accordingly, the result from the 20 samples gives the output of sensed signal energy of CDMA technology. In this work, we have considered 20 samples for sensing the signal from each technology. This process is repeated and for the next 20 samples and GSM signal energy is obtained. The process is continued for the remaining technologies and the sensed signal energies are determined in a similar fashion and the output of the integration block is obtained for all six technologies.

Figure 2 explains the sequence of operations in the sensing technique presented in this article. The received sensed signal S(x) is compared with the dynamic threshold value and the secondary user will identify the white spaces, as shown in Fig. 2. If the white space is found, i.e., if vacant channel is found, then the channel will be used for SU transmission. The detection model considered in this work is given in (8) as;

$$\begin{aligned} x_r^n(t) &= w_i^n(t);\ hyp.H_0 \\ x_r^n(t) &= |s_i[t]|^2.h_i^n + w_i^n(t);\ hyp.H_1 \end{aligned} \tag{8}$$

Cognitive Radio Fusion Center: Energy detection method is not very efficient in low SNR conditions and especially, in multipath fading and hidden-node conditions. For such situations, cooperative spectrum sensing would be useful, wherein the secondary users share the information and a centralized station called fusion center, will take a decision regarding the sensing result.

The fusion center applies fusion rule to arrive at a global decision to make a right decision regarding the spectrum sensing under fading conditions. This is explained in Fig. 3, in which CRN is considered with two primary users and secondary users

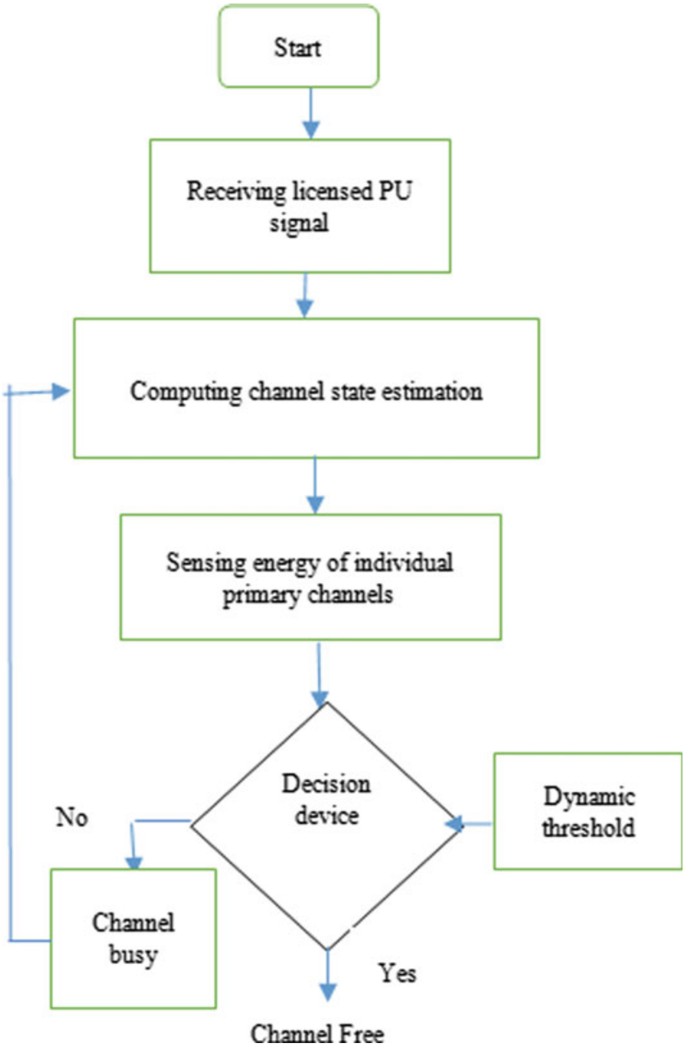
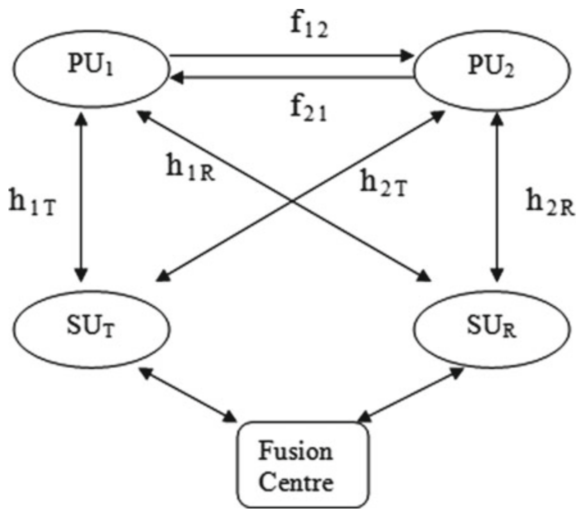


Fig. 2 Sequence of operations in the proposed technique

sharing their information with the fusion center. The secondary users, SU_T and SU_R share the sensed energy signal information of PU_1 and PU_2 with the Fusion Center. The fusion center facilitates the sharing of information among the SUs and using fusion algorithm takes decision about the vacant channel, for effective spectrum sensing.

In this method, optimum solution is formulated to detect the white spaces while ensuring minimal interference to the primary users. In case, vacant channel is wrongly detected and signal transmission is initiated by the secondary user, this will lead to interference with the primary user. Conversely, the false alarm may appear when it is

Fig. 3 Fusion center



decided on H_1 under H_0 condition. This will lead to inefficiency and throughput will be reduced. In order to address these problems, threshold has been varied dynamically in the proposed system. The channel estimator continuously checks the primary channels and fusion center collects and analyses the information. Based on the information collected, fusion center takes a decision on whether channel gain which is analogous to signal-to-noise ratio is poor or good. Accordingly, fusion center fixes the threshold dynamically, as per the channel state estimation. The threshold value is kept low for poor channel conditions and is kept high for good channel conditions.

5 Results

In the proposed system, impact of channel fading and AWGN in the channel has been considered. Simulation testing has been carried out using Matlab tool to analyze the spectrum autocorrelation function and associated features of the PUs. All the experiments were simulated using Monte Carlo simulations. Spectral Autocorrelation Function (SAF) plot of primary users is shown in Fig. 4. The vacant primary channel is shown in Fig. 5, which is available for secondary users to occupy.

The proposed technique performance is compared with general detection method and P_{fa} versus P_d plot is given in Fig. 6, which proves that the proposed technique performs better than the conventional method due to the CSI obtained using the channel filter parameters.

The proposed technique is tested for heterogeneous network scenarios with six mobile technologies, viz., CDMA, GSM, WCDMA, 802.11b, 802.11a and LTE. The frequency segment value and sampling frequency as determined in [7] are considered for the integration of these six mobile technologies. The RF signals are

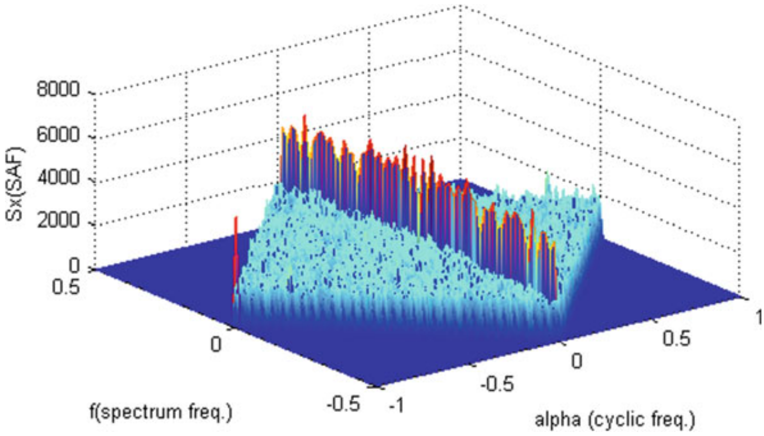


Fig. 4 Primary users' spectrum autocorrelation function

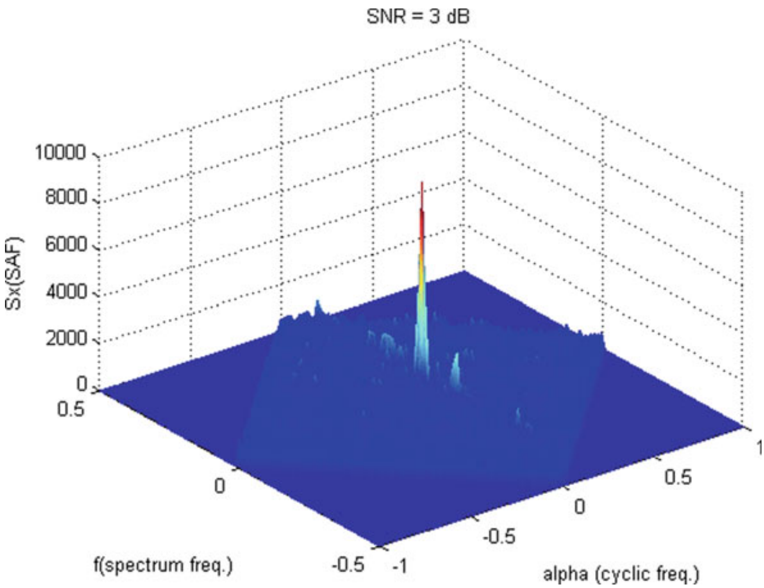


Fig. 5 Inactive primary channel

down-converted to their respective baseband signals using the respective intermediate frequency values for each of the six technologies.

The proposed method is found to be very effective for spectrum sensing in heterogeneous network scenarios, wherein multiple mobile technologies are considered. Figure 7 shows the $P_{fa}VsP_d$ performance curves obtained for the six wireless technologies considered in this work, viz. CDMA, GSM, WCDMA, 802.11b, 802.11a

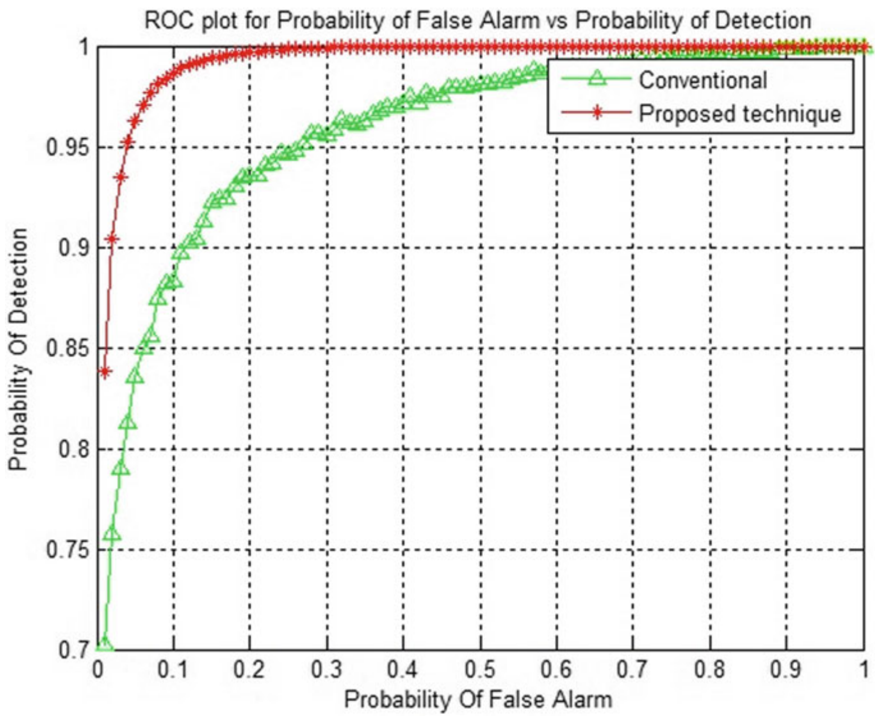


Fig. 6 P_{fa} versus P_d curve

and LTE. Neyman-Pearson threshold observer is applied to these technologies by fixing a threshold for P_{fa} and maximizing P_d. P_{fa} is kept at 0.1 for LTE technology, 0.12 for IEEE 802.11a, 0.14 for WCDMA, 0.16 for IEEE 802.11b, 0.18 for CDMA and 0.20 for GSM technologies. The proposed spectrum sensing technique is very efficient as it does not require scanning the entire frequency segment and using the f_s and f_{IF} of the respective mobile technology under consideration, the signal energy is sensed.

Adaptive Double Threshold (ADT) technique is implemented in many research works to improve the accuracy and the probability of detection. With a view to comparing the performance of the proposed method with ADT technique, work has been carried out with ADT method separately to determine the P_d for different SNR values. These results are shown in Fig. 8. The performance of the proposed method is found to be better than the existing ADT method for different SNR values.

Comparative chart of the values of probability of detection obtained from the proposed technique and the conventional Adaptive Double Threshold Technique (ADT) is given in Table 1.

The significance of the results from the proposed method is that it achieved increased probability of detection which is possible due to estimation of CSI in the proposed scheme. Also, using the technique proposed here, spectrum sensing

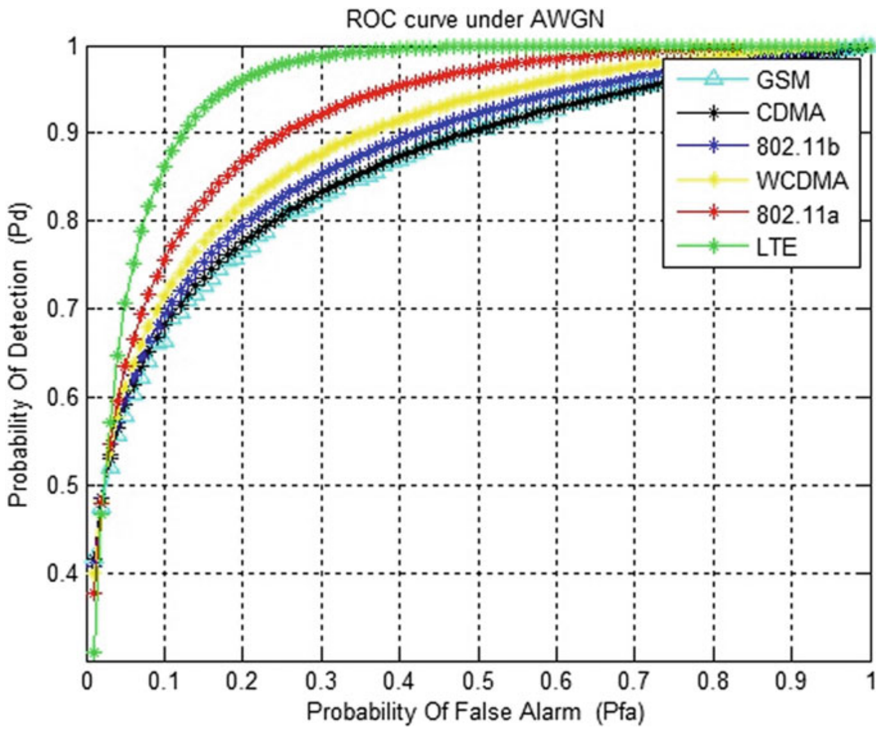


Fig. 7 ROC curves for six wireless technologies

could be efficiently carried out for various technologies. Using this method, better $P_d$value is obtained which also improves the performance and reduces the P_{fa} value. Nearly, 50% of improvement in $P_d$is achieved with the proposed method.

6 Conclusion

CSI estimation based energy-efficient spectrum sensing method is provided in this work. The sequential sample-size framework is used to improve the SU throughput while ensuring minimal interference to PUs. The spectrum sensing accuracy has been increased by combination of channel state estimation and by fixing the threshold dynamically in the energy detection technique. One of the important benefits of the proposed technique is that the system has shown improved performance under various SNR conditions including low SNR values. The results have been compared with the traditional adaptive threshold methods and it has been verified that the proposed system has shown better probability of detection. Another important feature of this technique is that it works in multiple network scenarios and the results are tested for six wireless technologies and given in the previous section. This technique performs

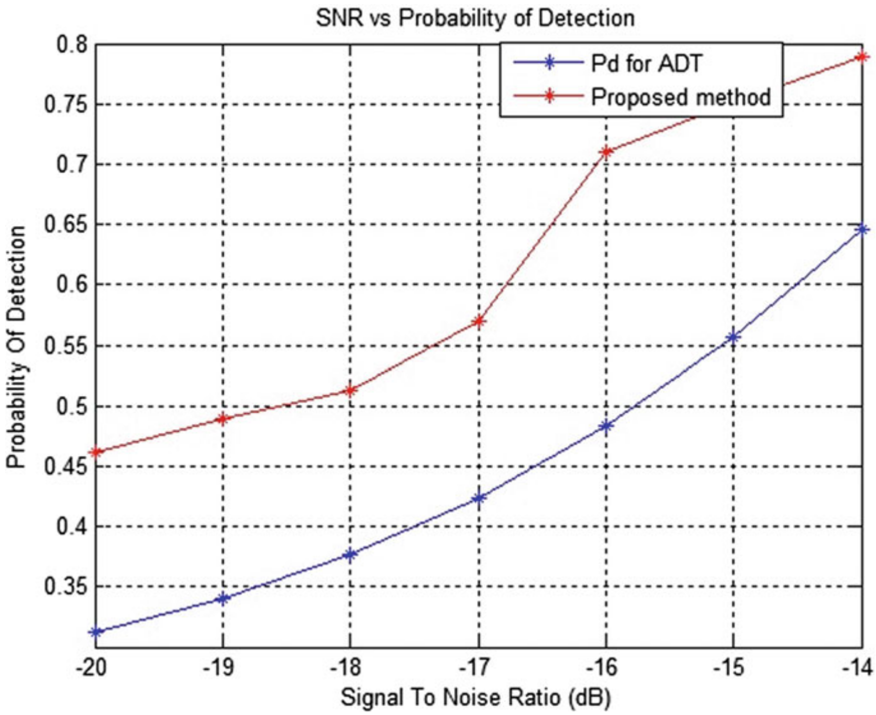


Fig. 8 SNR versus probability of detection

Table 1 Comparative chart

Sl. No	SNR (in dB)	P_d in conventional ADT method	P_d obtained from proposed technique
1	−20	0.316	0.462
2	−19	0.342	0.493
3	−18	0.377	0.514
4	−17	0.424	0.570
5	−16	0.485	0.711
6	−15	0.558	0.756
7	−14	0.648	0.792

satisfactorily under poor signal conditions and noisy channel conditions also which otherwise is not feasible using conventional schemes. The results showed that as the SNR values increase, performance of the technique proposed improved further. Further, the performance of the proposed technique is found to be good even in low SNR conditions.

References

1. Broderson RW, Wolisz A, Cabric D, Mishra SM, Willkomm D (2004) Corvus: a cognitive radio approach for usage of virtual unlicensed spectrum. Berkeley Wireless Research Center (BWRC) technical report, white paper
2. Budaraju SM, Bhagyaveni MA (2013) A novel method of band pass sampling for down conversion of multiple RF signals. In: Elsevier, Proceedings of international conference on advances in electrical and electronics, AETAEE, pp 636–644
3. Budaraju SM, Bhagyaveni MA (2016) Universal band pass sampling algorithm for down conversion of multiple radio frequency signals using software defined radio platform. Circuits Syst 7:497–505
4. Vilar GV, Valcarce RL, Sala J (2011) Multi antenna spectrum sensing exploiting spectral a priori information. IEEE Trans Wireless Commun 10(12):4345–4355
5. Peh E, Liang YC, Guan YL, Pei Y (2011) Energy-efficient cooperative spectrum sensing in cognitive radio networks. In: Proceedings GLOBECOM, pp 1–5
6. Evans T, Tossou K, Ye F, Shu Z, Qian Y, Yang Y, Sharif H (2016) A new architecture for application-aware cognitive multihop wireless networks. IEEE Wirel Commun 23(1):120–127
7. Shanshan Y, Liu J, Wang J, Ullah I (2020) Adaptive double-threshold cooperative spectrum sensing algorithm based on history energy detection. Hindawi Wireless Commun Mobile Comput 4794136. https://doi.org/10.1155/2020/4794136
8. Bagwari A, Singh B (2012) Comparative performance evaluation of spectrum sensing techniques for cognitive radio networks. In: Proceedings 4th IEEE international conference CICN, vol 1. pp 98–105

Control System and IOT

Design and Development of Virtual Simulation Lab with Technology-Enabled Self-assessment

Ajay Kant Upadhyay, Mayank P. Singh, and Amit Yadav

Abstract The paper presents the development of a virtual simulation lab that provides theoretical content, simulations, models, videos, animations, comparative simulations, simulation benches, and self-check quizzes for e-learning and self-study. Self-assessment evaluates the learning capability and knowledge of students. Virtual lab may play a role to teach students without any assistance of teachers or professors in this coronavirus disease 2019 (COVID-19) pandemic situation. The COVID-19 pandemic situation has barriers between students and universities in delivering lectures in classroom. Virtual simulation labs can change the traditional teaching practices of classroom teaching in the world. The virtual lab is developed due to the limited equipment of experiments and simultaneously assisting students. Web-based educational resources have taken an important place in e-education and self-study through an e-learning platform. Virtual simulation lab is an e-learning platform where students can perform experimentation without any direct involvement on instruments in physical lab. An interactive virtual lab is developed to provide web-based global access to everyone without any authentication. This lab does not require any credentials to login into website. This lab provides multiple options for experimentations of various simulations. This can be accessed through web address http://14.139.245.230/mfvlab/home.php and contains over 950 simulations. Each module contains theory as well as audio embedded videos to reduce time and effort required to understand and analysis of various process parameters. Comparative simulation compares among different materials and processed and different parameters with interactive graphs.

Keywords Virtual lab · Simulation lab · Self-assessment · Simulation bench

A. K. Upadhyay (✉)
Dayalbagh Educational Institute, Agra 282005, India
e-mail: ajaykant900@gmail.com

M. P. Singh
Department of Electrical Engineering, GLA University, Mathura, India

A. Yadav
Department of Electronics and Communication, RBSETC, Bichpuri, Agra, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_10

1 Introduction

The virtual lab is defined as "a computer simulation which enables essential functions of laboratory experiments to be carried out on a computer is called a virtual laboratory" [1]. Virtual lab replaces an experiment with a computer model and provides a simulation of a real experiment. It was important to keep on with different safety regulations to the possibility of personal injury or damage to equipment. Distance and lack of financial resources make real experiments difficult to perform, especially in cases where it is necessary to have some advanced and sophisticated tools. Virtual platforms increase availability to students from all over the world, promoting access across labs, institutions, and regions [2]. In this COVID-19 pandemic situation, lockdown has created requirement for developing new technologies besides regular classroom teaching.

The regular classroom teaching in universities/ Institutes or training in industries provides either the use of 2-D diagrams or physical experimentation which are insufficient in elaborating any multidimensional metal forming process [3]. The simulations used in virtual lab assist the existing teaching and training methods and contribute more effectively for better understanding of multidimensional processes and helps in providing skilled students to the industries using information and communication technologies.

The virtual simulation lab provides self-study without teacher's assistance. Default parameters are defined and theoretical content is provided to explain the processes. Students read and evaluate him/her selves with self-check quizzes. A number of "What happens if we change this?" questions are asked. Answering these questions requires going through a learning cycle involving the parameter change, observation of the effect, a check on theory compliance, understanding, and explanation. The use of MCQs in Engineering is probably the most widespread automatic evaluation technology, which consists of a multiple selection quiz where student would answer some or all questions by choosing one option from a list of available answers [4]. PHP programming is used for checking the answers selected by the student match with correct answer or not. A result sheet is provided that shows the selected answers of the student and correct answer for that question. In this result sheet score is also provided on the basis of correct answers matched with selected answers by the student [5]. Comparative simulations explain comparisons among various forming processes with different variables.

2 Methodology

In regular classroom teaching, it is quite difficult to make understand to the students about multidimensional processes. In chalk and board teaching a diagram can be drawn only in two Dimensions. It is difficult to see the effect of various parameters of multidimensional process. Simulation-based teaching is easy and effective for understanding the effect of multidimensional processes with change in various parameters.

Numerical methods are popular techniques for universities, institutes, and research departments with high-performance computations. These are used for analyzing the effect of various process parameters (temperature, forces, velocity, strain rate, friction, etc.) which cannot be seen in physical lab during experimentation. The change in these parameters can be seen in a simulation that increases the knowledge and interest of students studying through virtual simulation lab.

PHP programming is used for self-check quizzes. PHP program checks the answers selected by the student match with correct answer or not. A result sheet is provided that shows the selected answers of the student and correct answer for that question. In this result sheet score is also provided on the basis of correct answers matched with selected answers by the student. With the help of PHP program, only those question numbers are shown in result sheet for which the student opts the answers from multiple-choice questions.

Finite element method (FEM) is a numerical technique used for generating simulations describing solutions for different forming processes. The simulations made by using these numerical methods are converted into videos and embedded into web pages merging with subjective and interactive content. It is not mandatory to know about FEM knowledge to the students for this metal forming virtual simulation lab.

3 Development of Virtual Simulation Lab

Experimentation in physical labs to illustrate fundamental principles of metal forming processes requires a large investment in machinery and technology. It is difficult for a student to observe effects of process parameters and changes in internal structure. Virtual simulation lab makes this easy for students, scholars, and also for teachers.

A special feature of simulation bench in virtual simulation lab permits students/scholars to interact with different types of processes and do comparative analysis of basic forming processes. This work assist students and scholars in increasing their knowledge and better understanding of processes, concepts, and analysis of metal flow. The virtual lab covers all types of metal forming processes with proper theoretical content, simulation benches, and self-check quizzes. It assists the students in operating interactive simulation benches and self-check quizzes for better understanding and analysis of metal forming processes. In this metal forming virtual simulation lab all types of forming processes and their applications are covered with theoretical and interactive content. It can be accessed through http://14.139.245.230/mfvlab/home.php.

3.1 Home Page

The home page of virtual simulation lab website is shown in Fig. 1. The home page shows the processes and applications with menu through which each process can be accessed separately. There are 10 main menus used on this home page and 55 sub-menu are used on the home page for accessing the processes. Home menu has seven sub-menus with glossary that is fetched from database. Equipment's menu has two types of processes, upsetting processes and extrusion processes. There are eight types of upsetting processes and 10 types of extrusion processes in this menu. Processes menu have 16 sub-menu those are different kind of metal forming processes.

There are three types of rolling processes covered in rolling sub-menu. There are seven types of applications used in the application menu. Registration menu provides the facility for registration of visitors. It helps to get the number of visitors visiting this virtual lab. Registration is not mandatory for visitors. Feedback menu is used to get feedback from visitors (Table 1).

Feedback helps in improving the content of virtual simulation lab and shows the interest of visitors in virtual simulation lab. Contributors menu shows the supporting staff who are working and contributing to this virtual simulation lab. Workshops menu shows the outreach of virtual simulation lab where this lab has been presented. References menu shows the references of those papers from which we have taken

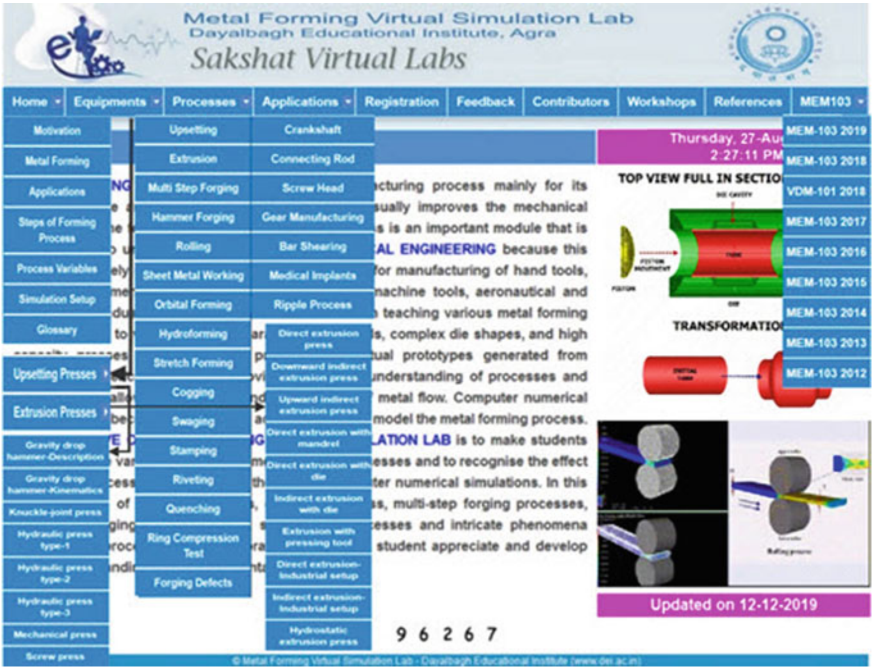


Fig. 1 Metal forming virtual simulation lab home page

Table 1 Menu for accessing processes and applications with sub-menu

S. No	Main menu	Sub menu	
1	Home	7	
2	Equipments—Upsetting	8	18
	Extrusion	10	
3	Processes—Main page	16	
	Rolling	3	
4	Applications	7	
5	Registration	0	
6	Feedback	0	
7	Contributor	0	
8	Workshops	0	
9	References	0	
10	MEM103—MEM103	8	9
	VDM101	1	
Total		60	

help for developing this metal forming virtual simulation lab. MEM103 menu has nine sub-menu that provides the content of Manufacturing Processes-I. These sub-menus are appeared on mouse hover to main menu. Cascading Style Sheet (CSS) is used with PHP programming for developing the interactive home page with menu.

3.2 Processes and Applications

Home page of virtual simulation lab shows the metal forming processes in processes menu. There are 16 types of metal forming processes used in this process's main menu.

1. Upsetting
2. Extrusion
3. Multi-Step Forging
4. Hammer Forging
5. Rolling
6. Sheet Metal Working
7. Orbital Forming
8. Hydroforming
9. Stretch Forming
10. Cogging
11. Swaging
12. Stamping
13. Riveting

14. Quenching
15. Ring Compression Test
16. Forging Defects.

Application menu provides the applications of metal forming processes. It has seven types of sub-menu.

1. Crankshaft
2. Connecting Rod
3. Screw Head
4. Gear Manufacturing
5. Bar Shearing
6. Medical Implants
7. Ripple Process.

These processes and applications have sub-menu for accessing the simulations, simulation bench, Comparative simulations, applications and self-check quizzes.

Upsetting Process: Virtual simulation lab provides knowledge to students, scholars, and teachers about upsetting processes. On visiting upsetting processes from home page an overview of upsetting process is appeared and provides menu for accessing the simulations, simulation bench, comparative simulations, applications, and self-check quiz. It has seven main menus. Upsetting menu has theoretical content and simulations menu provides audio embedded simulations with different parameters. Applications menu has applications of upsetting processes. Simulation bench, comparative simulations, and self-check quiz are other main menus of this module. Home menu is used to reach home page of this virtual simulation lab website. Upsetting process web page is shown in Fig. 2.

Rolling Process: Rolling process is accessed from rolling menu that appears when mouse is hovered on process menu on home page. When it is accessed, theoretical content about rolling process is appear. It has four main menus as rolling process,

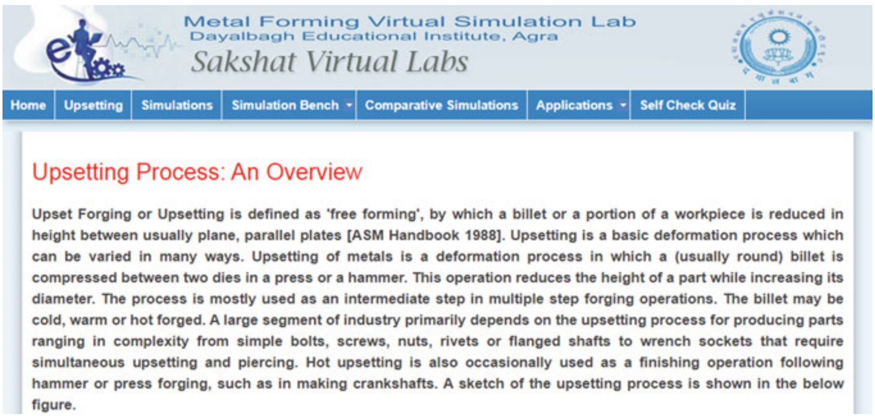


Fig. 2 Upsetting process web page

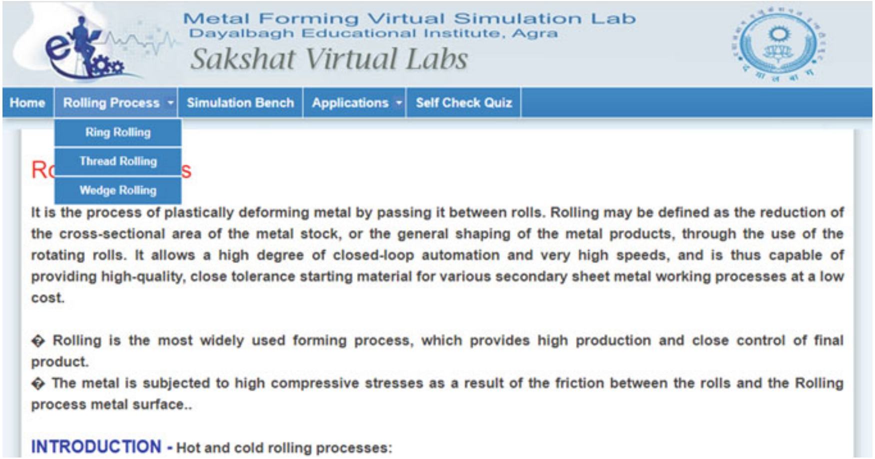


Fig. 3 Rolling process web page

simulation bench, applications, and self-check quiz. Home menu in this webpage is used to return to home page of virtual simulation lab.

There are three types of rolling processes used in this virtual simulation lab.

1. Ring Rolling
2. Thread Rolling
3. Wedge Rolling.

When mouse is hovered on processes menu, a sub-menu appears in which rolling sub-menu access the rolling processes. When mouse is hovered on rolling process then a sub-menu shows three types of rolling processes used in this virtual simulation lab. Figure 3 shows the web page of rolling process.

4 Simulation Bench

Simulation bench is an interactive feature for students to select parameters according to choice and find results. This allows students and scholars to study and analyze the result in a cost-effective manner with better visualization of miniscule details of process. Figure 4 shows the interactive simulation bench for mechanical press which can be accessed from the drop-down menu of simulation bench in upsetting process. Students select parameters such as material, L/D ratio of workpiece, coefficient of friction, and temperature of material to find the result.

On submitting the parameters a simulation is appear with graph as shown in Fig. 5. This simulation is embedded with audio and theoretical content shown below to simulation that explain the process of operation held in this simulation. Colors in this simulation show the temperature changing during the operation. Parameters

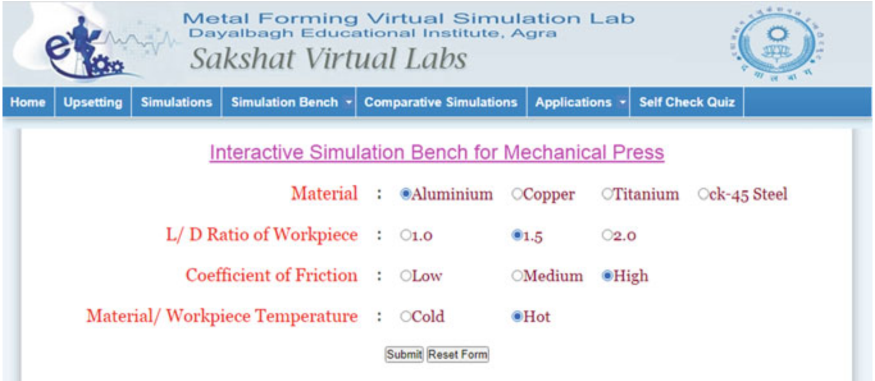


Fig. 4 Interactive simulation bench for mechanical press

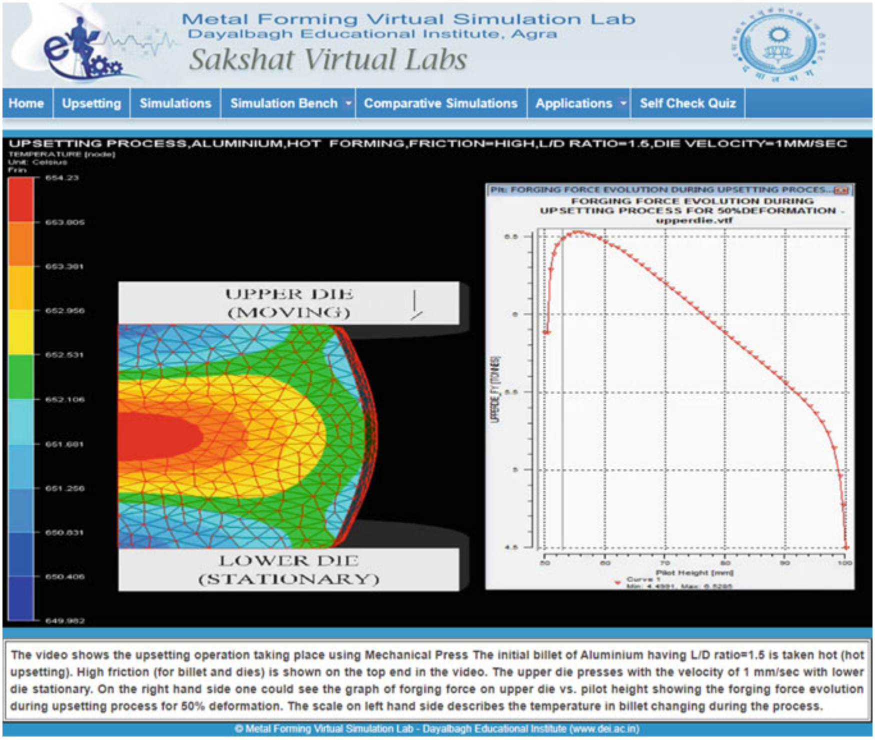


Fig. 5 Simulation according to selected parameters

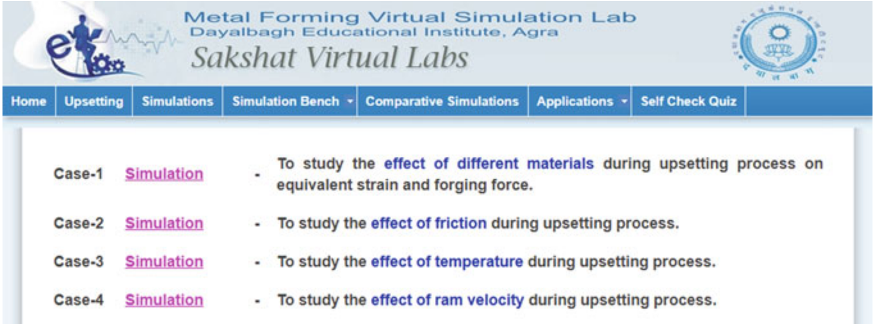


Fig. 6 Comparative simulation of upsetting processes

are shown at the top of the simulation by which students can verify parameters with selected parameters on the simulation bench.

5 Comparative Simulations

Comparative simulations explain the effect of different types of materials, effect of friction, effect of temperature, and effect of ram velocity in upsetting process. Figure 6 show the different type of comparative simulations for upsetting process. It compares among different types of materials with different kinds of parameters.

Figure 7 shows the simulation of comparison for effect of ram velocity. This simulation is embedded with audio that explains the comparison and theoretical content below this simulation explains the operation perform during the upsetting process. This simulation depicts the comparison for effect of ram velocity among copper, aluminum, titanium, and steel material.

Figure 8 shows the Comparative simulations for extrusion process. This compares the effect of solid and pipe extrusion, effect of radius, and effect of temperature.

Figure 9 shows the simulation that explains the effect of temperature for aluminum on cold forming at different die angles. First extrusion process has to die angle of 60 degrees with mandrel, second has 90 degrees with mandrel and third extrusion process has to die angle of 45 degrees without mandrel. This simulation is also audio embedded to explain the process and theoretical content define the operation performed in this simulation.

6 Self-check Quiz

Each process and application in this virtual simulation lab is supported by theoretical content and self-check quiz. This helps students to check their learning capability

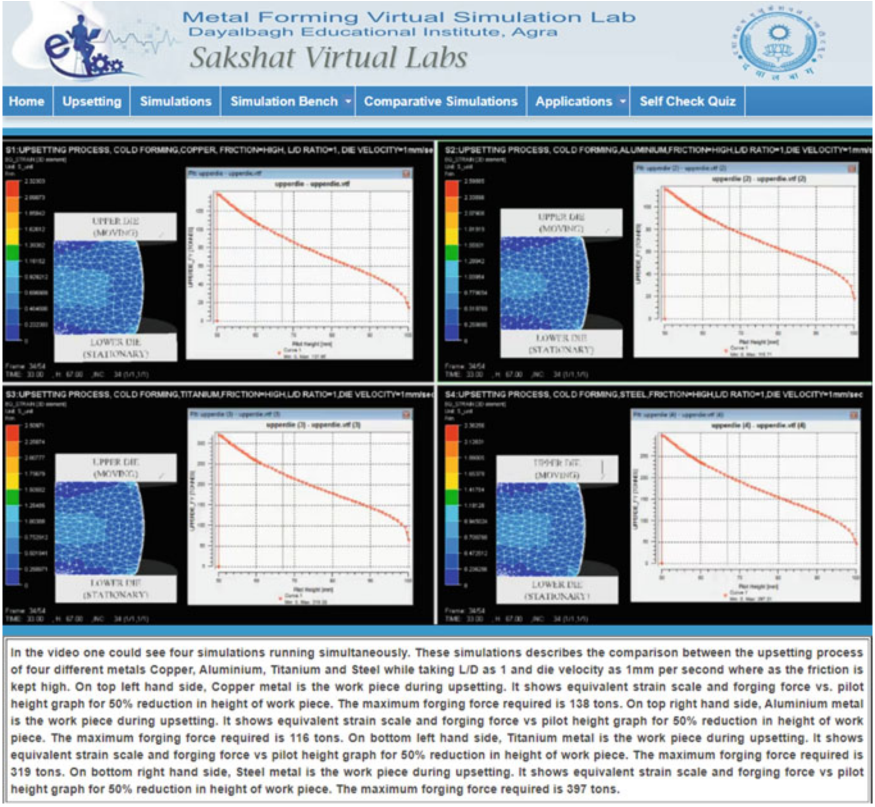


Fig. 7 Comparative simulation for effect of ram velocity

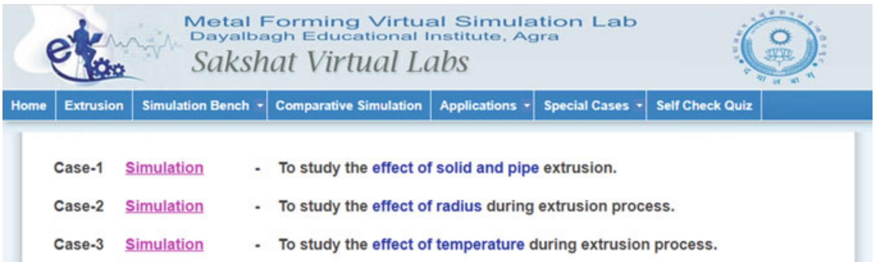


Fig. 8 Comparative simulation for extrusion processes

and knowledge. Figure 10 shows a sample self-check quiz for crankshaft forming. There are multiple-choice questions in self-check quizzes. Student read theoretical content, observe the simulations, and analyze their learning capability by using these self-check quizzes. There are multiple-choice questions in which student select one answer from four options. Students follow this process to all or some questions

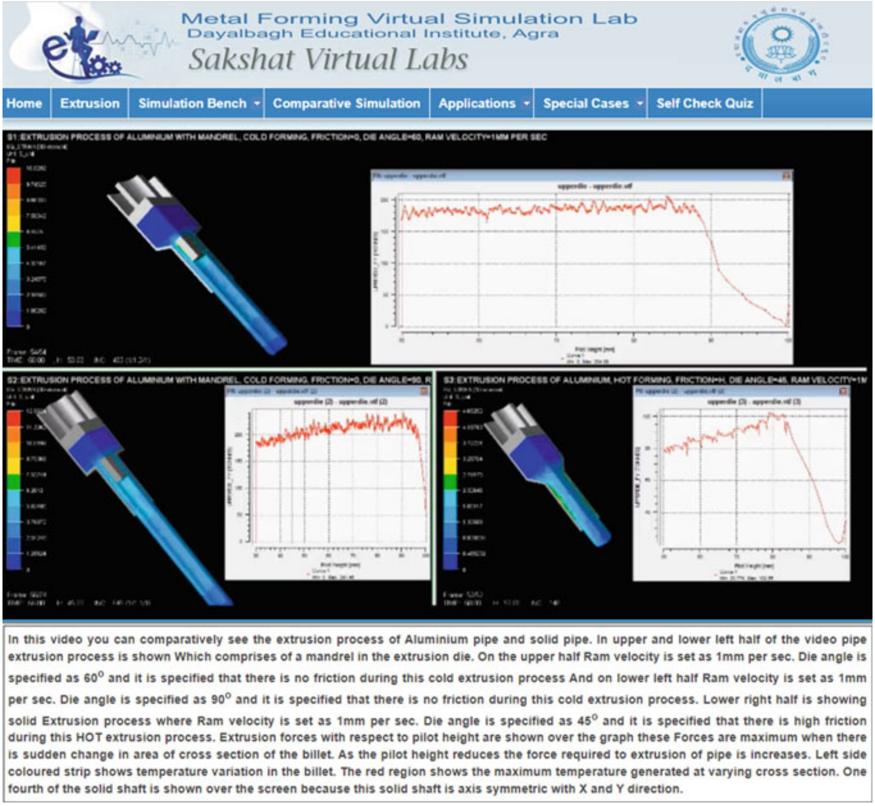


Fig. 9 Comparative simulation for effect of temperature

and submit the self-check quiz to find the result. It is not mandatory to answer all questions. Students can answer some questions or none of the questions to submit the self-check quiz.

After submitting the self-check quiz a result sheet is displayed. This result sheet displays the result of the student that he/ she scores for selecting the answer in self-check quiz. This result sheet displays the question number only for those the student has chosen answer, student selected answers, and the correct answers. Student score is also displayed that explains how many answers are correct of the student. Student can compare their answers with correct answers to analysis his/her learning capability with this virtual simulation lab. Students can take printout of result sheet (Fig. 11).

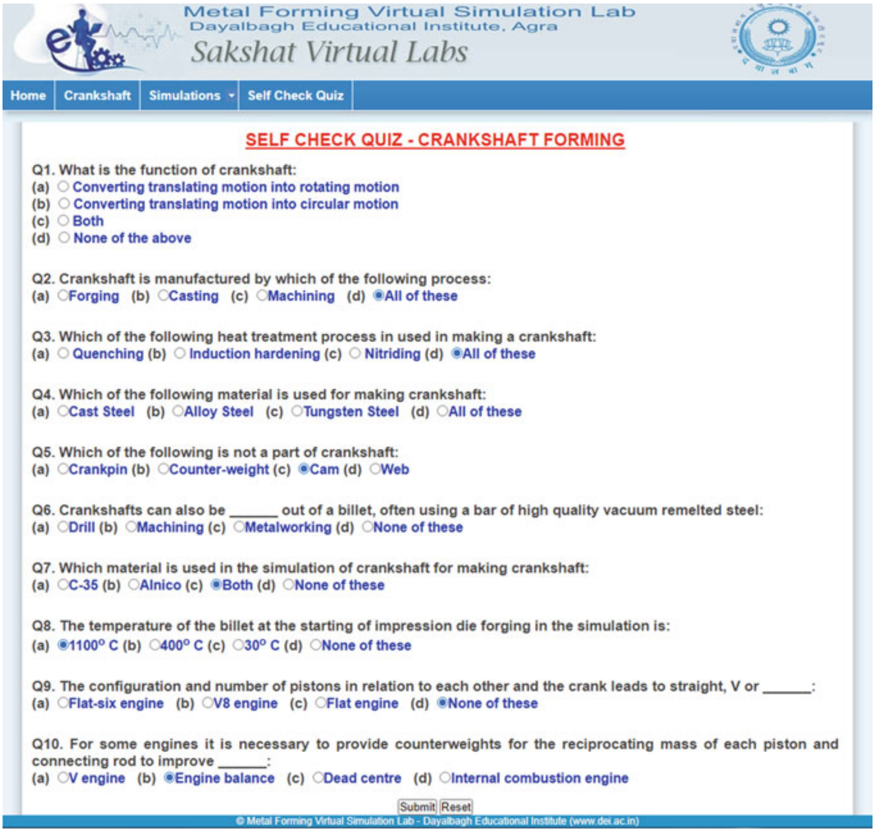


Fig. 10 Self-check quiz for crankshaft forming

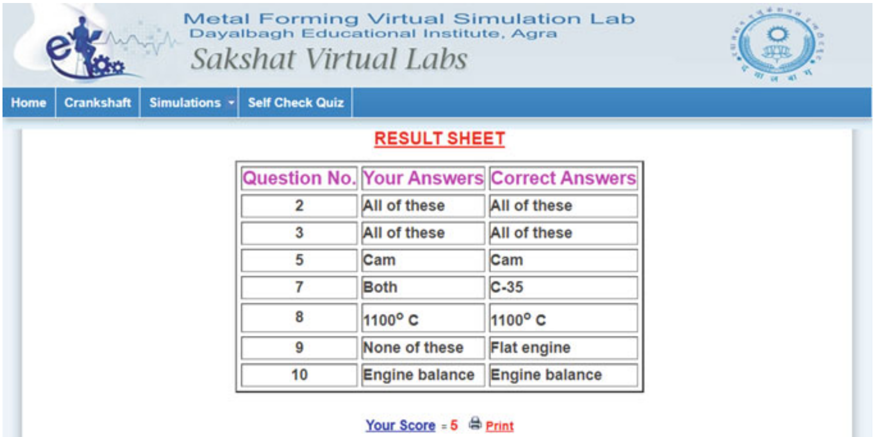


Fig. 11 Result sheet for crankshaft forming

7 Conclusion

This Virtual Simulation Lab is providing e-education and virtual training for students with web-based automatic evaluation mechanism for e-learning and self-study. This shows that learning with virtual labs increases the creativity and knowledge of students. This virtual simulation lab can be accessed with web address http://14.139.245.230/mfvlab/home.php.This virtual simulation lab provides teaching and training facilities to the students and also to the research scholars without any assistance of teacher/ trainer. It has presently more than 97,000 users and has more than 950 simulations. Each simulation is supported by adequate theory and self-check quizzes with automatic evaluation that display score of the student got in self-check quiz. The interactive simulation benches provide facility to test knowledge of students while changing the different parameters. All these features assist student in improving their knowledge with better understanding from observation and analysis of comparative simulations and simulation benches.

References

1. Harms U (2000) Virtual and remote labs in physics education, Physics teaching in engineering education, PTEE2000, Budapest
2. Achakulvisut T (2020) Improving on legacy conferences by moving online. eLife 9:e57892
3. Sharma RS (2016) Technology enabled learning of metal forming processes for engineering graduates using virtual simulation lab. SAGE Publications, Int J Mech Eng Educ 44(2):133–147 (2016)
4. Petridis V, Kazarlis S, Kaburlasos VG (2003) Aces: an interactive software platform for self instruction and self-evaluation in automatic control systems. IEEE Trans Educ 46(1):102–110
5. Faltin NB (2002) Distributed team-learning in an internet- assisted laboratory. In: International Conference on Engineering Education, Manchester, UK

Performance Evaluation of Metaheuristic Algorithms for Optimal Exoskeleton Controller Design

Anusha Kumar and Puneet Mishra

Abstract Exoskeletons are used for a wide variety of applications, with one of the most important being assistance and rehabilitation for those who have lost control of their limbs. Exoskeleton controllers must be robust and show stable behavior under varied conditions. Intelligent control schemes prove to be a good option for achieving this. However, there are no standard practices in literature to tune adaptive intelligent controllers which guarantee their performance. Keeping this in mind, an attempt has been made in this work to optimize the parameters of an adaptive fuzzy control scheme applied to a dynamic exoskeleton focusing on the knee and ankle joints. Three different algorithms, namely—Grasshopper Optimization Algorithm (GOA), Grey Wolf Optimizer (GWO), and Moth-Flame Optimization (MFO) are used for this purpose. The performance of the tuned controllers has been investigated and thorough comparative studies have been drawn. Based on the integral time of absolute error values (ITAE), it is concluded that GWO shows an overall better performance over the other two.

Keywords Metaheuristic algorithms · Swarm optimization · Exoskeleton · Adaptive fuzzy control · Direct adaptive control

1 Introduction

Exoskeletons [1] have gained immense recognition over the past few decades due to their application in multiple fields such as military, industrial, and medical. One of their most prominent uses includes assistance and rehabilitation for those who have suffered injuries or strokes and have lost partial/complete control of their limbs.

A. Kumar (✉) · P. Mishra
Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, Pilani 333031, India
e-mail: f20170195@pilani.bits-pilani.ac.in

P. Mishra
e-mail: puneet.mishra@pilani.bits-pilani.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_11

Multiple control schemes have been discussed in literature for exoskeleton controller design. PID [2, 3] controllers are commonly used but prove to be inefficient as systems become more complex and non-linear. Another alternative is the use of coupled oscillators [4]; however, these may be highly system parameter dependent. Lyapunov stability criterion [5] has been widely used to test system stability, but the criterion does not always ensure performance. Thus, there is a need for an intelligent control system to get a good balance of stability and performance. However, there exist no standard methods in literature for designing such controllers.

Recent works have shown the immense potential of global optimization algorithms for tuning parameters in multiple fields such as buildings, aircraft, machinery, and various artificial intelligence problems [6]. In particular, nature-inspired metaheuristic algorithms have become one of the most popular choices for the same. Metaheuristics involve a degree of randomness, allowing them to avoid local optima entrapment. Metaheuristics also offer flexibility because they solve problems only by looking at the inputs and outputs. Another reason for their popularity can be attributed to a reduction in the time taken for obtaining acceptable solutions. The "trial and error" approach is the reason for this reduction and performs well even in very complex scenarios. These solutions may not be the most optimum ones, but the idea behind such algorithms is that they are efficient, practical, and applicable to many problems [7].

Metaheuristics can be roughly divided into two categories—evolution-based and swarm intelligence-based [8]. The former is based on evolution in nature. One of the most common examples in this domain is the Genetic Algorithm (GA) [9] which is based on the Darwinian theory of evolution. Other evolution-based algorithms include genetic programming [10] and differential evolution [11].

The latter is based on modeling the swarm behavior of various species. Due to fewer number of controlling parameters, these are less complex than evolution-based algorithms. Search agents navigate using the simulated collective and social intelligence of creatures. Some commonly used swarm-based algorithms are particle swarm optimization (PSO) [12], Cuckoo Search Algorithm (CSA) [13], Moth-Flame Optimization Algorithm (MFO) [14], Grey Wolf Optimizer (GWO) [15], Grasshopper Optimization Algorithm (GOA) [16]. Further, each swarm optimization consists of two phases—exploration and exploitation. Exploration refers to searching the entire available search space to find the most optimal solution, whereas exploitation focuses on the area around the most optimal solution for further improvement.

In this paper, three metaheuristic algorithms, namely GWO, GOA, and MFO are employed to optimize the parameters of an adaptive fuzzy control scheme (AFCS) [17]. The controller is tested under three conditions for comparison: namely nominal condition, parametric uncertainty, and introduction of disturbances. The optimization problem is modeled such that it minimizes the ITAE value for trajectory tracking, which is the parameter used for comparison. ITAE integrates the absolute error over time. As a result, errors existing for a more extended period are given more weightage than the initial ones—resulting in systems achieving stability faster than those optimized using integral absolute error (IAE) or integral squared error (ISE). The remaining paper is structured as follows: the dynamic model and control scheme employed is described in Sect. 2, followed by an introduction of the algorithms in Sect. 3. Results are discussed in Sect. 4 and conclusions are presented in Sect. 5.

2 System Description

2.1 Knee-Ankle Modeling

This work deals with the dynamic modeling of the knee and ankle joints adapted from [4] and is shown in Fig. 1. The knee and ankle joints are represented by 1 and 2, respectively. The subject is assumed to be in a sitting position, with the range of movements for the knee and ankle joint took as 0 $0^\circ \le \theta_1 \le 134.65^\circ$ and 0 $0^\circ \le \theta_2 \le 49.85^\circ$, respectively. The upper limits for the knee and ankle joint correspond to their maximum flexion and maximum plantarflexion, respectively [4].

The joints are modeled separately, and thus there are two degrees of freedom with each joint generating two separate trajectories. They may be given as

$$\ddot{\theta}_i = I_i^{-1}\left(-m_i g l_i \sin\theta_i - d_i\dot{\theta}_i + \tau_i\right) \tag{1}$$

where, $i = 1, 2$ represent the trajectories for knee and ankle joints, respectively. I_i refers to the moment of inertia (modeled as cylindrical links) of the joint, having mass m_i. l_i is the length, d_i is the damping constant and g is the gravitational acceleration constant. $\ddot{\theta}_i$, $\dot{\theta}_i$, and θ_i represent the angular acceleration, velocity, and position of the joint. τ_i is the net torque applied to the system, and is taken as a sum of human torque and assisting torque.

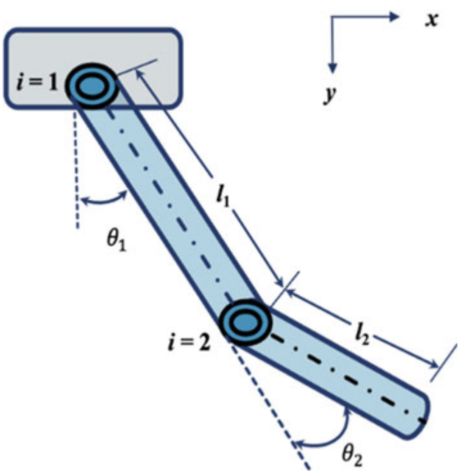


Fig. 1 Dynamic model

2.2 Control Scheme

Torque input (τ) to the dynamic model is a sum of human (τ_{H}) and assisting torque (τ_{A}). The latter is generated by the AFCS, whereas human torque is generated using a PID controller [17]. Figure 2 shows the representative model of the control structure.

The fuzzy controller is based on the Takagi–Sugeno inference model [18], which takes a scaled sum of the error (θ_e) and its first-order derivative ($\dot{\theta}_e$) as input to the system. σ and ρ are the scaling constants and the input (Λ_e) is given as

$$\Lambda_e = \sigma\left(\theta_e + \rho\dot{\theta}_e\right) \tag{2}$$

The scaled input, Λ_e varies between -1 to 1. The rule base consists of five rules, which are stated in (3). $\tilde{\Lambda}_e^k$ is the membership function defined for input Λ_e and $\tilde{\tau}_\alpha^k$ is the singleton output for the corresponding rule.

$$\text{If}\,\Lambda_e \text{ is } \tilde{\Lambda}_e^k \text{ then } \overline{\tau}_A \text{ is } \tilde{\tau}_\alpha^k \tag{3}$$

Triangular membership functions are used, which are uniformly divided over $[-1, 1]$ and are symmetrical. Singleton consequents for the membership functions defined in (3) are varied in run time according to the adaptive law given in (4)

$$\dot{\tilde{\tau}}_\alpha^k = \left(\gamma\Lambda_e\mu^k\right)/\left(\Sigma\mu^k\right) \tag{4}$$

where γ is the learning parameter. Finally, the crisp output given in (4) is multiplied by a constant scaling factor η, which results in the assisting torque for each joint. The resulting assisting torque for a single controller is given in (5)

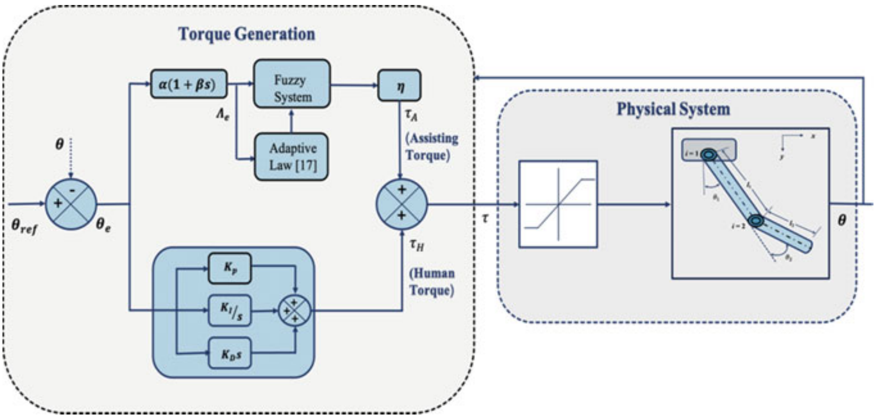


Fig. 2 Controller Implementation (Single Joint)

$$\tau_A = \eta \sum \left(\frac{\mu_k \tilde{\tau}_\alpha^k}{\Sigma \mu_k} \right) \tag{5}$$

3 Algorithms

3.1 Grasshopper Optimization Algorithm (GOA)

The swarming activity of grasshoppers in their nymph phase inspires GOA [16]. The main characteristics during this phase are the slow movement and small steps taken to seek a food source. During this algorithm, the fittest grasshopper is assumed to be the target, as the global optima are unknown. This target may change over iterations, and the swarm moves toward it accordingly. A unique feature in this algorithm is the introduction of a "comfort zone", which is where there is no attraction or repulsion toward any other grasshopper. This zone is gradually reduced over iterations to ensure that the swarm converges toward a solution and is governed by a factor c, referred to in [16]. The algorithm is given as [16]

Initialize swarm X_k *(k* ϵ *[1, number of search agents (K)])*
Initialize max, min and maximum iterations
Calculate fitness of X_k $\forall$ *k*
X_α *= fittest agent*
while *(m < Maximum iterations)*
  Update c using $c = max - m\frac{max-min}{M}$
  for *each* X_i *(such that i* ϵ *[1, K])*
    Normalize distances between grasshoppers in [1,4]
    Update position of X_i *using*
    $X_i^d = c\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} c\,\frac{ub_d - lb_d}{2}\, s\left(\left|x_j^d - x_i^d\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + \hat{T}_d$ *[16]*
    Bring X_i *back if position is outside the boundaries*
  end for
  Update X_α
  m=m+1
end while
Return X_α

3.2 Grey Wolf Optimizer (GWO)

GWO [15] closely mimics the behavior of grey wolves. The algorithm focuses specifically on two aspects—their chain of command and hunting pattern. Each pack consists of the alpha (leader of the pack), beta (second in command), delta (subordinate to alpha and beta), and the omegas (lowest ranking members of the pack). This hierarchy is implemented in the algorithm as well—the best or fittest solution is known as alpha, and it is saved along with the beta and delta solution in every iteration. Hunting consists of three stages—tracking the prey, encircling it to stop its movement, and attacking it. The algorithm for the *m*th iteration is given as below [15]

Initialize population X_k ($k \in$ [1, number of search agents])
Initialize random parameters [15]
X_α = fittest agent
X_β = 2[nd] fittest agent
X_δ = 3[rd] fittest agent
***while** (m < Maximum iteration)*
    ***for** each agent*
        Update the position of the current search agent using $\vec{X}(m+1) = \frac{\vec{X_1}+\vec{X_2}+\vec{X_3}}{3}$
        ($\vec{X_1}$, $\vec{X_2}$ and $\vec{X_3}$ may be referred to from [15])
    end for
    Update random parameters
    Calculate the fitness of X_k $\forall$ k
    Update X_α, X_β and X_δ
    m = m+1
end while
return X_α

3.3 Moth-Flame Optimization (MFO)

MFO [14] is based on the travel mechanism used by moths to travel in a straight line, called transverse orientation. This is very effective for traveling over long distances. However, moths often confuse the artificial light (such as streetlights) for the moon, which causes them to spiral inwards and eventually converge with the light, as the former is much closer. MFO, as defined in [14], is given as.

X = f ( );
***while** P(X) is false*
X = R(X)
end

The purpose of function f is to generate the earliest solutions and calculate the corresponding objective function values. The default is given as in [14]. Till the *P* function returns true, the *R* function is executed repeatedly. Once the *R* function is terminated, the best moth is returned as the best solution. The *R* function is given as [14]

Update flame number using: $round\ (N - l * \frac{N-1}{M})$ *[14]*
OX = Fitness_Func (X);
***if** 1^st^ iteration*
T = sort (X);
OT = sort (OX);
else
T = sort (X_{m-1}, X_m);
OT = sort $(X_m,\ X_{m-1})$;
end
***for** i = 1: n*
***for** j = 1: d*
Update parameters [14]
Calculate D using $D_i = |\, T_j - X_i \,|$ *w.r.t corresponding moth*
Update X(i, j) using $X_i = D_i * e^{bt} * \cos(2\pi t) + T_j$ *w.r.t corresponding moth*
end
end

4 Results and Discussion

This section presents a comparative analysis of the three algorithms on the basis of ITAE values. All simulations are carried out in MATLAB 2020a. Values for the dynamic model of the knee-ankle joint are given in Table 1. g is taken as 9.8 m/s^2.

Table 1 Values for knee-ankle orthosis

Parameter	Shank	Foot
Mass (m)	2.80 kg	1.17 kg
Length (l)	0.2 m	0.08 m
Moment of Inertia (I)	0.075 kg-m^2	0.012 kg-m^2
Damping coefficient (d)	0.4 m/s^2	0.6 m/s^2

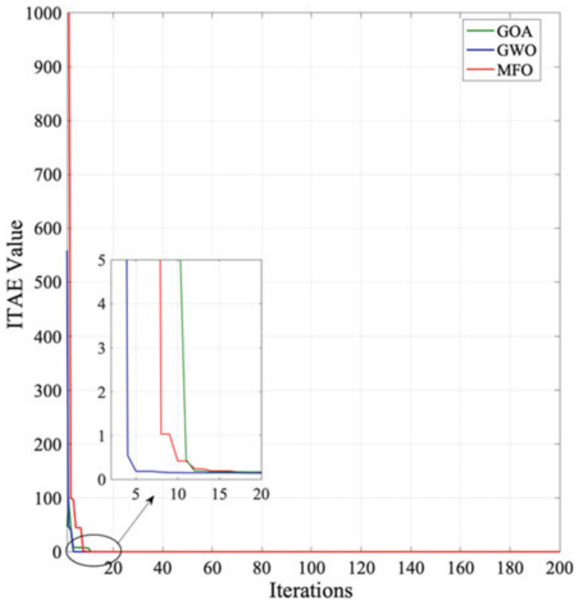


Fig. 3 Convergence curves for GOA, GWO and MFO

4.1 Convergence Curves

For each algorithm, controller was tuned for trajectory tracking with 50 search agents and 200 iterations with the aim of minimizing the sum of ITAE values for knee and ankle joints. The period of each simulation is taken as 80 s. The convergence curves for all have been shown in Fig. 3. The ITAE values obtained are 0.1307, 0.1305, and 0.1292 for GOA, GWO, and MFO, respectively. As can be seen from Fig. 3, GWO shows the fastest exploration, whereas the exploitation period is similar for all three.

4.2 Parametric Uncertainty

The second case considered is parametric uncertainty. All parameters of the physical system are varied, and the controller is tested for those conditions. Figure 4 shows a comparative analysis of the ITAE values for system parameter variation from − 20 to + 20%. While all three algorithms maintain relatively good performance, it is seen that MFO has the least variation in ITAE values, which is 17.59%, whereas GOA and GWO both show 18% variation. Despite the slight difference, it may be concluded that MFO shows the most stable performance.

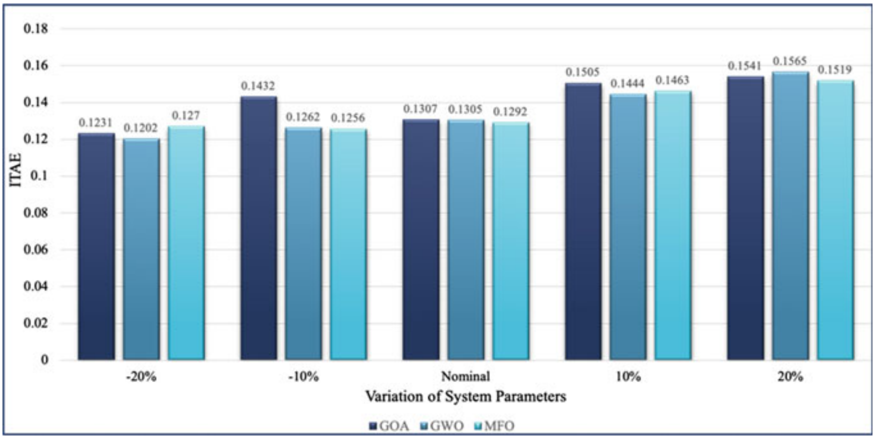


Fig. 4 Parametric uncertainty comparison

4.3 Disturbance Analysis

A step signal of three units is introduced to the torque input of both the joints at $t = 30$ s. The error variation is given in Fig. 5. It can be seen that GWO maintains its performance, whereas GOA and MFO show significant system deviations. This is reflected in their ITAE values—0.1443 for GWO, 34.47 for GOA, and 31.14 for MFO. This can be attributed to the fact that GOA and MFO cannot sustain unmodeled dynamics, whereas GWO can show quick recovery. As the ankle joint shows deviation, the torque response for the ankle joint will be discussed.

It can be seen in Fig. 6, the corrective action in the case of GWO is taken by the assisting device, which is in line to assist the patient. GOA and MFO, on the other hand, result in a combination of human and assisting torque and show performance degradation.

5 Conclusion

In this paper, a performance comparison between GOA, GWO, and MFO for optimizing the parameters of AFCS for a knee and ankle joint exoskeleton was carried out. All three algorithms show impressive results, with MFO showing the minimum ITAE value of 0.1292. In terms of parametric uncertainty, all algorithms show variations close to 18% from the nominal conditions. However, in the case of disturbance analysis, GWO shows a significantly better result, with an improvement of 99.6% over GOA and 95.5% over MFO. Thus, it may be concluded that GWO is overall best suited for tuning the parameters of the considered adaptive fuzzy control scheme.

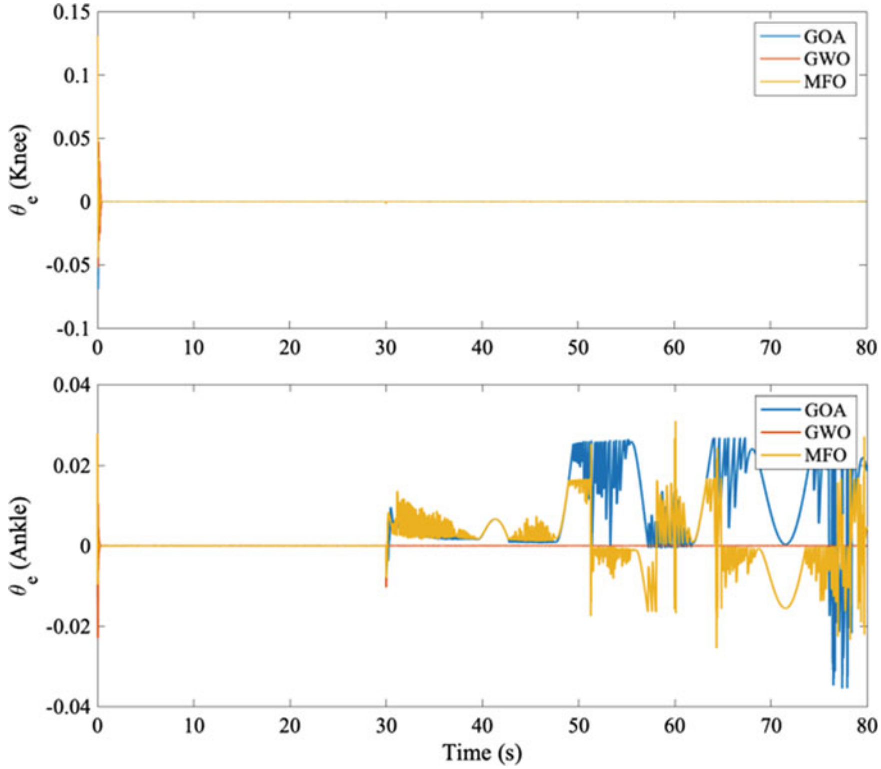


Fig. 5 Trajectory tracking for both joints

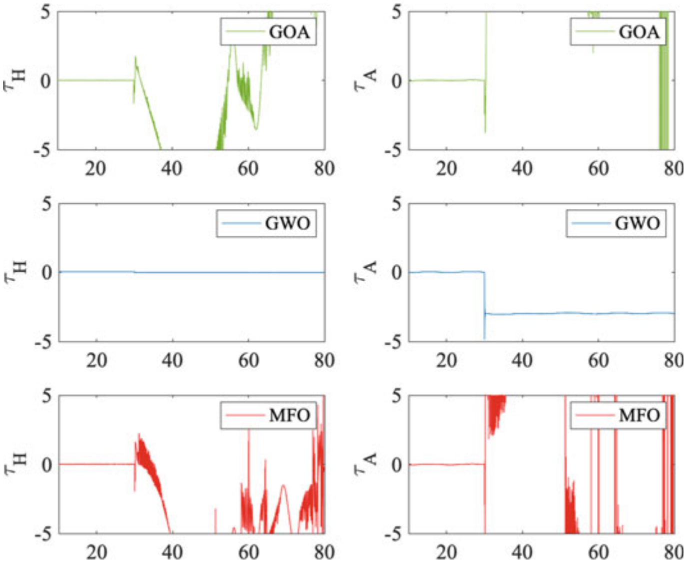


Fig. 6 Human and assisting torque for ankle joint

References

1. Veneman JF, Kruidhof R, Hekman EEG, Ekkelenkamp R, Van Asseldonk EHF, van der Kooij H (2007) Design and evaluation of the LOPES exoskeleton robot for interactive gait rehabilitation. IEEE Trans Neural Syst Rehabil Eng 15(3):379–386
2. Al Rezage G, Tokhi MO (2016) Fuzzy PID control of lower limb exoskeleton for elderly mobility. In: IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), pp 1–6
3. Yu W, Rosen J (2013) Neural PID control of robot manipulators with application to an upper limb exoskeleton. IEEE Trans Cybern. 43(2):673–684
4. Ajayi MO (2016) Modelling and control of actuated lower limb exoskeletons: a mathematical application using central pattern generators and nonlinear feedback control techniques. In: General Mathematics - Université Paris-Est
5. Liu X (2009) Stability analysis of switched positive systems: a switched linear copositive lyapunov function method. IEEE Trans Circ Syst II Exp Briefs 56(5):414–418
6. Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp swarm algorithm: a bio-inspired optimizer for engineering design problems. Adv Eng Softw 114:163–191
7. Yang XS (2010) Nature-inspired metaheuristic algorithms. Luniver Press
8. Luthra I, Chaturvedi SK, Upadhyay D, Gupta R (2017) Comparative study on nature inspired algorithms for optimization problem. In: International conference of Electronics, Communication and Aerospace Technology (ICECA)
9. Yuan X, Elhoseny M, El-Minir HK, Riad AM (2017) A genetic algorithm-based, dynamic clustering method towards improved WSN longevity. J Netw Syst Manag 25(1):21–46
10. Tran B, Xue B, Zhang M (2016) Genetic programming for feature construction and selection in classification on high-dimensional data. Memet Comput 8(1):3–15
11. Das S, Mullick SS, Suganthan PN (2016) Recent advances in differential evolution–an updated survey. Swarm Evol Comput 27:1–30

12. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95—International Conference on Neural Networks, vol. 4, pp 1942–1948
13. Yang XS, Deb S (2009) Cuckoo search via Lévy flights. In: World Congress on Nature & Biologically Inspired Computing (NaBIC)
14. Mirjalili S (2015) Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm. Knowl Based Syst 89:228–249
15. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
16. Saremi S, Mirjalili S, Lewis A (2017) Grasshopper optimisation algorithm: theory and application. Adv Eng Softw 105:30–47
17. Kumar A, Mishra P (2020) Analysis of an adaptive fuzzy control scheme: application to a lower limb exoskeleton. In: 17th IEEE India council international conference
18. Lin CM, Hsu CF (2003) Self-learning fuzzy sliding-mode control for antilock braking systems. IEEE Trans Control Syst Technol 1(2):273–278

Nonlinear PID Controller for Three-Link Robotic Manipulator System: A Comprehensive Approach

Jitendra Kumar, Devbrat Gupta, and Vishal Goyal

Abstract The main objective of this work is to showcase how a Nonlinear Proportional plus Integral plus Derivative (NPID) controller can be utilized to analyze the output of a highly coupled multi-input multi-output (MIMO) nonlinear system called the Three-Link Robotic Manipulator System for setpoint monitoring or trajectory tracking performance. In general, the three-link planar robots are employed in the production sector and are exposed to a variety of disturbances, such as heavy equipment movements, manipulator's arm variations, a loud industrial climate, and so on. In order to cater for these limitations, A proposed NPID controller is incorporated into the three-link revolute joint robotic manipulator system and its gain parameters are tuned using the GA (Genetic Algorithm) to minimize the weighted sum of the integral of absolute error (IAE) signal. Based on this chosen objective function, the proposed NPID controller is then compared to NPI, PID, and PI controllers for testing trajectory tracking accuracy in order to conduct a complete comparative performance analysis.

Keywords PID · PI · NPID · NPI · GA · Three-link robotic manipulator

1 Introduction and Literature Studies

Robotic manipulators are being developed and industrialized for seven decades to replace humans in hazardous manufacturing environments. Using robot manipulators instead of a conveyor system during the process of material movement that shifts the target position is more cost-effective. This is because the robot manipulator only requires thin material, which is lighter, more energy-efficient, has a compact actuator, is easier to operate, and is less costly to manufacture. Robotic manipulators have recently been used in medical devices as well as military missions due to their ability to increase accuracy and robustness [1]. These are nonlinear unpredictable plants with mechanically coupled arms and segment joints that capture and pass material

J. Kumar (✉) · D. Gupta · V. Goyal
Department of Electronics and Communication Engineering, GLA University, Mathura 281406, India
e-mail: jitendra.kumar@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_12

along a fixed path. One of the most common uses of industrial robotic manipulators is to pick and place materials. Welding, assembling, manufacturing, painting, and other automotive-related uses of industrial manipulators include robotically-assisted automated surgery, management of toxic and bio-hazardous ingredients, and many more [2]. Proportional plus Integral plus Derivative (PID) controllers, on the other hand, are utilized and have been deployed in 90% of process control industries, but PID controllers struggle to provide efficient control action for dynamic nonlinear coupled systems like manipulators[3]. In the early '90s, John Holland found a method named genetic algorithm to solve this mystery, genetic algorithm (GA) is an optimization tool for solving both controlled and unhindered optimization problems that use a natural selection mechanism similar to biological evolution [4]. The performance of feedback-driven complex nonlinear systems is hugely affected by a perfect control strategy. Researchers and scientists are continuously looking for the smartest adaptive servo and regulatory mode control solutions in complicated systems like revolute joint robotic manipulators. Some of the most current research studies are further described.

Kazemian, in his paper [5] presents a comparison of a self-organizing fuzzy PID (SOF-PID) controller with a regular PID controller for route tracking on a highly nonlinear two-link revolute joint robot arm. The SOF-PID controller was found to follow the prescribed direction more strictly and efficiently than the traditional PID controller in this study. The classical approach was incorporated to tune the gain parameters of the PID controller, and then the SOF scheme was employed to alter the gains in real-time. Because of this improvement, the SOF-PID controller's performance trajectories for the trajectory tracking task followed the appropriate direction more accurately than the self-organizing fuzzy controller (SOFC) as well as PID. He suggested to utilized Lyapunov stability theory on a PID controller for studying semi-global asymptotic stability in order to operate a two-DOF robot arm. The gain parameters of a PID were tuned online in this research, and real-time investigational results were showcased to demonstrate the utility of the proposed method [6]. In his paper, Rana, K.P.S., [7] has suggested incorporating a new fractional-order FPID (FOFPID) controller and its hybrid constraint combinations, which have been tuned by GA. The suggested control methods have been evaluated on nonlinear time-delayed, fractional-order, as well as fractional-order time-delayed processes. While fractional-order calculus has been around 300 years, its applications in areas such as control and simulation are currently being researched. As the degrees of freedom rise, fractional-order simulation helps the models to be identical to the standard system dynamics [8]. Hultmann Ayala [9] has introduced the IO-PID controller in 1911 to incorporate into automatic ship steering. Due to its simplicity of design and execution, PID has remained relevant over time. A proportional, integral, and derivative control actions of a PID controller, have a definite physical value, making tuning spontaneous.

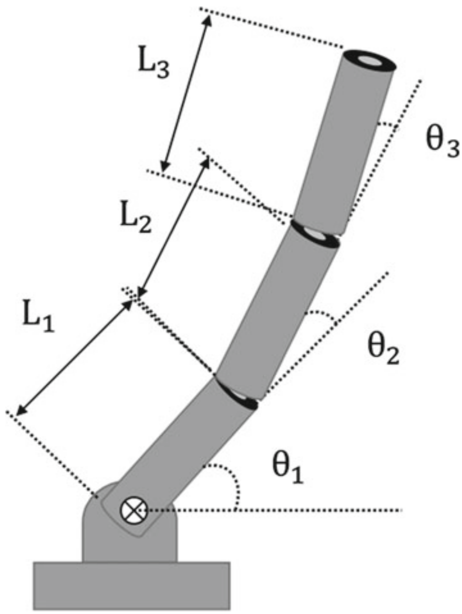
2 Problem Formulation and Mathematical Modeling of a Three-Link Robotic Manipulator System

Because of their lightweight, low cost, larger work capacity, high speed of operation, higher power usage, and a vast number of applications; 3-link robotic manipulators have received considerable attention. Maximizing stiffness to reduce vibration and achieve good position and velocity accuracy is a key element in the design of the most current robotic manipulators. Heavy material and a broad design are used to achieve high stiffness [10–13].

A three-link revolute joint planar manipulator has three revolute joints (RRR), and the process workspace is confined to a plane. There are three degrees of freedom (DOFs) in the specified system, one for each joint. It contains three insignificant mass connections. A large portion of the mass of these linked manipulators is expected to be dispersed at the joints, which is a plausible hypothesis given how light the mechanical connections are compared to each joint heavy actuator. Three torques are given to the joints of link-1, 2, and, 3, respectively. The measurements of joint angles θ_1, θ_2, and θ_3 link-1, 2, and, 3, respectively, define the manipulator outputs. In order to determine the error signal value for each link, the manipulator outputs are subtracted from the reference value of joint angles [2, 14]. Figure 1 describes the schematic representation of a three-link revolute joint robotic manipulator.

The expression for the dynamical modeling of the three-link robotic manipulator [2] is showcased in Eq. (1) as,

Fig. 1 Three-link robotic manipulator



$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}\begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \\ \ddot{\theta}_3 \end{bmatrix} + \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} + \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} \tag{1}$$

where terms A_{11}–A_{33} denotes the second-order derivative of the angular position of three links $\ddot{\theta}_1$, $\ddot{\theta}_2$, and $\ddot{\theta}_3$ as expressed in Eqs. (2)–(10). C_1, C_2, and C_3 denotes the centrifugal terms for link-1, 2, and, 3, respectively as expressed in Eqs. (11)–(13). R_1, R_2, and R_3 denotes Coriolis terms for link-1, 2, and, 3, respectively as expressed in Eqs. (14)–Eq. (16). g_1, g_2, and g_3 denotes acceleration due to gravity for link-1, 2, and, 3, respectively as expressed in Eqs. (17)–(19).

$$\begin{aligned} A_{11} = {} & (M_1 + M_2 + M_3)L_1^2 + (M_2 + M_3)L_2^2 + (M_3)L_3^2 \\ & + 2M_3L_1L_3\cos(\theta_2 + \theta_3) + 2(M_2 + M_3)L_1L_2\cos(\theta_2) \\ & + 2M_3L_2L_3\cos(\theta_3) \end{aligned} \tag{2}$$

$$\begin{aligned} A_{12} = {} & (M_2 + M_3)L_2^2 + (M_3)L_3^2 + M_3L_1L_3\cos(\theta_2 + \theta_3) \\ & + (M_2 + M_3)L_1L_2\cos(\theta_2) + 2M_3L_2L_3\cos(\theta_3) \end{aligned} \tag{3}$$

$$A_{13} = (M_3)L_3^2 + M_3L_1L_3\cos(\theta_2 + \theta_3) + M_3L_2L_3\cos(\theta_3) \tag{4}$$

$$\begin{aligned} A_{21} = {} & (M_2)L_2^2 + (M_3)L_2^2 + (M_3)L_3^2 + M_3L_1L_3\cos(\theta_2 + \theta_3) \\ & + M_2L_1L_2\cos(\theta_2) + M_3L_1L_2\cos(\theta_2) + 2M_3L_2L_3\cos(\theta_3) \end{aligned} \tag{5}$$

$$A_{22} = (M_2)L_2^2 + (M_3)L_2^2 + (M_3)L_3^2 + 2M_3L_2L_3\cos(\theta_3) \tag{6}$$

$$A_{23} = (M_3)L_3^2 + M_3L_2L_3\cos(\theta_3) \tag{7}$$

$$A_{31} = (M_3)L_3^2 + M_3L_1L_3\cos(\theta_2 + \theta_3) + M_3L_2L_3\cos(\theta_3) \tag{8}$$

$$A_{32} = (M_3)L_3^2 + M_3L_2L_3\cos(\theta_3) \tag{9}$$

$$A_{33} = (M_3)L_3^2 \tag{10}$$

$$\begin{aligned} C_1 = {} & -L_1\{M_3L_3\sin(\theta_2 + \theta_3) + M_2L_2\sin(\theta_2) + M_3L_2\sin(\theta_2)\}\dot{\theta}_2^2 \\ & - M_3L_3\{L_1\sin(\theta_2 + \theta_3) + L_2\sin(\theta_3)\}\dot{\theta}_3^2 \end{aligned} \tag{11}$$

$$\begin{aligned} C_2 = {} & L_1\{M_3L_3\sin(\theta_2 + \theta_3) + M_2L_2\sin(\theta_2) + M_3L_2\sin(\theta_2)\}\dot{\theta}_1^2 \\ & - M_3L_2L_3\sin(\theta_3)\dot{\theta}_3^2 \end{aligned} \tag{12}$$

$$C_3 = M_3L_3\{L_1\sin(\theta_2+\theta_3) + L_2\sin(\theta_3)\}\dot{\theta}_1^2 + M_3L_2L_3\sin(\theta_3)\dot{\theta}_2^2 \tag{13}$$

$$\begin{aligned} R_1 &= -2L_1\{M_3L_3\sin(\theta_2+\theta_3) + (M_2+M_3)L_2\sin(\theta_2)\}\dot{\theta}_1\dot{\theta}_2 \\ &\quad - 2M_3L_3\{L_1\sin(\theta_2+\theta_3) + L_2\sin(\theta_3)\}\dot{\theta}_2\dot{\theta}_3 \\ &\quad - 2M_3L_3\{L_1\sin(\theta_2+\theta_3) + L_2\sin(\theta_3)\}\dot{\theta}_1\dot{\theta}_3 \end{aligned} \tag{14}$$

$$R_2 = -2M_3L_2L_3\sin(\theta_3)\dot{\theta}_1\dot{\theta}_3 - 2M_3L_2L_3\sin(\theta_3)\dot{\theta}_2\dot{\theta}_3 \tag{15}$$

$$R_3 = 2M_3L_2L_3\sin(\theta_3)\dot{\theta}_1\dot{\theta}_2 \tag{16}$$

$$\begin{aligned} g_1 &= (M_1+M_2+M_3)gL_1\cos(\theta_1) \\ &\quad + (M_2+M_3)gL_2\cos(\theta_1+\theta_2) \\ &\quad + M_3gL_3\cos(\theta_1+\theta_2+\theta_3) \end{aligned} \tag{17}$$

$$g_2 = (M_2+M_3)gL_2\cos(\theta_1+\theta_2) + M_3gL_3\cos(\theta_1+\theta_2+\theta_3) \tag{18}$$

$$g_3 = M_3gL_3\cos(\theta_1+\theta_2+\theta_3) \tag{19}$$

In the MATLAB environment simulation study, the parameters of the system are considered as follows: masses for link-1, 2, and, 3 are considered as, $M_1 = 0.1$ kg, $M_2 = 0.1$ kg, and $M_3 = 0.1$ kg, respectively. The length of the links of the manipulator system for link-1, 2, and, 3 are considered as $L_1 = 0.8$ m, $L_2 = 0.4$ m, and $L_3 = 0.2$ m, respectively. The acceleration due to gravity is taken as $g = 9.8\ \text{m/s}^2$.

3 Proposed Controller Design

The term "high-performance controller" refers to a controller with a fast torque response time and high precision. To distinguish such controllers, the current controller can use vector control, which can immediately control torque, as well as specific rpm, load disturbance compensation, and motor and machinery parameter protection in the speed controller [15]. Some Controllers are described as follows:

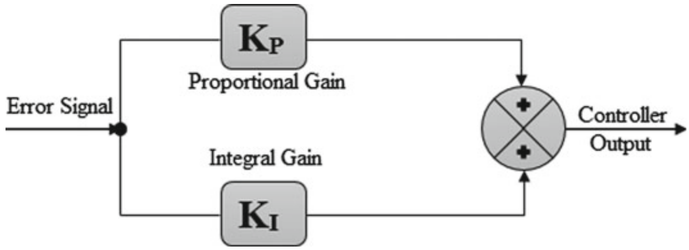


Fig. 2 Block diagram of PI control structure

3.1 Proportional Plus Integral (PI) Controller

Proportional plus Integral (PI) controllers are consisting of two control actions, one is proportional action and the other is integral action. In general, it is a kind of controller that combines proportional and integral control actions into one. As a result, it is known as a PI controller Both proportional and integral control actions are used in the proportional-integral controller. This fusion of two separate controllers results in a more powerful controller that reduces the drawbacks of both. The main purpose of PI control is to reduce or remove steady-state errors at the cost of inaccurate transient responses. The error signal is fed to proportional and integral action and gives combined control signal as controller output [16]. The block diagram of the PI controller is shown in Fig. 2.

3.2 Proportional Plus Integral Plus Derivative (PID) Controller

Proportional plus Integral plus Derivative (PID) control is a closed-loop control mechanism that was used. The additive, integral, and derivative control systems are combined in this control system. Setpoint values are used as signals in this framework (the desired output values). Because of the offset, i.e., the distance (difference) between the output value that exists and the desired output value, the controlled system's response is less acceptable when using only this control system. As a result, we used a Proportional-Integral Derivative (PID) control system that combined the three control systems. The output response of this control system will be fast, with no offset, and will not oscillate since it combines the characteristics of its constituent control systems [17].

The PID controller employs feedback features in the control system to establish the accuracy of a tool system. Three modes of operation are available to the PID control system: proportional, derivative, and integral with their own set of benefits and disadvantages. During implementation, each control system can be utilized alone or in combination. The parameters *P*, *I*, or *D* are always in the specification of a PID

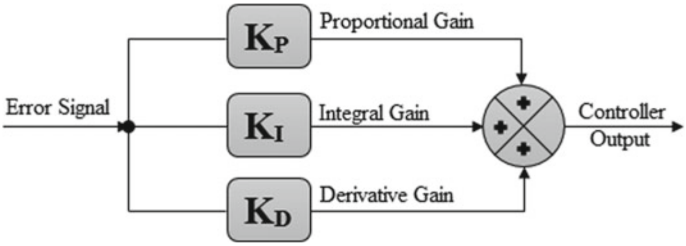


Fig. 3 Block diagram of PID control structure

control system for achieving the required response of the output signal to a given input [18].

Due to its ease of design and execution, PID has remained relevant over time. In addition, a physical value is defined for the gains of the PID controller: the proportional (K_P), the integral (K_I) and the derivative (K_D). The use of proportional action in any system reduces the steady-state error response. Hence in order to eliminate this error in the output response, the integral action comes into the picture and, then subjected to optimization. Finally, the derivative component is used to boost the transfer function and minimize the overshoot generated in the device output. PID operation is tuned for individual output based on the system application by adjusting the gains of these respective controllers [14]. The block diagram of the PID control structure is illustrated in Fig. 3.

3.3 Nonlinear Proportional Plus Integral (NPI) Controller

The benefit of the nonlinear PI controller varies nonlinearly in response to error. The aim of the nonlinear PI controller in this analysis is to reduce the deterioration caused by load variance when the system's load is modified in a steady-state. The bandwidth should be maximally expanded or the load variance should be correctly measured to achieve a power greater than the current PI controller [19]. A controller aims to reduce errors over time. The bandwidth can be comparatively expanded in the part of the spectrum where the error is limited, but narrowed in the part where the error is high so that the bandwidth can be widened under steady-state and disturbances can be remarkably regulated [15]. The block diagram of the NPI control structure is shown in Fig. 4.

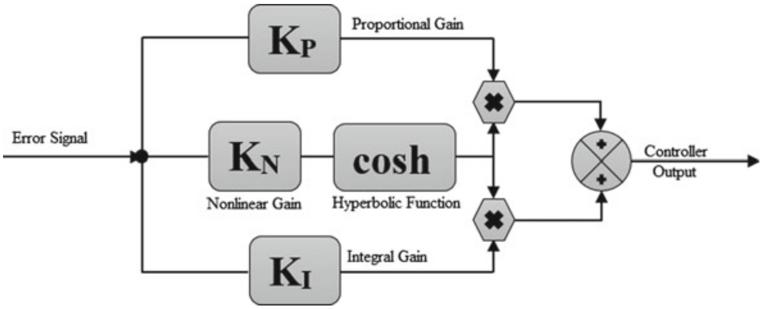


Fig. 4 Block diagram of NPI control structure

3.4 Nonlinear Proportional Plus Integral Plus Derivative (NPID) Controller

For industrial applications, nonlinear PID (NPID) control is regarded as one of the most efficient and simple methods. Any control structure of the following type as shown in Fig. 5 can be used for nonlinear PID controller.

The mathematical equation of NPID control structure can be illustrated as;

$$U_{\mathrm{NPID}}(t) = \left[K_{\mathrm{P}}e(t)f(e) + K_{\mathrm{I}}\int f(e)e(t)\mathrm{d}t + K_{\mathrm{D}}\frac{\mathrm{d}e(t)}{\mathrm{d}t}\right] \tag{20}$$

where $f(e)$ is a nonlinear gain function in terms of error e and further it can be expressed as;

$$f(e) = \cosh(K_N e) \tag{21}$$

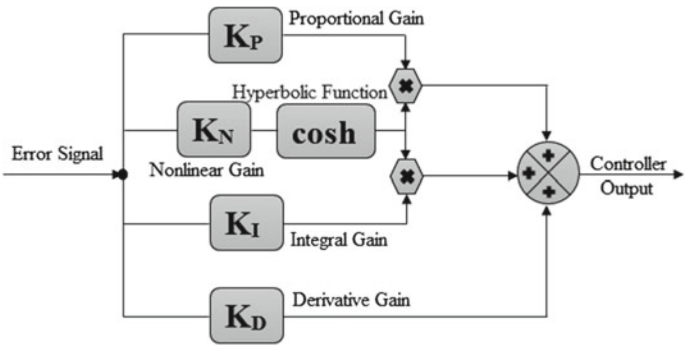


Fig. 5 Block diagram of NPID control structure

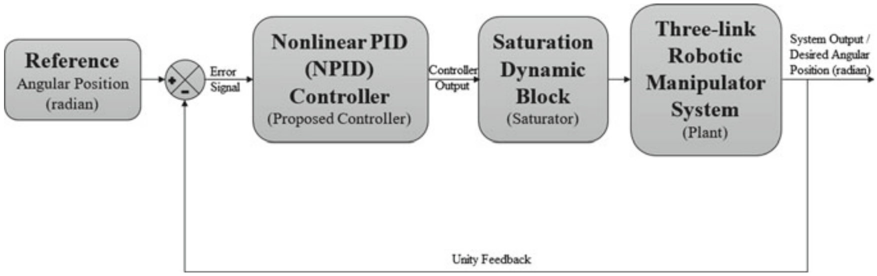


Fig. 6 Complete system illustration with the reference signal, output, and NPID Controller

$$\text{or,}\ \ f(e) = \frac{\exp(K_N e) + \exp(-K_N e)}{2} \tag{22}$$

$$\text{Here, } e = \begin{cases} e;\ |e| \le e_{\max} \\ e_{\max} * \operatorname{sgn}(e);\ |e| > e_{\max} \end{cases} \tag{23}$$

In this research work, for NPID controller parameter K_N and $e_{\max}$ are positive constants taken by the user. The lower limit of $f(e)$ would be 1, when $e = 0$. Here, K_N is optimized by GA algorithm where $e_{\max} = 50$ is assumed.

The gain parameters K_{P}, K_{I}, and K_{D} are time-varying, which may depend on system input, state, or other variables, The proposed controller output is denoted as $U_{\mathrm{NPID}}(t)$ and the error signal is considered as $e(t)$ for this work [20–22].

The overall system with the reference signal, control structure, saturator, and three-link robotic manipulator is shown in Fig. 6.

3.5 Gain Tuning with Genetic Algorithm (GA)

The genetic algorithm is a classical evolutionary algorithm that is based on randomness. By random, we say that in order to use the GA to find a solution, random modifications are made to existing solutions in order to produce new ones. The genetic algorithm (GA) is an efficient, stochastic-based search/optimization approach that imitates biological evolution, as described by Goldberg, Eiben, and others [23–25].

This is an evolutionary philosophy based on the ideas of Darwin. It is a long and progressive process involving tiny, progressive improvements. GA also improves the overtime suggestions significantly before it gets the perfect one. The measures are as follows:

1. population (set of individuals) initialization
2. fitness function (criterion) measurement for each person in the population
3. finish if expiration objectives are satisfied; otherwise, proceed to next stage
4. deciding on a relative (more fit individuals have a higher probability to be selected),

5. cross-pollination and mutation of parents = offspring
6. Proceed with stage 2 after completing the new population (children + selected unchanged individuals).

The central idea of GA selection, like natural selection, is fitness. The fit chromosomes have a higher chance of surviving. Fitness is a parameter that takes into account the performance of the chromosome-represented solution. Each chromosome in the population, in particular, represents the input parameters [26].

Chromosomes form pairs of parents for reproduction during the selection process. Each infant inherits traits from its parents. Effectively, the offspring is a recombination of its parents 'characteristics: Such traits are inherited from one parent and others from the other. Any of the characteristics will mutate in addition to recombination.

Since fitter chromosomes grow more offspring, the fitness of each subsequent generation would improve. A generation will produce a chromosome that will represent a better enough solution to our challenge at some stage.

GA is a simple tool that could be used to tackle a plethora of challenges. There is a broad class of optimization problems that traditional optimization methods like PID controllers struggle to solve. Genetic algorithms are effective algorithms with a nearly perfect solution [27].

The mathematical expression of an objective function to minimize the weighted sum of the integral of absolute error is given in Eq. (24).

$$J_{\mathrm{IAE}} = \int_0^t |e(t)|\mathrm{d}t \tag{24}$$

where J_{IAE} denotes objective function value of IAE and $e(t)$ is the error signal which is minimized.

4 Results and Discussion on the Comparative Performance of the Controllers

All of the results were generated using the MATLAB/Simulink system. For all simulation experiments, a well-configured 64-bit operating system with an Intel Core i5 CPU running at 3.33 GHz, 8 GB RAM, and 1 TB storage capacity is employed. The Runge–Kutta technique with order four is used to determine the solution of a mathematical model of plant dynamics with differential equations, and the sampling time for simulation studies is set at 1 ms. Furthermore, tailored gains are left unchanged throughout the simulation research.

The objective function curve for PID, PI, NPI, and NPID controllers with 100 generations is shown in Fig. 7.

The objective function values for link-1, 2, and 3 as well as overall IAE values for all the three links of the three-link robotic manipulator for NPID, NPI, PID, and PI controllers is shown in Table 1. And it can be clearly observed that the objective

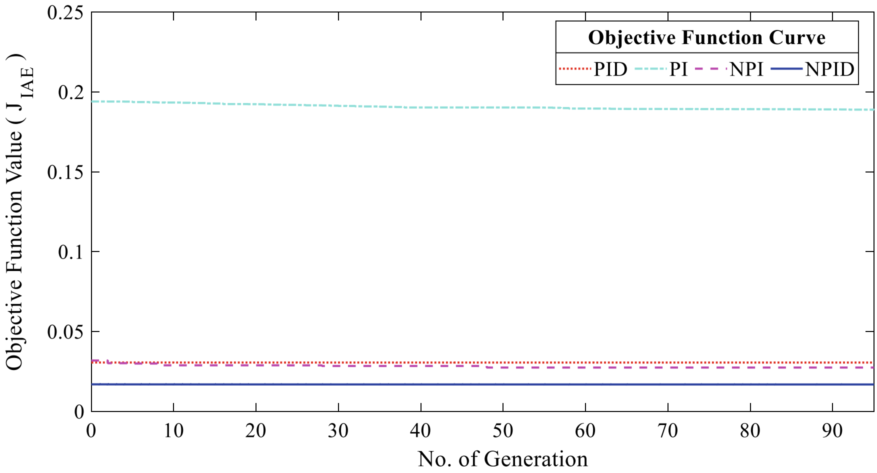


Fig. 7 Objective function curve vs. no. of generation

Table 1 Objective function values for NPID, NPI, PID, and PI controllers

Controllers	Objective function values (J_{IAE})			
	IAE values of link-1	IAE values of link-2	IAE values of link-3	Overall IAE values of link-1, link-2, and link-3
NPID	0.0080	0.0042	0.0046	0.0168
NPI	0.0081	0.0082	0.0111	0.0274
PID	0.0183	0.0060	0.0063	0.0306
PI	0.1377	0.0417	0.0090	0.1884

function value for the NPID controller for the individual as well as all the three-link is minimum among all the controllers which means the NPID controller gives optimum and robust results.

The Bar chart for objective function values for NPID, NPI, PID and PI controllers of Link-1, 2, and 3 of the three-link robotic manipulator is illustrated in Fig. 8.

The tuned controller parameters for Link-1, 2, and 3 for the proposed controller, i.e., NPID controller as well as NPI, PID, and PI controllers are given in Table 2.

Similarly, the trajectory tracking curves for PI, PID, NPI, and NPID controllers incorporated on Link-1, 2, and 3 of the robotic manipulator are shown in Fig. 9.

The PI, PID, NPI, and NPID Controllers' output curve for Link-1, Link-2, and Link-3 of the robotic manipulator is shown in Fig. 10. The Error signal for PI, PID, NPI, and NPID controllers incorporated on Link-1, 2, and, 3 of the robotic manipulator is shown in Fig. 11. The reference signal for PI, PID, NPI, and NPID controllers incorporated on Link-1, 2, and 3 of the robotic manipulator is shown in Fig. 12.

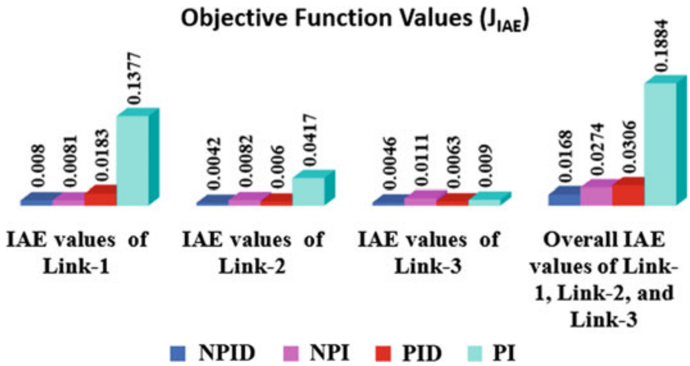


Fig. 8 Bar chart for objective function values for NPID, NPI, PID, and PI controllers of Link-1, Link-2, and Link-3 of robotic manipulator

Table 2 Tuned controllers' parameters

Controllers' gains		Gain values of tuned controller's parameters			
		NPID	NPI	PID	PI
Link-1	K_{P_1}	728.9883	1292.1769	728.9970	1461.3495
	K_{I_1}	88.9983	67.8521	88.9999	0.0650
	K_{D_1}	47.9978	–	47.9985	–
	K_{0_1}	1499.9847	1407.5899	–	–
Link-2	K_{P_2}	733.9836	1186.9601	733.9978	1499.6128
	K_{I_2}	35.9562	67.2142	35.9978	21.1827
	K_{D_2}	16.9865	–	16.7549	–
	K_{0_2}	1499.9622	1252.7776	–	–
Link-3	K_{P_3}	178.9989	1223.7743	178.9999	1165.2849
	K_{I_3}	23.9745	82.7406	23.9996	50.0749
	K_{D_3}	4.9968	–	4.1831	–
	K_{0_3}	1492.9605	1446.8386	–	–

5 Conclusion and Future Scope

A nonlinear proportional plus integral plus derivative (NPID) controller is proposed and successfully tested in simulation to effectively control a nonlinear, multi-input, and multi-output, complex, and coupled 3-link planar rigid robotic manipulator in a virtually simulated industrial environment in the current study. By minimizing the weighted sum of the integral of the absolute error (IAE) signal, a Genetic Algorithm (GA) optimization approach is utilized to adjust the nonlinear, proportional, integral, and derivative gain parameters. Under a simulated industrial environment, intensive simulations were conducted to evaluate and compare the performance of

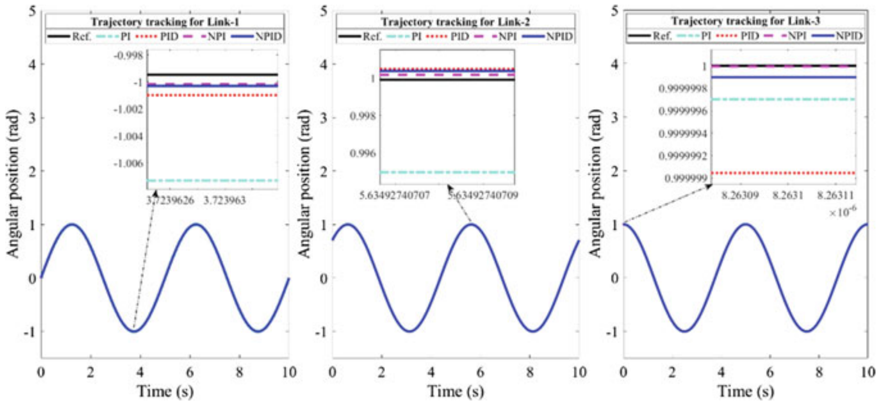


Fig. 9 Trajectory tracking curves for PI, PID, NPI, and NPID controllers incorporated on Link-1, Link-2, and Link-3 of robotic manipulator

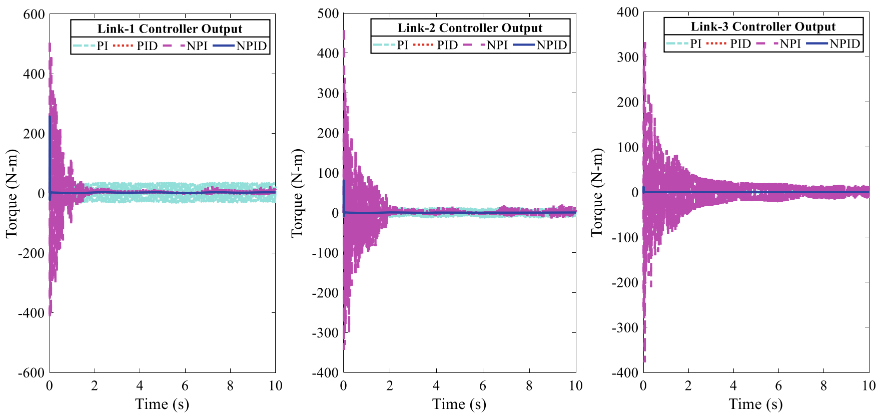


Fig. 10 PI, PID, NPI, and NPID Controllers output curve for Link-1, Link-2, and Link-3 of robotic manipulator

PI (proportional plus integral), PID (proportional plus integral plus derivative), NPI (nonlinear proportional plus integral), and NPID controllers for reference trajectory tracking. For trajectory tracking, the NPID controller outperformed the NPI, PID, and PI controllers by 38.69%, 45.10%, and 91.08% in objective function (OBF) values. It was discovered that, under typical conditions, the proposed NPID controller not only outperformed its PI, PID, and NPI counterparts but also provided a good reference path-following performance with superior and robust performance in an industrial environment. In the future, it is recommended that the results of simulation investigations be empirically confirmed by implementing the proposed control method in real-time.

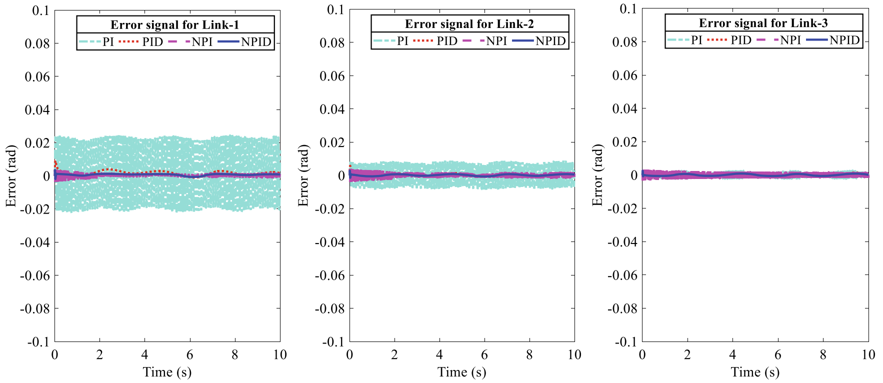


Fig. 11 Error signal for PI, PID, NPI, and NPID controllers incorporated on Link-1, Link-2, and Link-3 of robotic manipulator

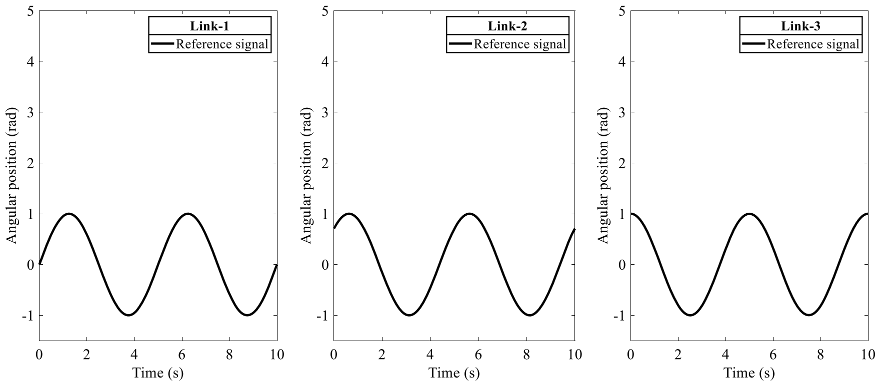


Fig. 12 The reference signal for PI, PID, NPI, and NPID controllers incorporated on Link-1, Link-2, and Link-3 of robotic manipulator

References

1. Kathuria T, Gupta A, Kumar J, Kumar V, Rana KPS (2017) Study of optimization methods for tuning of PID gains for three link manipulator. In: 2017 7th international conference on cloud computing, data science & engineering—confluence
2. Kumar J, Kumar V, Rana KPS (2020) Fractional-order self-tuned fuzzy PID controller for three-link robotic manipulator system. Neural Comput Appl 32:7235–7257. https://doi.org/10.1007/s00521-019-04215-8
3. Gupta D, Goyal V, Kumar J (2019) An optimized fractional order PID controller for integrated power system. presented at the (2019). https://doi.org/10.1007/978-981-13-8461-5_76
4. Holland JH (1984) Genetic algorithms and adaptation BT—adaptive control of Ill-defined systems. Presented at the. https://doi.org/10.1007/978-1-4684-8941-5_21
5. Kazemian HB (2002) The SOF-PID controller for the control of a MIMO robot arm. IEEE Trans Fuzzy Syst 10:523–532. https://doi.org/10.1109/TFUZZ.2002.800658

6. Meza JL, Santibañez V, Soto R, Llama MA (2009) Stable fuzzy self-tuning PID control of robot manipulators. In: Proceedings of the 2009 IEEE international conference on systems, man and cybernetics. IEEE Press, pp 2624–2629
7. Rana KPS, Kumar V, Mittra N, Pramanik N (2016) Implementation of fractional order integrator/differentiator on field programmable gate array. Alexandria Eng J 55:1765–1773. https://doi.org/10.1016/j.aej.2016.03.030.
8. Boulkroune A, Hamel S, Azar AT, Vaidyanathan S (2016) Fuzzy control-based function synchronization of unknown chaotic systems with dead-zone input BT—advances in chaos theory and intelligent control. Presented at the. https://doi.org/10.1007/978-3-319-30340-6_29
9. Hultmann Ayala HV, dos Santos Coelho L (2012) Tuning of PID controller based on a multi-objective genetic algorithm applied to a robotic manipulator. Expert Syst Appl 39:8968–8974. https://doi.org/10.1016/j.eswa.2012.02.027
10. Xiong P, Lai X, Wu M (2016) Position control strategy based on energy attenuation for planar three-link underactuated manipulator. In: 2016 35th Chinese Control Conference (CCC), pp 704–708. https://doi.org/10.1109/ChiCC.2016.7553167
11. Zhang A, Lai X, Wu M, She J (2015) Global stabilization of underactuated spring-coupled three-link horizontal manipulator using position measurements only. Appl Math Model 39:1917–1928. https://doi.org/10.1016/j.apm.2014.10.010
12. Seraji H (1989) Decentralized adaptive control of manipulators: theory, simulation, and experimentation. IEEE Trans Robot Autom 5:183–201. https://doi.org/10.1109/70.88039
13. Kim E (2004) Output feedback tracking control of robot manipulators with model uncertainty via adaptive fuzzy logic. IEEE Trans Fuzzy Syst 12:368–378. https://doi.org/10.1109/TFUZZ.2004.825062
14. Yavin Y, Frangos C (2004) Control of a three-link manipulator: collision avoidance of a moving obstacle. Comput Math Appl 47:1627–1639 https://doi.org/10.1016/j.camwa.2004.06.010.
15. Oh T-S, Kim I-H, Park C-W (2006) Design of nonlinear PI controller for velocity control of induction motor. J Ind Technol 26:227–231
16. Cheng M, Sun Q, Zhou E (2006) New self-tuning fuzzy PI control of a novel doubly salient permanent-magnet motor drive. IEEE Trans Ind Electron 53:814–821. https://doi.org/10.1109/TIE.2006.874269
17. Kong H, Clara S (2011) Advances in PID control (Advances in Industrial Control)
18. Knospe C (2006) PID control. IEEE Control Syst 26:30–31. https://doi.org/10.1109/MCS.2006.1580151
19. Mishra P, Kumar V, Rana KPS (2015) An online tuned novel nonlinear PI controller for stiction compensation in pneumatic control valves. ISA Trans 58:434–445. https://doi.org/10.1016/j.isatra.2015.05.005
20. Su YX, Sun D, Duan BY (2005) Design of an enhanced nonlinear PID controller. Mechatronics 15:1005–1024. https://doi.org/10.1016/j.mechatronics.2005.03.003
21. Moreno-Valenzuela J, Perez-Alcocer R, Guerrero-Medina M, Dzul A (2018) Nonlinear PID-type controller for quadrotor trajectory tracking. IEEE/ASME Trans Mechatronics 23:2436–2447. https://doi.org/10.1109/TMECH.2018.2855161
22. Kasim I, Riyadh W (2016) On the improved nonlinear tracking differentiator based nonlinear PID controller design. Int J Adv Comput Sci Appl 7:234–241 (2016). https://doi.org/10.14569/ijacsa.2016.071032
23. Eiben AE, Michalewicz Z, Schoenauer M, Smith JE (2007) Parameter control in evolutionary algorithms BT—parameter setting in evolutionary algorithms. Presented at the. https://doi.org/10.1007/978-3-540-69432-8_2
24. Goldberg DE, Korb B, Deb K (1989) Messy genetic algorithms: motivation, analysis, and first results. Complex Syst 3:493–530
25. Deb K (1999) An introduction to genetic algorithms. Sadhana 24:293–315. https://doi.org/10.1007/BF02823145

26. Števo S, Sekaj I, Dekan M (2014) Optimization of robotic arm trajectory using genetic algorithm. IFAC Proc 47:1748–1753. https://doi.org/10.3182/20140824-6-ZA-1003.01073
27. Liu GP, Daley S (2000) Optimal-tuning nonlinear PID control of hydraulic systems. Control Eng Pract 8:1045–1053. https://doi.org/10.1016/S0967-0661(00)00042-3

An Intelligent and Effective Cyber-Secured Smart-Home Automation System with Embedded AI

Mohd Kamran Ikram, Karama M. K. Senan, Mohd Mustaqeem, and Mohammad Sarfraz

Abstract With technological advancement, the concept of controlling household equipment remotely via the internet from anywhere in the world at any time is now a reality. This paper presents intelligent smartphone-based home automation and cyber-secured system by using an Arduino microcontroller. The system has two modes of operation, the manual mode, and automatic mode, i.e., the system can operate on its own without any interference of humans or in a dehumanized way. The hardware of the proposed method is successfully developed and its various subsystems are successfully tested. With the help of the ESP866 Wi-Fi module, the whole system can be taken to the cloud (online) mode and the system can be accessed from anywhere in the world. When the appliances start exchanging information with smartphones, and this makes the system vulnerable to security threats and hence threatens the privacy and security of the end-user. The security requirement will be displayed in second place. Our proposed method can provide additional protection in network security, which is easily integrated with heterogeneous IoT devices and protocols.

Keywords Home automation · Cyber security · Vulnerability · Threat · Arduino · Microcontroller · Sensor

1 Introduction

Automation is an integral part of human life and it brings ease to the users by giving them the freedom to control and monitor their home appliances by just using a smartphone. It also gives tons of advantages to the old aged and handicapped

M. K. Ikram (✉) · K. M. K. Senan · M. Sarfraz
Department of Electrical Engineering, Aligarh Muslim University, Aligarh, UP 202002, India
e-mail: mki.kamran@gmail.com

M. Sarfraz
e-mail: msarfraz@zhcet.ac.in

M. Mustaqeem
Department of Computer Science, Aligarh Muslim University, Aligarh, UP 202002, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_13

people. A smart home is the integration of twenty-first century people. It is basically deploying technology into the home to improve quality of life [1]. Moreover, it also provides home security [2]. The improved lifestyle is one of the benefits of home automation. Household automation allows customers to manage and monitor many home appliances through a single system, providing more convenience, improved security, and increased energy. Consumers will regulate their home systems and conserve energy more effectively when home automation technologies are integrated into future smart grids [3]. Also, with the advancement of IoT, The Internet of Things is widely regarded as one of the most crucial areas of future technology and the IoT-connected devices are exponentially increasing [4, 5]. Smart homes have been recognized as one of the industrial areas with the greatest potential for IoT adoption, such as home automation and energy management [6].

Plenty of systems of home automation is documented in the literature. ZigBee-based home automation systems are proposed in [7, 8]. Although the ZigBee technology is inexpensive, the integration of different devices of different manufacturers is the main disadvantage associated with this technology [9].

Several methods based on Bluetooth technology and android are reported [10, 11]. Most of the methods require an Arduino board and a cell phone to form the system's hardware architecture, and the Arduino board and cell phone communicate wirelessly using Bluetooth technology. The major issue with these Bluetooth-based smart-home systems is that they can only operate household appliances within Bluetooth range [12].

The voice recognition-based home automation system is also studied and documented [13, 14]. A smartphone-based application is developed to control home appliances by recognizing the voice commands of the user. The android application communicates with microcontroller boards via Bluetooth or Wi-Fi technology. The voice recognition-based system is user-friendly. The major limitation of this method is that it requires a noise-free environment to operate. Moreover, it also demands the correct pronunciation; otherwise, the system will fail to execute the commands.

The global system of mobile communication (GSM) based home automation systems are also presented [15]. The GSM modem is equipped with a microcontroller to decode the SMS sent from the user's mobile device. It has been established that using GSM in a home automation system provides the highest level of security [16]. Moreover, the GSM technology is vulnerable to power failure [9].

In wireless technology, Wi-Fi-based systems are also studied [17, 18]. The Wi-Fi technology enables the user to control home appliances from anywhere in the world. Wi-Fi technology makes home automation systems easier to adopt than ZigBee technology since Wi-Fi has lower connection latency than ZigBee [9].

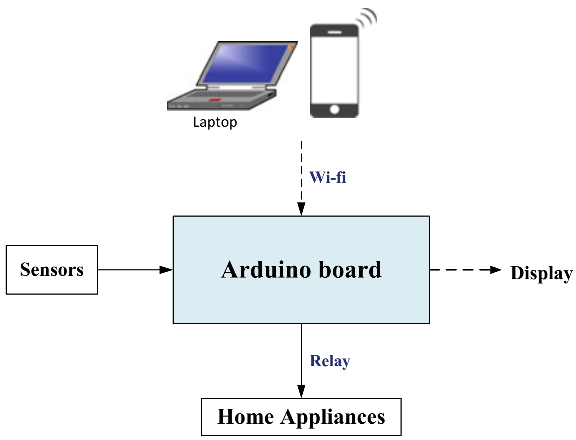
It is known that IoT-based smart-home automation systems also suffer from the problem of security threats which is a very big challenge to protect the system from cyber-attacks and risk. According to [19], smart-home automation with the learning of human behavior can be secure by reducing the computation overheads, but they have used traditional and conventional methods. In the study [20] author has used the network security approach to protect the system from cyber-attack. They have used the central router concept, but if the attack on the main router becomes successful, all the connected devices will be easily manipulated. In the manuscript [21] author has told about the reference architecture of the smart home that facilitates the security analysis, along with the security attacked surfaces by providing the three viewpoints. Still, they have used old approaches to demonstrate the scenario. In comparison to others, our proposed system will contain multi-level security points that can provide the next level of security to the IoT devices. We will give multi-layered network security so that if the cracker can get the primary level, it cannot break the other layer easily.

In this work, an Arduino microcontroller-based dual-mode home automation system is being designed. The system is controlled via the smartphone/laptop through a Wi-Fi module. In addition to this, the user has control over the system to use it either in manual mode or in automatic mode. Four subsystems are designed, temperature control monitor, lighting control, door security, and fire-fighting systems. This research work is based on machine learning algorithms with the internet of things. IoT devices constitute a large number of devices and sensors that can monitor/control different physical quantities. With the dawn of Big Data, there is a need for proper automated data storage solutions and cloud-based applications to analyze and extract the required information from the data.

2 System Overview

The system is designed so that four subsystems are coupled with an Arduino microcontroller via sensors and the home appliances are connected to the Arduino microcontroller via a relay system for their respective operation. The system is controlled via smartphone/laptop with the help of a Wi-Fi module. The basic system design is shown in Fig. 1.

Fig. 1 System architecture



The proposed methodology has two modes; one is manual mode, in which the user has a choice to operate appliances on his own. The other is an automatic mode in which the home appliances will automatically operate without any interference of humans or can be called a dehumanized mode.

The ESP8266 takes the whole system to the Wi-Fi system mode, where all the developed algorithms and methods are being utilized. The information exchanged among the heterogeneous devices will be taking place in the network-secured manner to fulfill the CIA triad of security.

3 Subsystems

The components of the system are connected with a microcontroller. A 16 × 2 Liquid Crystal Display (LCD) is used to display the data like room temperature, object distance, and the information about the number of devices connected, etc. The hardware of the system is successfully implemented and its various subsystems are verified. An android application Blynk is used to monitor the connected home appliances.

3.1 Temperature Control Monitoring System

In a temperature control system, an LM35 temperature sensor is used along with the infrared (IR) sensor, the LM35 is used for the purpose of measuring the ambient temperature of the room. The IR sensor is used to improve the accuracy. The temperature is adjusted to be less than 30 °C; if the temperature exceeds 30 °C, the microcontroller will turn on the fan or Air Conditioner to reduce the temperature.

3.2 Lighting Control System

A LED will turn on according to the place's luminosity. It is performed through a light-dependent resistor (LDR). As the light impinging on the photoresistor gets stronger, the resistance decreases, and the voltage output of the divider increases. The reverse happens when the impinging light gets weaker.

3.3 Door Automatic Opening System

The HC-SR04 is a complete ultrasonic distance measurement device. The HC-SR04 works simply by sending a pulse to the trigger input and waiting for a pulse from the echo output. The length of this pulse allows the calculation of the distance to the detected object.

3.4 Fire-Fighting System

In the fire-fighting system, an MQ-2 gas sensor is to operate when the amount of smoke or gas exceeds 400%, when this condition occurs, the Arduino will operate the buzzer first and then the fire-fighting system will operate.

3.5 System of Other Appliances

The other appliances like refrigerators, microwave ovens, and geysers can be connected with the home automation system in such a way, they operate in accordance with the user.

4 Cyber-Attack on Home Automation System

The problem with the Wi-Fi-connected home automation devices is that it is remotely controlled when we are away from home. To communicate with them, they need to be connected to the internet (cloud database) from outside the home network. They are controlled by smartphones or laptops that detect the devices and are connected. This availability of the device on the internet through Wi-Fi makes it vulnerable to hackers and attackers to manipulate. It can be done in the following manners, as shown in Fig. 2.

4.1 Man-In-The Middle

In this attack, the attacker can breach, interrupts or spoofs communications between two systems. The attacker can send the fake temperature data generated by the device to be spoofed and sent to the cloud; similarly, the attacker can disable vulnerable devices during heat waves, creating a disastrous situation for the system users.

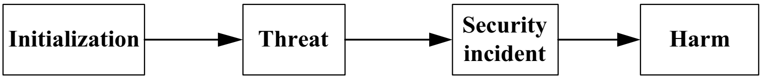


Fig. 2 Process leading to security happenings

4.2 Data and Identity Stealing

When unprotected wearable and smart devices generate data, they can provide important personal information to the hackers that can be exploited for fraud, transactions, and identity hacks.

4.3 Device Hijacking

In this type of attack, the attacker can hijack the device and control the device effectively. These attacks are quite difficult to detect because the attacker does not change the device's basic functionality. It simply controls a single major device and can get access to others as well.

4.4 Distributed Denial-of-Service

It is the advanced version of the Denial-of-service (DoS) attack. The attacker can flood the incoming traffic coming from multiple sources to a targeted network. It can block the simple source to get access to the data. This attack is now common in IoT devices nowadays due to the lack of security of the devices. Another version called Permanent Denial-of-service (PDoS) attack is a lethal attack. It can damage the device badly the device cannot recover, it can replace, or reinstallation takes place.

5 Protection from Cyber-Attacks of Home Automation System

The IoT devices can be protected by comprehensive IoT security solutions that do not disrupt home or service providers' smart devices. It can include the following things to get a more secure system.

5.1 Multi-level Authentication

As the smart device is ready to connect with the network, it should be authenticated before information exchange with the cloud. This can ensure the data generated from the device is not fraud and corrupt. Hashing algorithms SHA-256 can be used

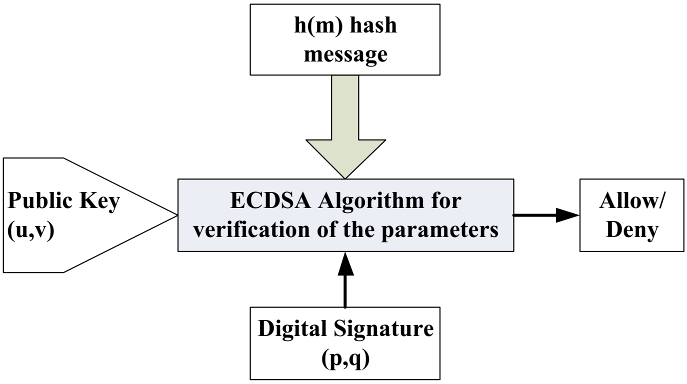


Fig. 3 ECDSA algorithm

for symmetric keys and Elliptic Curve Digital Signature Algorithm (ECDSA) for asymmetric keys cryptography (Fig. 3).

5.2 Network Data Communication Security

Data transmission between the device and cloud can be done in encryption key authentication mode. It can be protected by ensuring or allowing those with the secret decryption key to communicate with the cloud data. The information sent to the cloud by the smart device must be protected from digital eavesdropping (Fig. 4).

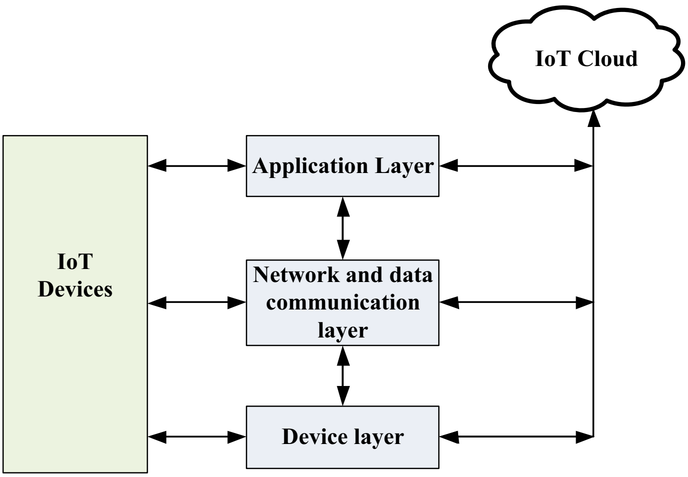


Fig. 4 Security layered model of IoT

5.3 Analysis of Security Monitoring

The cloud system or smartphone will capture the data from the various smart devices, including endpoint devices and network traffic data. Our system can analyze whether any threat or vulnerability occurs or not if necessary actions can be formulated for the system security policy like isolation of vulnerable devices. This analyzing cycle can be run in real-time or later to detect the attack pattern and scenarios.

5.4 Machine Learning Approach

Detecting the anomalies and vulnerability of attacks in the smart-home systems through a traditional approach like traffic pattern recognition, it is better to use dynamic machine learning methods to classify the data flows. It is because the unknown attack detection possibility can also be enhanced by using this approach.

Our proposed model can be secured in the techniques mentioned above. The overall demonstration of the cyber security portion of the proposed model can be seen in Fig. 5. The device present at the device layer and the most important layer, i.e.,

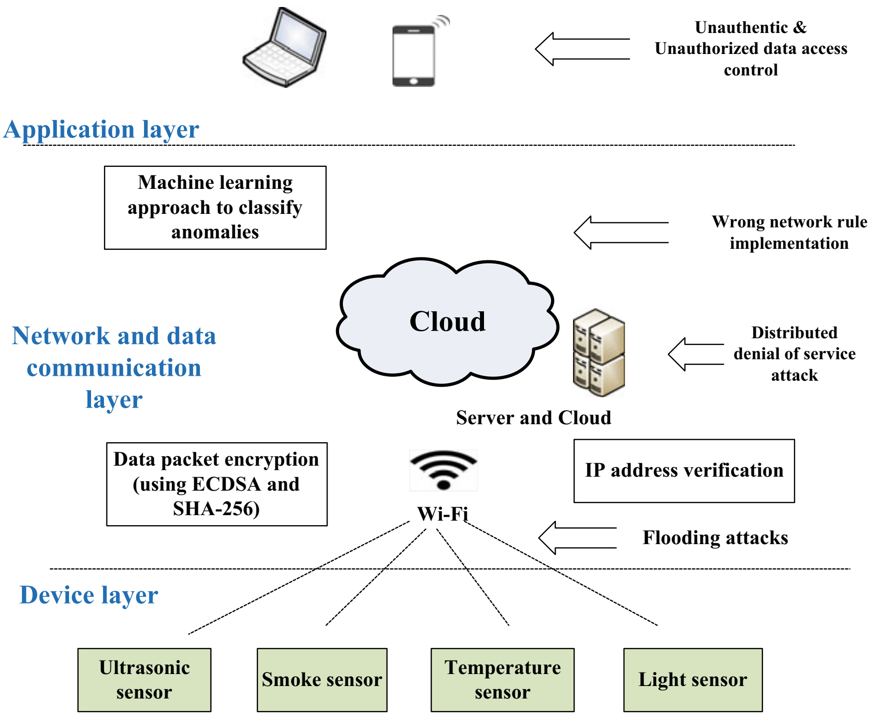


Fig. 5 Overall cyber security demonstration of home automation devices

network, has the vulnerability from multi-level attacks. We have to apply security in data transmission on the second layer by using ECDSA and SHA-256 algorithms for encrypted data transfer and IP-address verification through verification code generation technique along with the anomaly detection using machine learning approach in which we can apply Support Vector Machine (SVM) technique to classify the anomalies in the unknown data. It can classify defective and non-defective datasets, which can save our time to manual and traditional classification and more accuracy so that the information exchange can occur securely.

6 Experimental Evaluation

The Arduino is powered via the USB cable through a laptop or PC. The relays are connected to the output ports of the Arduino. The appliances are connected to the relays; the relays are behaving just like a switch. The system has three main components, the Arduino board, sensors, and a Wi-Fi module. Figure 6 shows an experimental evaluation of the Hardware model of the proposed home automation system. The hardware model is successfully tested on various smartphones and laptops.

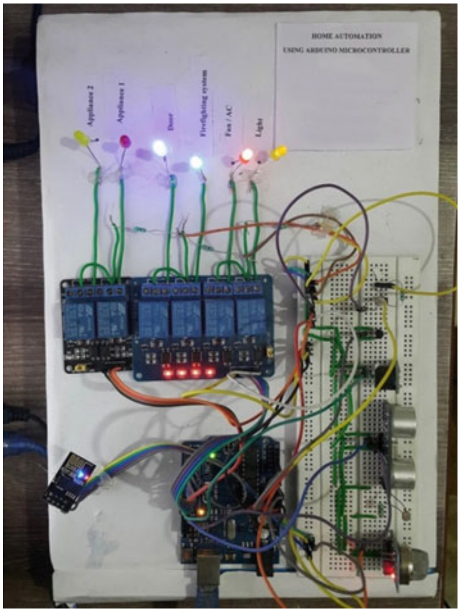


Fig. 6 Experimental setup of proposed home automation system

7 Conclusion and Future Research Challenges

This paper proposes a fundamental design of a cyber-secured and smart-home automation system based on Arduino microcontroller and the hardware model is successfully implemented. The key features of this design are low cost, effective, scalable, fast responsive, straightforward implementation, and online compatibility. Moreover, the user can use the system with the help of a smartphone by just downloading an application named Blynk. The system is capable of performing tasks on its own. In security aspects, the proposed system is smart and cyber-secured with multi-level security that can produce a tangible innovation roadmap with real impact, and the system can maintain the security triad, i.e., Confidentiality, Integrity, and Availability (CIA). If anyone wants to get unauthorized access or manipulate the exchanged information, it cannot be done easily; it has to crack the multi-layered security of the system.

In the future, there will be the use of some AI-based techniques to automate the system's security that can detect and correct if any vulnerabilities occur in the system. Because the Arduino MEGA has more pins than the Arduino Uno, the same work may be expanded. Controlling only a few items can be addressed by expanding automation to all other home appliances. Security cameras may be operated, allowing the user to watch what is going on around their home or company. Motion sensors may be used in security systems to detect any illegal movement and alert the user. Through Wi-Fi and sensor signals, the scope of this work may be expanded to numerous locations, rather than being limited to only the home.

References

1. Bromley K, Perry M, Webb G (2003) Trends in smart home systems. Connectivity and services, www.nextwave.org.uk.
2. Das SR, Chita S, Peterson N, Shirazi BA, Bhadkamkar M (2011) Home automation and security for mobile devices. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). IEEE, Seattle, WA, USA, pp 141–146. https://doi.org/10.1109/PERCOMW.2011.5766856
3. Ahmim A, Le T, Ososanya E, Haghani S (2016) Design and implementation of a home automation system for smart grid applications. In: 2016 IEEE International Conference on Consumer Electronics (ICCE). IEEE, Las Vegas, NV, USA, pp 538–539. https://doi.org/10.1109/ICCE.2016.7430721
4. Ding J, Nemati M, Ranaweera C, Choi J (2020) IoT connectivity technologies and applications: a survey. IEEE Access. 8:67646–67673. https://doi.org/10.1109/ACCESS.2020.2985932
5. Lee I, Lee K (2015) The internet of things (IoT): applications, investments, and challenges for enterprises. Bus Horiz 58:431–440. https://doi.org/10.1016/j.bushor.2015.03.008
6. Santoso FK, Vun NCH (2015) Securing IoT for smart home system. In: 2015 International Symposium on Consumer Electronics (ISCE). IEEE, Madrid, Spain, pp 1–2. https://doi.org/10.1109/ISCE.2015.7177843.
7. Gill K, Yang S-H, Yao F, Lu X (2009) A zigbee-based home automation system. IEEE Trans. Consumer Electron 55:422–430. https://doi.org/10.1109/TCE.2009.5174403

8. Byun J, Jeon B, Noh J, Kim Y, Park S (2012) An intelligent self-adjusting sensor for smart home services based on ZigBee communications. IEEE Trans. Consumer Electron 58:794–802. https://doi.org/10.1109/TCE.2012.6311320
9. Danbatta SJ, Varol A (2019) Comparison of Zigbee, Z-Wave, Wi-Fi, and bluetooth wireless technologies used in home automation. In: 2019 7th International Symposium on Digital Forensics and Security (ISDFS). IEEE, Barcelos, Portugal, pp 1–5. https://doi.org/10.1109/ISDFS.2019.8757472
10. Asadullah M, Ullah K (2017) Smart home automation system using Bluetooth technology. In: 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT). IEEE, Karachi, Pakistan, pp 1–6 https://doi.org/10.1109/ICIEECT.2017.7916544
11. Husain MdI, Alam M, Rashed MdG, Haque MdE, Rashidul Hasan MAFM, Das D (2019) Bluetooth network based remote controlled home automation system. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE, Dhaka, Bangladesh, pp 1–6. https://doi.org/10.1109/ICASERT.2019.8934500
12. Al-Ali AR, AL-Rousan M (2004) Java-based home automation system. IEEE Trans Consumer Electron 50:498–504. https://doi.org/10.1109/TCE.2004.1309414
13. Eric T, Ivanovic S, Milivojsa S, Matic M, Smiljkovic N (2017) Voice control for smart home automation: evaluation of approaches and possible architectures. In: 2017 IEEE 7th International Conference on Consumer Electronics-Berlin (ICCE-Berlin). IEEE, Berlin, pp 140–142. https://doi.org/10.1109/ICCE-Berlin.2017.8210613
14. Mittal Y, Toshniwal P, Sharma S, Singhal D, Gupta R, Mittal VK (2015) A voice-controlled multi-functional Smart Home Automation System. In: 2015 Annual IEEE India Conference (INDICON). IEEE, New Delhi, India, pp 1–6. https://doi.org/10.1109/INDICON.2015.7443538
15. Yuksekkaya B, Kayalar AA, Tosun MB, Ozcan MK, Alkar AZ (2006) A GSM, internet and speech controlled wireless interactive home automation system. IEEE Trans Consumer Electron 52:837–843. https://doi.org/10.1109/TCE.2006.1706478
16. Alheraish A (2004) Design and implementation of home automation system. IEEE Trans Consumer Electron 50:1087–1092. https://doi.org/10.1109/TCE.2004.1362503
17. Jakovljev S, Subotic M, Papp I (2017) Realisation of a smart plug device based on Wi-Fi technology for use in home automation systems. In: 2017 IEEE International Conference on Consumer Electronics (ICCE), IEEE, Las Vegas, NV, USA, pp 327–328. https://doi.org/10.1109/ICCE.2017.7889340
18. A JC, Nagarajan R, Satheeshkumar K, Ajithkumar N, Gopinath PA, Ranjithkumar S (2017) Intelligent smart home automation and security system using arduino and Wi-Fi. IJECS. Department of Electrical and Electronics Engineering, Gnanamani College of Technology, Namakkal India. https://doi.org/10.18535/ijecs/v6i3.53
19. Chaurasia T, Jain PK (2019) Enhanced smart home automation system based on internet of things. In: 2019 Third international conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp 709–713. https://doi.org/10.1109/I-SMAC47947.2019.9032685
20. Saha R, Bera JN, Sarkar G (2018) An alternate approach for power quality computation using sample shifting technique towards load characterisation. Measurement 129:642–652. https://doi.org/10.1016/j.measurement.2018.07.037
21. Ghirardello K, Maple C, Ng D, Kearney P (2018) Cyber security of smart homes: development of a reference architecture for attack surface analysis. In: Living in the internet of things: cybersecurity of the IoT—2018. Institution of Engineering and Technology, London, UK, p 45 (10 pp)–45 (10pp). https://doi.org/10.1049/cp.2018.0045

IoT and GSM Based Real-Time Water Monitoring System with Low Power Consumption

Divesh Kumar

Abstract Increasing water pollution is a massive issue in today's world. It not only affects humans but other species on earth as well. Many researchers are working in this field on how the quality of water can be improved and for that, it needs to be monitored carefully. The proposed system presents a real-time water quality monitoring system that uses IoT and GSM technology for transferring data to the authorities and is also powered by solar energy which makes it different from others and more efficient for usage. Apart from that, the microcontroller used in the proposed system is capable to operate in ultra-low power mode, so it enables the compactness and durability of the system. The sensors connected to the microcontroller will provide different parameter's values and allow the authorities to have incessant checks on the water quality.

Keywords ARM cortex m4F · Cloud computing · Internet of things (IoT) · GSM technology · Solar energy · Water quality monitoring

1 Introduction

About 71% of the earth's surface is water-covered and so is the planet endowed with the name "Blue Planet". Out of the total percentage, 96.5% of water is salty and acquired by the oceans, the remaining percentage of water is fit for drinking purposes but only 1.2% of it is available to us rest is locked in glaciers, ice caps, and permafrost.

Water is one of the basic necessities of survival for all the living beings on the earth. But due to the increasing pollution, everyone is supposed to suffer. Due to the consumption of unhygienic water, not only humans but the other species also develop many diseases. In Fact, a lot many species have already got extinct in recent years. The main reason for this surging pollution is the vast development of industries in the country. All the wastes and by-products of these factories are directly dumped into

D. Kumar (✉)
Department of ECE, GLA University, Mathura, U.P., India
e-mail: divesh.kumar@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_14

the rivers without any pre-treatment, since these wastes contain both chemical and solid impurities, they create a major problem for aquatic animals. Aquatic animals swallowed up the solid waste and the chemicals which had emerged out from the firms disturb the chemical composition of water and decrease the level of oxygen dissolved in the water which is the basic requirement for these marine creatures to survive.

As a result, our rivers are getting deprived of aquatic animals and also losing their essence day by day. The same water from the rivers is used for agriculture and thus makes vegetation unhealthy. Since river is the source of fresh drinking water all living beings consume it and get diseased. Apart from the factories, rivers are getting polluted by several human activities as well. The government has initiated many projects for the cleaning of water bodies but neither gets success up till now. For the treatment of the water, there is a conventional method of collecting samples and then checking the quality of water on the basis of various parameters. This method is appropriate but the only drawback with this is that a long time period is required for this along with the requirement of other resources as well which simply goes to vain.

For resolving this problem many systems have been proposed with various features.

Dandekar et al. proposed a water grade monitoring system as good water quality is a necessity in everyday life. In this paper, various sensors such as ph sensor, turbidity sensor are used to test different parameters of water which are compared with standard values that are stored in database. If the measured values do not match with standard values then the concerned water regulating authorities are informed using GPRS and web-based monitoring system is used for observing water quality. The main benefit of this system is that it is operated using solar energy therefore it helps in saving electricity [1].

Geetha and Gouthami proposed a survey of existing water quality monitoring systems. This paper provides a simple solution for in-pipe water quality monitoring using IoT. Data from the controller are uploaded over internet and analyzed. It also gives alert to user if there is any deviation from standard drinking water values [2]. Soumya and Preeti Gadgay proposed a system that solves the problem of water pollution by using a cleaning robot in rivers that uses belt drive mechanism for lifting debris from water [3]. Raghunathan and Chou proposed a research on various issues, tradeoffs, system design issues, and power management that are involved in designing energy harvesting embedded systems. It also provides various system design techniques that are used for extracting most energy from environment and making it available for consumption as there is more demand for low power system designs [4].

Brunelli et al. proposed a research on solar harvesting circuits that are used for increasing efficiency of embedded systems and also provides ways for optimizing solar energy collection when there are no stationary light conditions [5]. Brunelli et al. proposed a method to overcome the problem of limited battery lifetime that frequently requires battery replacement or recharging by developing a solar harvesting circuit for battery-less embedded system [6]. Vijayakumar and Ramya proposed a low-cost system for real-time monitoring of the water quality. The system measures various physical and chemical parameters of water using different types of sensors which then transfer the recorded data to the authorities via microcontroller. Since the system is equipped with a Wi-Fi module so with the help of IoT the data can be accessed anywhere in the world [7].

Wang and Liu proposed a river water level remote monitoring system using a GPRS module. The system is designed with various features so that there is no manual requirement of inspection rather the data, i.e., the values of water level can be accessed from a remote area at any moment of time and according to that surrounding residents can be notified. With the help of a water level sensor data is recorded and transmitted to the server terminal with the help of GPRS technology [8]. Yuchun et al. proposed a research on GSM technology. GSM (global system for mobile communications) is an open and digital cellular technology used for transmitting mobile voice and data services that operates at 850, 900, 1800, and 1900 MHz frequency bands. With the help of GSM technology, we can send and receive SMS (short message service) and even emails. GSM is the most widely accepted standard in telecommunications globally [9].

Rahim et al. proposed a research on IoT and cloud convergence. The internet of things (IoT) refers to a system of interrelated, internet-connected objects that are able to collect and transfer data over a wireless network without human intervention. Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user. The cloud computing and IoT have become two very closely affiliated future internet technologies with one providing the other a platform for success [10].

This paper proposes the model "Real-time water quality monitoring using IoT and GSM Technology". The model provides a system where the water quality is monitored without any manual support or in other words, an unmanned system where all the parameters are measured by the system itself and then transferred to the required authorities and also to the locals in case of any emergency. The model will run on solar power and become more effective in terms of power requirement as the need for external power source is demolished. The different sensors connected to the microcontroller TM4C123GH6PM, pick up the values of different parameters and send the data to the launchpad for further process. The model uses IoT and GSM modules for transferring the data, to the authorities and locals via cloud computing and GSM mobile telephone technology, respectively.

2 Model Descriptions

Clean water is one of our most exquisite assets, but nowadays due to the pollution, finding clean water is very rare. Rivers, the main source of fresh water are polluted above the mark which is simply dragging our society toward a waterless life that is impossible to live. To survive on the planet we need fresh water and for that, the rivers should be cleaned but that is not so easy. The customary process of recognizing

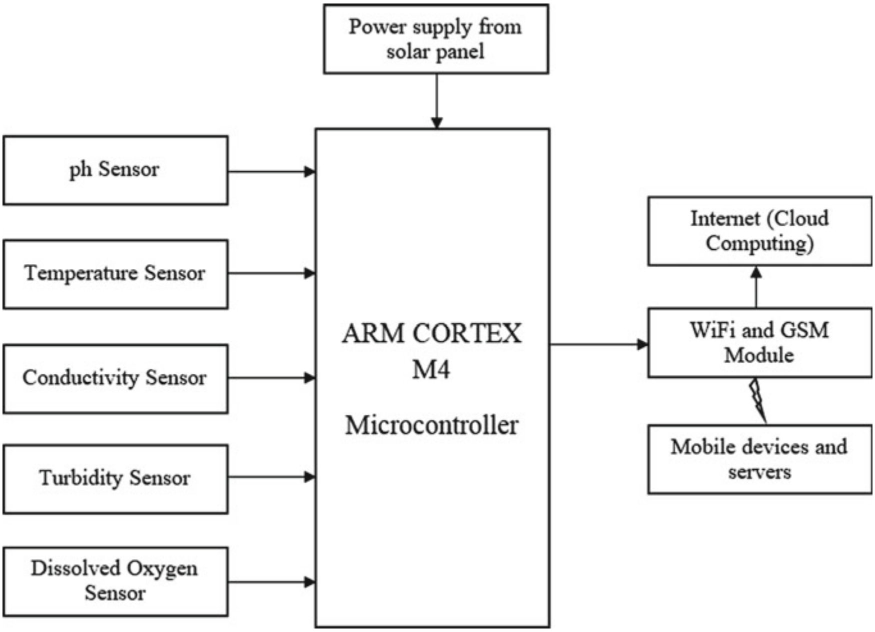


Fig. 1 Block diagram of real time water grading system

the sources of pollution by taking samples and then performing various tests is quite expensive and time-consuming and in some cases like an emergency, it is futile.

The proposed model comes up with a solution for the above-mentioned issues. As shown in Figs. 1 and 2 the model comprises a system where various sensors, an ESP32 SIM800L wifi module, and a charging circuit are connected with the microcontroller (tm4c123gh6pm). Data input from the sensors to the microcontroller is transferred to the cloud so that the authorities can have a continuous check on the quality of water.

2.1 Block Diagram

Figure 1 gives a brief view of the proposed model. In the model, five different sensors are used for measuring the water quality. Temperature sensor, ph sensor, turbidity sensor, dissolved oxygen sensor, and conductivity sensor are connected to the microcontroller. A charging circuit, which uses solar energy, is connected to the microcontroller for providing the input power to the microcontroller. An IoT cum GSM module is also connected to the launchpad for the transmission of data to the cloud and also for the communications with locals and authorities.

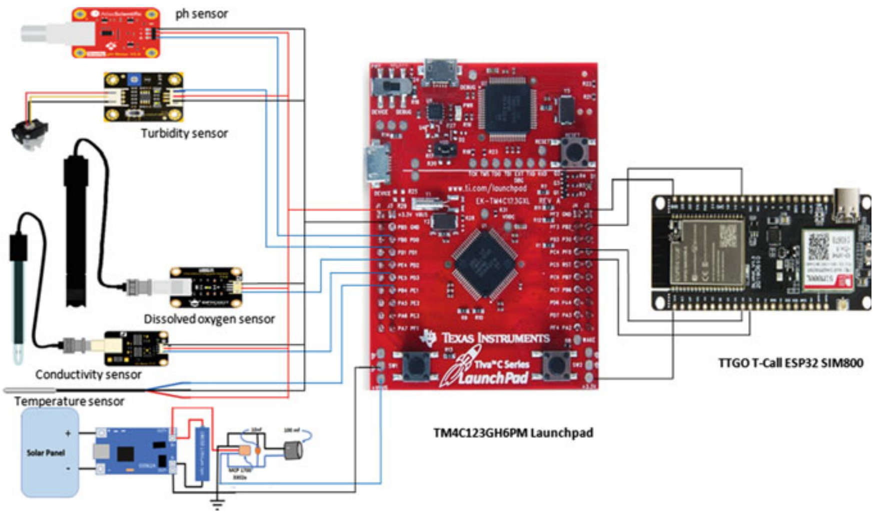


Fig. 2 Schematic diagram of real time water grading system

2.2 Schematic Diagram

Figure 2 gives a more proximate view of the model of how the components are connected together. All the sensors ph sensor (to check the ph level of the water), turbidity sensor (to check the turbidity amount in water), conductivity sensor (to check the conductivity in water), dissolved oxygen sensor (to check the dissolved oxygen amount I water), and temperature sensor (to check the temperature of water), are connected to the analog pins of the microcontroller, i.e., to the ADCs of the microcontroller which will convert the sensor data to digital data and then transfer it for further analysis on the cloud. The wifi module used in the model is TTGO T-Call ESP32 SIM800L, which is an IoT cum GSM module. Through this wifi, cum gsm module data is uploaded on the ThingSpeak cloud for the monitoring of the water. For input power to the microcontroller a charging circuit is used which consists of a solar panel, lithium battery cells, charging board, voltage regulator, and capacitors.

2.3 Flow Chart

Figure 3 gives a view of workflow of the whole model. First of all the system is initialized. After that, the microcontroller will collect data from the sensors and convert them into their digital value. Then after the conversion, the data is sent to the cloud for further analysis. If any parameter's value, sensed by the sensors, is greater than or equal to the threshold value, i.e., if more than or equal to the standard value

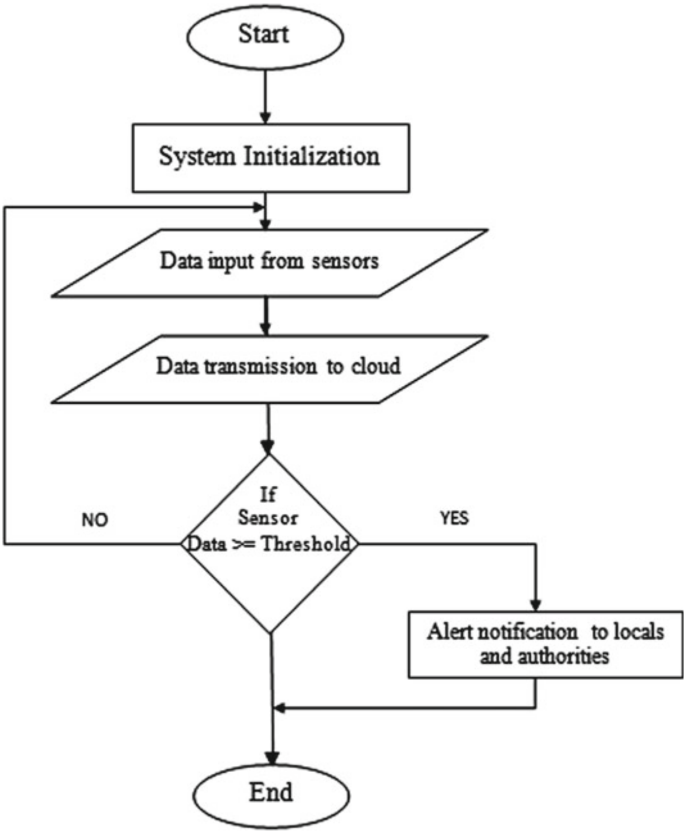


Fig. 3 Flow diagram of real-time water grading system

then an immediate alert notification will be sent to the locals and authorities so that appropriate actions can be taken in due time.

3 Results

The proposed system is tested and the data from the sensors is stored on the cloud so that it can be accessed from anywhere and at any time. Four different sensors were tested. Figure 4 shows the reading of these sensors. Figure 4a shows the current values of pH of the water. Figure 4b shows the temperature readings. Figure 4c shows the conductivity variations of the water and Fig. 4d shows the TDS values. Data received from the sensors are plotted for better comparative analysis. This data can be stored for analysis of the water quality over a long span of time.

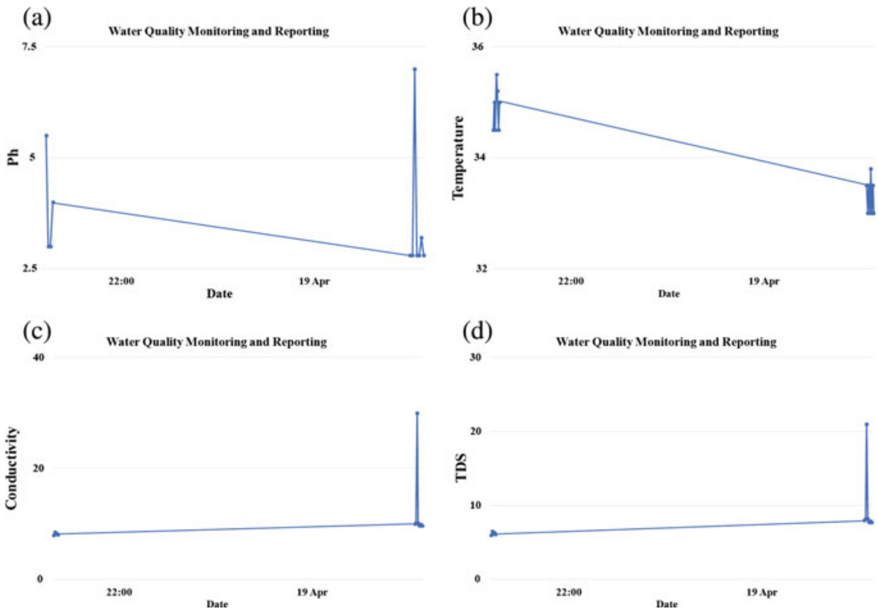


Fig. 4 **a** pH sensor data. **b** Temperature Sensor data. **c** Conductivity sensor data and **d** TDS sensor data

4 Conclusion and Future Scope

This paper proposed the real-time water quality monitoring system. The proposed system was successfully tested and obtained good results overcloud. The proposed system is solar powered so it can be used at any location irrespective of the accessibility and as the microcontroller used in the proposed system can run in ultra-power saving mode so overall power consumption is also very less compared to the other systems. The only limitation is the periodic maintenance/cleaning of the sensors connected to the system. For longer duration of the operation and better accuracy, optical sensors can be used in place of traditional sensors.

References

1. Dandekar S et al (2018) IoT based real time water grade tracking system using solar energy. In: 2018 3rd International Conference on Communication and Electronics Systems (ICCES). IEEE
2. Geetha S, Gouthami S (2016) Internet of things enabled real time water quality monitoring system. Smart Water 2(1):1
3. Soumya HM, Gadgay B (2018) Pond Cleaning Robot. Int Res J Eng Tech 5(10):2395–0056
4. Raghunathan V, Chou PH (2006) Design and power management of energy harvesting embedded systems. In: Proceedings of the 2006 international symposium on low power electronics and design
5. Brunelli D et al (2008) An efficient solar energy harvester for wireless sensor nodes. 2008 Design, Automation and Test in Europe. IEEE
6. Brunelli D et al (2009) Design of a solar-harvesting circuit for batteryless embedded systems. IEEE Trans Circ Syst I: Regular Papers 56(11):2519–2528
7. Vijayakumar N, Ramya R (2015) The real time monitoring of water quality in IoT environment. In: 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE
8. Wang Y, Liu Z (2017) GPRS based river water level monitoring and measuring system. 2017 29th Chinese Control and Decision Conference (CCDC). IEEE
9. Yuchun M et al (2011) General application research on GSM module. In: 2011 International Conference on Internet Computing and Information Services. IEEE
10. Biswas AR, Giaffreda R (2014) IoT and cloud convergence: opportunities and challenges. In: 2014 IEEE World Forum on Internet of Things (WF-IoT). IEEE

Application of Internet of Things and Cloud Computing to Enhance the Agro-productivity

Anju Upadhyay and Indresh Yadav

Abstract The growing global population needs more food production in all areas, particularly agriculture. Still, demand and supply will not always match up. In order to increase agricultural productivity, it is difficult to maintain and sustain capital resources and labor. With this research, the function of the Internet of Things (IoT) in agriculture is examined. As a result of GPS and sensors, modern agriculture is able to exchange, break down, and trade information. IT benefits agriculture in the same way that cloud computing does. Agriculture cloud and IT service provide outstanding ability management to ranchers in terms of yield development, estimating, composts, and maladies specific remedy approach to be used. Scientists working in agriculture will present their findings, as well as recommendations for cutting-edge development techniques and manure usage, so that the history of the region may be learned. Agricultural cloud computing applications were used as the basis of the evaluation. As agro-cloud enhances agricultural production and access to research-related information develops in the future, expenses and time are saved, and correspondence is facilitated. In the topic of IoT in agriculture, this study would help to forward a great deal of research.

Keywords Smart agriculture · Internet of things · Cloud computing · Agronomies

1 Introduction

India is a country whose economy is based on agriculture. Agriculture will benefit from IoT-based sensor systems in the future. Devices are used in the power grid, railroad, link, and burrow streets, buildings, dams, oil and gas pipelines, and other sectors. Then find out how to operate it remotely by connecting to the internet, which

A. Upadhyay (✉) · I. Yadav
Department of Electrical Engineering, GLA University, Mathura, U.P, India
e-mail: anju.upadhyay@gla.ac.in

I. Yadav
e-mail: indresh.yadav@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_15

runs services. RFID approaches, web working communication, new sensor innovation, and sensor organization innovation are all examples of IoT methods [1]. IoT is a brilliant recognition, detection, and intelligence invention. Cloud computing has five advantages, including lower startup costs, unrestricted asset allocation, backend maintenance, and upgrades, and the absence of collaboration with other cloud frameworks, as well as the possibility for global administration improvement. To make data innovation-related operations more easily accessible, cloud computing hides the complexities of the services. Customers do not necessarily need to know who is delivering services. Services do not have to be disclosed to the customer [2]. It not only provides maintenance but also lowers the cost of employing such services. It provides services to clients as and when they are required. This is a one-of-a-kind cloud that was created using MAD-cloud design [3]. A country like India, which is primarily reliant on agriculture, has a rural sector that accounts for 20% of the country's GDP and feeds around 66% of the people. The main objective of our study is to unambiguously describe arrival potential, as well as to advance for special and always exhibiting indicators of change soil and climatic conditions. Our study's major goal is to define arrival potential in a clear and unambiguous manner, as well as to identify distinctively and always altering indicators of climatic and soil conditions.

With cloud-based administrations, ranchers may communicate with the cloud through less expensive ways such as sensors, mobile phones, scanners, etc. An online connection can be used to pose a query to the customer. Data is saved in line with the harvest's coordinate, physical, and chemical requirements [4]. The data is organized in a methodical manner, and it is updated by the administrator. Sensors and GPS are used to gather data. In addition to soil surface parameters, rainfall totals and wind speed are included in the data. Information about comparable goods may be saved by the consumer, which will be useful in the future when the generation grows. In addition to choosing a location, the customer can add personalization such as a name, address, and so on. On the other hand, it illustrates an insignificant disease and how to treat it. Agriculturists benefit from the use of the cloud in agriculture, and it can also be used to enhance our financial status by offering target approaches such as essential quality, reliability, and security [5]. As part of this study, which takes into consideration farmers' prior experiences [6], predictive assessment, IoT sensors with cloud administration, and security units for multi-culture farming are all included. It also highlights the challenges and complications that may arise when integrating modern technologies into traditional farming techniques.

2 Scope of Internet of Things (IoT) in the Agricultural Sector

Agricultural production provides the bulk of India's people with income and jobs. A few nations practice exact farming, however, cloud computing must be included in farming in order to boost edit production and, more crucially to expand our economy.

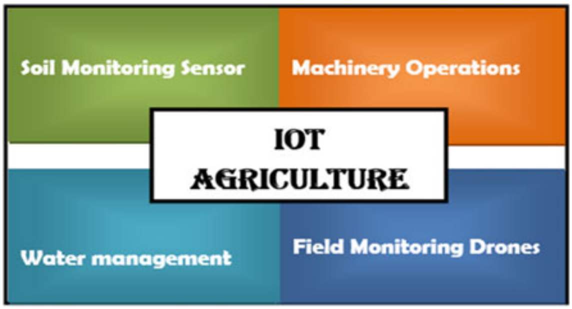
In many ways, the Internet of Things (IoT) represents a foreshadowing of what is to come in computing and communication [7]. An autonomous remote-control system and an interrogation system are hinted at. In both rich and emerging nations, the return of global subsidence has sparked a boom in economic activity and prices. The agriculture sector will have a far better chance of ensuring global food security. Other IoT applications may be found in agriculture, human services, and retail as well as environmental management as well as production network monitoring [5]. Agricultural uses include land and crop surveillance, greenhouse surveillance, and natural pecking order supply control framework monitoring. For IoT to be accepted and widely utilized, the key administration of things must be sophisticated and beneficial to target consumers.

Advantages of applying IoT in Agriculture incorporate.

- Sustainability
- Lower generation costs
- Food wellbeing
- Increased benefits
- Environment protection

In the last ten years, the rural framework has produced incredible results in the sector of agriculture. It offers the agricultural facts benefit the basement. In terms of sophisticated structures, IoT offers a variety of options, including batch cultures, cultural identity configuration management, artificial photosynthesis, maturing energy management, smart water system control, and more. A general application of IoT in agriculture is shown in Fig. 1. Illustration of IoT components used by farmers in order to improve plant production lines, ranches are utilizing IoT innovation. For agriculture, this allows for year-round growth of crops in a very precise and manageable environment [8].

Fig. 1 IoT in agriculture



3 Cloud Computing Applications in Agriculture Information Systems

In 2025, the United Nations Food and Agriculture Organization predicts that the population of the United States would rise to 8 billion people, and to 9.6 billion by the year 2050. It is not uncommon for cloud computing to be used as a way to connect with demonstrations throughout the globe. There is also the usage of remote sensors to keep tabs on what is going on. In order to meet the rising demands of a growing population, this may be used to enhance harvest production [3]. Together with the Internet of Things technology, cloud computing enables pay-per-use pricing as well as a decrease of the cast. This may be created to increase the quality, quantity, manageability, and economic viability of agricultural production. Due to the difficulty in managing agricultural data, cloud computing has become a necessity in this area. After years of advancement in the invention, the agricultural industry has gained prominence in recent years [4]. Ranchers can get the information they need from the agricultural data asset and administration, such as changes in the environment and soil dependency. It is possible that cloud computing will be implemented in the agriculture sector; it will supply data and commodities that are suited for that scenario. During the cloud computing era, different firms can utilize the basic space to store messages and to use the framework offered by cloud providers. Agricultural data assets are kept on a cloud server, eliminating the need to acquire equipment. In addition to expert data, it allows for more exact invoicing. It is characterized by five characteristics: reduced initial fetched, unconstrained asset distribution, support and redesigns are done in the backend and they do not work with other cloud frameworks, and probable outcomes of worldwide administration advancement to make IT-related administrations easily available, cloud computing hides the administrations' problems.

In Fig. 2, you can see how cloud computing is used. Customers do not necessarily need to know who is delivering services. Because it decreases the cost of profiting those administrations and offers upkeep for them, it is beneficial to them. Our main objective is to expressly and strongly describe arrive potential for one-of-a-kind and constantly changing soil and climate circumstances [6]. Agriculturists can use cloud applications to extend their agricultural fields. This approach will also promote faster, more comprehensive merging and dissemination of adjacent and logical data. This program provides excellent administration and useful information to clients at any time and in any location. It provides a larger scope zone. The scope of this article is to deal with the representation and evaluation of agricultural data in the face of changing climatic and soil conditions. Cloud enables agriculturists to connect to the cloud using less expensive methods such as sensors, portable devices, scanners, and so on. A web connection can be used to ask the client a question [9]. The exploration program is entirely based on the MAD-cloud architecture, and the data is saved according to coordinates, as well as physical and material harvesting requirements [10]. The data is saved in a methodical framework, which is then reviewed by an administrator, and the data is acquired using sensors and GPS. In addition, the data characterizes soil surface, moisture, wind speed, and rain total.

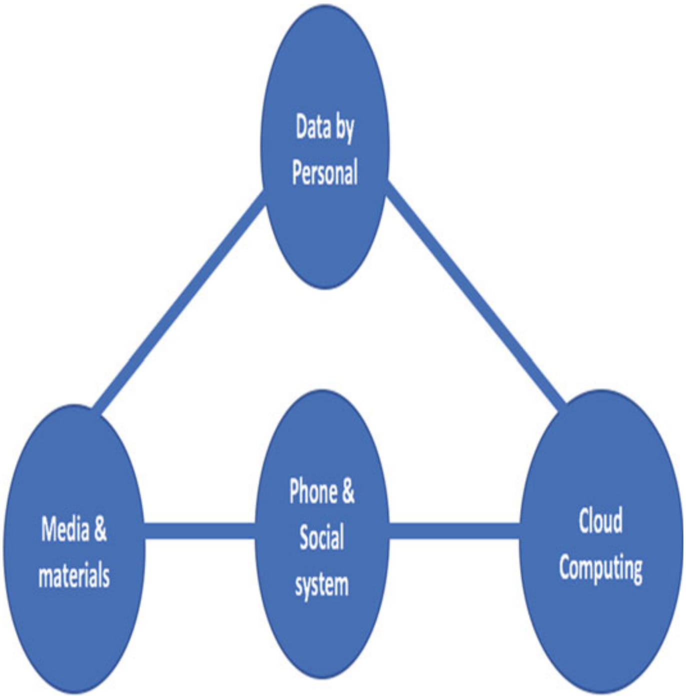


Fig. 2 Cloud computing approaches to agriculture

4 Agriculture Virtualized Information Database

We see a dispersed and monstrous stockpile of agricultural data assets like interactive media sheets that can recognize photos remotely and validate information streams. Good development is the improvement of cutting-edge agricultural data processes and technology, as well as the utilization of farming data assets and information sharing. With the help of distant detecting picture data, the harvesting growth observation element was also developed. These include data from agricultural land inspection and application areas, atmospheric composition, and land investigation; cloud computing; circulated registering; and structure filtering, which highlights virtualization as a highly reliable, commonplace technology and services to the request; cloud computing; circulated registering, and framework processing. Enrolment data, legal approval, and status screen are examples of fundamental services that are included in data administrations. Agricultural data may be retrieved and accessed by ordinary clients as well [11]. The asset server's hub may be monitored in real-time by the administrator client. A virtualized pool of asset capacity is cloud storage. It has element distribution, smooth augmentation stockpiling, and correspondence capabilities when linked to at least one programming or equipment [12]. Cloud storage is concerned with the aggregation of assets that have been sent separately. Capacity cloud assets are stored in multiple asset hubs. To accomplish high-quality information transfer, distributed capacity computing is utilized. A client's class should be specified when they include a blueprint of agricultural data that has been enlisted. According to free source programming ganglia, Hadoop cloud storage was used to build an appropriate hub for the element inspection step. Virtualized data databases for agriculture are now being created and deployed in a number of nations. It addresses the issue of blocking large amounts of agricultural data, as well as multi-duplicate capacity and agriculture asset data transmission simultaneity [9]. Server security was thoroughly checked, and farm asset security should be addressed in every research company.

5 Sensors' Importance in Preciseness of Agriculture

India's economy is one of the world's fastest-growing economies, according to the World Bank. The open agricultural collection is practiced by 58.4% of the world's population. Only 30–60% of the entire budget in India is allocated to agriculture. Indian agriculture is presently the world's second-best performer in terms of crop output. Moreover, half of all workers were employed in agriculture and related businesses, which contributed 13% to the country's GDP in 2014. Our economy relies on agriculture in many regions of the country. Despite this, the agriculture sector continues to fall behind, according to the present research. As a farming administration concept, precision agriculture (PA) is centered on analyzing bury and intra-field harvest changes and adjusting accordingly [13]. Both the spatial and transient segments of the edit fluctuation are included. A few agriculture experts have seen an increase in demand for sensor data in the agrarian framework and have looked into how to combat the practical advice, which includes the government and its policies. Information examiners, farming specialists, and others use these platforms to provide step-by-step guidance for achieving better results at reduced prices. Information advancements over the previous decade have generated a unique notion in data innovation, termed Big Data [14]. By pulling together statistics and communiques from many frameworks, this technology is helping to improve the performance of rural frameworks by minimizing opportunities and anticipated disappointments, as well as improving crop administrations and agro-leaderships. Since the volume of information and the speed at which it is being generated is so large, e-Agriculture benefit data may be described as Huge Data [15]. E-Agriculture advantage Big Data solutions include cutting-edge technologies such as HDFS, Map Reduce, Hadoop, and storm.

- The following are the focal points that help improve the agricultural framework execution and efficiency:
- Increase producer–consumer satisfaction by improving preventative care.
- To improve yield quality, measure, store, and analyze data.

- Reduce the probability of yield disappointment by keeping an eye on income costs.

6 Big Data in Food and Agriculture

Farming has become more digitalized in recent years. Data from precise rural devices and huge databases are used by even small-scale farmers to conduct precision research. Sensors are installed on the majority of John Deere tractors, which collect data on soil and product conditions. Ranchers can use this information to determine where to plant crops [16]. "Big data" refers to enormous data sets and automated techniques for obtaining, separating, and breaking them down. It used to be done through observation, but the information gathered was not digitized [17, 18]. However, however, logging data via the program should be feasible more efficiently, and agriculturists should be able to access the data [18]. At this point, Monsanto and its partners have created a set of sophisticated equipment for acquiring and analyzing ranch data, called Coordinated Field Systems (IFS). After acquiring a computerized device engineer atmosphere in 2013, Monsanto now has access to the company's expertise in crop-level monitoring devices and services. Canadian Agriculture and Agrifood's National Agricultural Information Administrations (NAIS) created Agro-climate Impact Reporter/AIR, a web-based tool. All agriculturists and volunteers' atmospheric data is gathered and made available to the agriculturists who use this device. The creators expect that the specialized rural frameworks will be used in the planning of big data as well as the marketing of big data breakthroughs [19]. Half-breed seed development, agronomy, data analysis, and precision farming will provide ranchers with crossover matches and increase the production of yield opportunities. The rancher's work is believed to be changing as a result of big data apparatuses. It is the link between innovation and social biology and human conflicts [20]. A model framework for cloud computing is shown in Fig. 3.

7 Conclusion

The role of the Internet of Things (IoT) in the agriculture industry is investigated in this study. The frequency of challenges and limits is determined by the number of IoT-based smart agricultural equipment. IoT devices are designed to be cost-effective in terms of hardware and software without sacrificing system precision. Imported devices ignore the compromise regardless of the component's price. The process' data format will be standardized, which will increase device uniformity as well as the process' execution speed. While increasing the system's goods or services, the initial process barrier providers for active farmers are governed. Farmers can use satellites and sensors to engage, analyze, and communicate with one another.

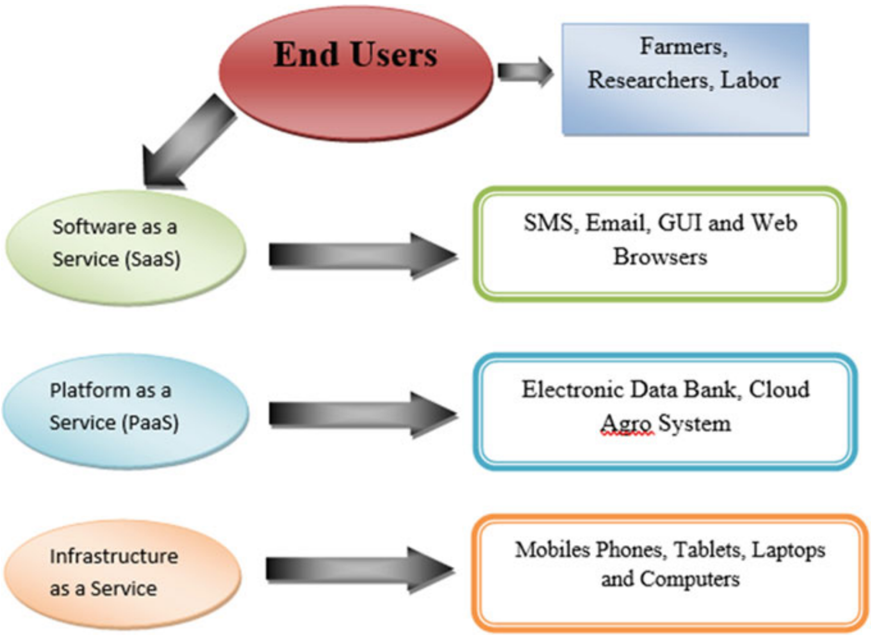


Fig. 3 Model framework for cloud computing

Cloud technology, which is a kind of IT, has agricultural advantages. With the use of agriculture cloud and IT services, agriculturalists may take benefit of a one-of-a-kind expertise administration in terms of yield development, evaluation, manures, and disease-specific treatment plan agricultural scientists will present their results and recommendations on improved development techniques, manure utilization, and the historical background of the tragedy.

The evaluation was based on the use of a cloud-based application in agriculture. We can credit the cloud for this because it boosts agricultural productivity and makes it easier for researchers in the field to obtain crucial data. As a result, money and time will be saved, and correspondence will become easier and faster. IoT in agriculture is a burgeoning discipline, and this study will contribute to a substantial body of literature on the subject.

References

1. Nath B, Chaudhuri S (2012) Application of cloud computing in agricultural sectors for economic development. In: Interplay of economics, politics and society for inclusive growth-international conference organized by RTC and GNHC. Accessed: Aug. 03, 2021. [Online]. Available:https://www.researchgate.net/profile/SomnathChaudhuri/publication/233910963_Application_of_Cloud_Computing_in_Agricultural_Sectors_for_Economic_Development/

links/0fcfd50cc89109a30b000000/Application-of-Cloud-Computing-in-Agricultural-Sectors-for-Economic-Development.pdf.
2. Choudhary S, Jadoun R, Mandoriya H (2021) Computing, and undefined 2016. Role of cloud computing technology in agriculture fields. Computing 7(3). Accessed: Aug. 03, 2021. [Online]. Available: https://core.ac.uk/download/pdf/234645129.pdf.
3. Goraya MS, Kaur H (2021) Cloud computing in agriculture. In: HCTL Open International Journal of Technology Innovations and Research (IJTIR). Accessed: Aug. 03, 2021. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.2449&rep=rep1&type=pdf.
4. Kaur H, Goraya S (2016) Role of cloud based technologies in agriculture process across countries: a review. Int J Comput Sci Inf Secur 14(9):397. [Online].Available: https://www.academia.edu/download/51103830/42_Paper_31081683_IJCSIS_Camera_Ready_pp._397-405.pdf.
5. Mekala MS, Viswanathan P (2017) A survey: smart agriculture IoT with cloud computing. In: 2017 international conference on microelectronic devices, circuits and systems (ICMDCS). Ieeexplore.ieee.org. Accessed: Aug. 03, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8211551/
6. Maharana SB, Purusottam K, Bali DK, Limma KC (2015) Application of cloud computing in agriculture development. Int J Eng Manage Res (IJEMR) 5(6):87-89. Accessed: Aug. 03, 2021. [Online]. Available: https://www.indianjournals.com/ijor.aspx?target=ijor:ijemr&volume=5&issue=6&article=014.
7. Alreshidi E (2019) Smart sustainable agriculture (SSA) solution underpinned by internet of things (IoT) and artificial intelligence (AI). [Online]. Available: http://arxiv.org/abs/1906.03106
8. Ray PP (2017) Internet of things for smart agriculture: technologies, practices and future direction. J Amb Intell Smart Environ 9(4):395–420. https://doi.org/10.3233/AIS-170440
9. Chen S, Xu H, Liu D, Hu B, Wang H (2014) A vision of IoT: applications, challenges, and opportunities with china perspective. IEEE Int Things J 1(4):349–359. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6851114/
10. Doshi J, Patel T, Kumar Bharti S (2019) Smart Farming using IoT, a solution for optimally monitoring farming conditions. Proced Comput Sci 160:746–751. Accessed: Aug. 03, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919317168.
11. Malik AD, Jamil A, Omar KA, Abd MH (2021) Implementation of faulty sensor detection mechanism using data correlation of multivariate sensor readings in smart agriculture. Ann Emerg Technol Comput (AETiC) 5(5):1–9. https://doi.org/10.33166/AETiC.2021.05.001.
12. Mercier-Laurent E (2020) The future of AI or AI for the future. IFIP Adv Inf Commun Technol 555:20–37. https://doi.org/10.1007/978-3-030-64246-4_3
13. Salim C, Mitton N (2021) K-predictions based data reduction approach in WSN for smart agriculture. Computing 103(3):509–532. https://doi.org/10.1007/S00607-020-00864-Z
14. Nóbrega L, Gonçalves P, Pedreiras P, Pereira J (2019) An IoT-based solution for intelligent farming. Sensors19(3):603. https://doi.org/10.3390/s19030603
15. Suji Prasad SJ, Thangatamilan M, Suresh M, Panchal H, Rajan CA, Sagana C, Gunapriya B, Sharma A, Panchal T, Sadasivuni KK (2021) An efficient LoRa-based smart agriculture management and monitoring system using wireless sensor networks. Int J Amb Energy 1–4. https://doi.org/10.1080/01430750.2021.1953591
16. Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX (2017) A review on the practice of big data analysis in agriculture. Comput Electron Agric 143:23–37.https://doi.org/10.1016/j.compag.2017.09.037
17. White EL et al (2021) Report from the conference, 'identifying obstacles to applying big data in agriculture.' Precis Agric 22(1):306–315. https://doi.org/10.1007/S11119-020-09738-Y
18. Rejeb KR, Zailani S (2021) Big data for sustainable agri-food supply chains: a review and future research perspectives. J Data, Inf Manag https://doi.org/10.1007/S42488-021-00045-3

19. Roux J, Escriba C, Fourniols JY, Soto-Romero G (2019) A new bi-frequency soil smart sensing moisture and salinity for connected sustainable agriculture. J Sens Technol 9(3):35–43. Accessed: Aug. 03, 2021. [Online]. Available: https://www.scirp.org/journal/paperinformation.aspx?paperid=95804
20. Cadavid H, Garzón W, Pérez A, López G, Mendivelso C, Ramírez C (2018) Towards a smart farming platform: from IoT-based crop sensing to data analytics, vol 885. Springer, pp 237–251. https://doi.org/10.1007/978-3-319-98998-3_19

Analyzing the Sources of Air Pollution and Comparing Its Impact During the Phases of COVID-19 Pandemic and the Scope of IoT in Monitoring Air Quality

V. Sahaya Sakila and A. R. Kavitha

Abstract Air Pollution is a serious threat in the enlightened globe that impacts atmospheric change and human health. Air pollutants like Particulate Matter $PM_{2.5}$, Particulate Matter PM_{10}, ground-level ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2) and carbon monoxide (CO) lead to air pollution either in the form of gases or particles suspended in the air. The contemporary study aimed to analyze the source of air pollution and change in air quality during lockdown periods due to 1st and 2nd wave of COVID-19 pandemic within six different states namely, Uttar Pradesh, Madhya Pradesh, Delhi, Haryana, Rajasthan and Punjab. The selected states were analyzed in four different phases (i) Before First Covid Wave Lockdown (January 1, 2020 to March 24, 2020), (ii) During First Covid Wave Lockdown (March 25, 2020 to May 31, 2020), (iii) After First Covid Wave Lockdown (June 1, 2020 to April 19, 2021), (iv) During Second Covid Wave Lockdown (April 20, 2021 to May 26, 2021) from 45 air quality monitoring stations. This study also explains the limitations in stationary Ambient Air Monitoring Stations and the scope of IoT technology to improve the standards of monitoring air quality.

Keywords Air pollution · AQI · Particulate matter ($PM_{2.5}$, PM_{10}) · NO_2 · O_3 · SO_2 · CO · COVID-19 · IoT

1 Introduction

Air Quality is one of the most essential needs for human beings, plants and animals. In general, 14,000 L of air are inhaled by a person every day. Air is a substance that contains 78.09% nitrogen, 20.95% oxygen, 0.93% argon, 0.39% carbon dioxide and a small volume of other gases [1]. But in a real scenario, elements/compounds like Particulate Matter $PM_{2.5}$ (A particulate matter $\leq 2.5\ \mu m$ in aerodynamic diameter), Particulate Matter PM_{10} (A particulate matter $\leq 10\ \mu m$ in aerodynamic diameter), ground-level ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon

V. Sahaya Sakila (✉) · A. R. Kavitha
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai 600026, India
e-mail: sv5969@srmist.edu.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_16

monoxide (CO) are also indirectly added to the pure air through different means that has put enormous pressure on the environment in the form of air pollution and these pollutants lead to more serious respiratory disorders which reduce the efficiency of lung functioning [1, 2]. State of Global Air 2018 report took most evidences from Global Burden of Disease (GBD) project and published a 'Global ranking list for the number of deaths that occurred among all the age groups and all sexes around the world, 2016', in which the deaths caused due to Ambient particulate matter ($PM_{2.5}$) had been ranked 6th among 33 risk factors ranked by it. PM $_{2.5}$ is one of the major pollutants of air pollution and it was responsible for over 4.1 million deaths that occurred worldwide from the diseases like respiratory infections, heart disease and stroke, chronic lung disease and lung cancer [3].

According to the World Health Statistics 2016, it has described that both the indoor and outdoor air pollution were the world's largest environmental health risk factor which caused a death rate as nearly as 7 million or 11.6% of total deaths in the year 2012 [4]. Outdoor air pollution emerges from natural and anthropogenic sources. Natural source of air pollution occurs in the form of forest fires and dust storms whereas anthropogenic sources are defined by human activities such as heat and power generated by power plant, Industrial facilities with Manufacturing Yard, burning of agricultural waste, fuel consumption from automobiles [5]. Due to urbanization, transport has become the main contributor of outdoor air pollution, which has become as a major health hazard [4, 6].

Indoor air pollution is also one of the major health hazards and it is caused due to high-risk substances released from the human activities such as heating, lighting and cooking with polluting fuels or pollutants that are released from construction materials, buildings and from indoor equipments [7]. Based on GBD analysis, it had stated that household air pollution had caused 2.1 million of attributable deaths in 2016 and it was also ranked 8th for premature deaths across the globe [3].

The novel corona virus (SARS-CoV-2) named as COVID-19 disease, has become a global pandemic as it was declared by World Health Organization (WHO) on 11 March 2020 [8]. The first human affected for COVID-19 was found in Wuhan City, China in December 2019, subsequently the outbreak of disease COVID-19 from Wuhan City started to spread almost all the cities across the world [9]. According to the current evidence SARS-CoV-2 virus is mainly transmitted from one person to another person through close contact (within one meter) by airborne respiratory droplets which are between 5 and 10 μm in diameter [10].

In India, first positive case for COVID-19 disease was found in Kerala who was a 20-year-old female student who returned from Wuhan, China [11]. To avoid the rapid spread of COVID-19 epidemic, Government of India issued an order for National wide lockdown, which was initially announced for 21 days as phase-1 and later it got extended till phase-4 during the first wave of COVID-19 [12, 13]. In order to mitigate the infection caused by COVID-19 disease, some containment measures got implemented in India during first wave of COVID-19, where Autonomous Bodies, Corporation, Industrial Establishment, all Transport Services (Air, Rail, Roadways), Hospitality, all Educational Institutions, all Worship Centers, all gatherings (social, political, entertainment, academic, cultural, religious functions) were kept barred

[14]. When the cases of COVID-19 disease were getting considerably decreased, the second wave of COVID-19 had already started during June–August 2020 across the European and other regions of the World [15]. In India, second wave of COVID-19 disease had started in the middle of March 2021, subsequently the highest number of cases was recorded in India as 144,829, on April 09 2021. Europe and India were experiencing massive increase of COVID-19 cases and fatality rate [15]. During the second wave of COVID-19 disease, as the nation-wide lockdown was not announced to reduce the escalation of infection, states in India imposed many lockdown measures like complete/partial lockdown with strict counter measures [16].

Lockdown has helped us discover a chance to find out the major pollutants present in the air due to the human activities and to identify the sources of poor air quality [17]. In this study, it has been estimated how the air quality has improved because of lockdown across the different states (Delhi, Uttar Pradesh, Madhya Pradesh, Rajasthan, Haryana and Punjab) in India and exposes the limitations of fixed ambient air quality monitoring stations as well as introduces the IoT [18] based atmospheric air pollution monitoring system.

2 Materials and Method

2.1 Selection of Study Region

This study mainly focusses on the change in quality of air observed during the implementation of lockdown measures. Thus, the intention of this work is to (I) Evaluate the impact of lockdown strategies in major states like Delhi, Madhya Pradesh, Punjab, Rajasthan, Haryana and Uttar Pradesh (II) Analyze and compare the air quality during four phases such as BFL (Before First Wave Lockdown), DFL (During First Wave Lockdown), AFL (After First Wave lockdown) and DSL (During Second Wave Lockdown) (III) Assess the impact of AQI (Air Quality Index) during the first and second wave of COVID-19 disease (IV) Introduce IoT technology to improve the efficiency of air quality monitoring methods.

2.2 Sampling and Analysis

In order to evaluate the quality of air, ambient air sampling and analysis was done for the pollutant levels of major air pollutants such as Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), Carbon Monoxide (CO) and Particulate Matters ($PM_{2.5}$, PM_{10}) through 45 Air Quality Monitoring Stations positioned at different towns/cities of the chosen states. The impact of lockdowns had significant effects on mainly NO_2, SO_2 and CO as vehicle emissions act as the primary source for these pollutants. The sources of $PM_{2.5}$ and PM_{10} are from different means such as industrial processes, operations in agriculture, construction activities, demolition activities, dust, burning of wood, vehicles, etc.,

2.3 Data Collection

Due to the outbreak of COVID-19, the World Air Quality Index Project team provided a new dedicated dataset for air quality covering 380 major cities in the world. Delhi, Uttar Pradesh, Madhya Pradesh, Rajasthan, Haryana and Punjab are chosen to evaluate the impact of air quality during the lockdown period (https://aqicn.org/data-platform/covid19/).

2.4 Calculation Ratio of AQI

AQI is the standard given to measure the air quality and it ranges from 0 to 500. When the value of the AQI is higher, it leads to health issues. The range specified for the AQI is categorized into six divisions represented by different colors as described in Fig. 1.

Calculation of AQI

Date	DD-MM-YYYY
Station	NSIT
City	Delhi
State	Delhi

Pollutants		concentration in µg/m3 (except for CO)	Sub-Index	check
PM_{10}	24-hr avg	121.00	114	1
$PM_{2.5}$	24-hr avg	34.00	57	1
SO_2	24-hr avg	0.00	0	0
NO_2	24-hr avg	8.00	10	1
*CO (mg/m3)	max 8-hr	0.00	0	0
O_3	max 8-hr	57.00	57	1

Air Quality Index

AQI = 114

* Concentrations of minimum three pollutants are required; one of them should be PM10 or PM2.5
* The check displays "1" when a non-zero value is entered

Good (0–50)	Minimal Impact	Poor (201–300)	Breathing discomfort to people on prolonged exposure
Satisfactory (51–100)	Minor breathing discomfort to sensitive people	Very Poor (301–400)	Respiratory illness to the people on prolonged exposure
Moderate (101–200)	Breathing discomfort to the people with lung, heart disease, children and older adults	Severe (>401)	Respiratory effects even on healthy people

Fig. 1 Describes the calculating index of AQI by CPCB

AQI is calculated with the sub-indices of the pollutant's 24-h average concentration level taken from the air quality monitoring stations except for CO and $\mathbf{O_3}$. AQI can be calculated with a minimum of three pollutant's concentration data in which it is mandatory that one pollutant should be either $PM_{2.5}$ or PM_{10}. Similarly, the pollutant's data gathered from the air quality monitoring station should be minimum of 16-h data. With all the necessary data, it is possible to calculate the AQI and the worst Sub-index is considered to be the AQI for that particular area and its calculation factor has been described in the Fig. 1.

3 Results

3.1 Analysis of NO_2

The desired range for NO_2 to be considered as 'Good' should be between 0 and 40 $\mu g/m^3$, 41 and 80 $\mu g/m^3$ to be considered 'Satisfactory', 81 and 180 $\mu g/m^3$ to be considered 'Moderate', will be considered as 'Poor' if the range is between 181 and 280 $\mu g/m^3$, 'Very poor' if the range is between 281 and 400 $\mu g/m^3$ and 'Severe' if the range exceeds 400 $\mu g/m^3$.

Analysis of $\mathbf{NO_2}$ in Delhi. Among all the air monitoring stations in Delhi, Jahangirpuri station recorded huge reduction in NO_2 levels approximately by 60.5% and 30.9% DFL and DSL, respectively. Overall, the NO_2 levels in Delhi had seen huge reduction of 40.7% and 32.8% DFL and DSL, respectively. Notably, the average NO_2 observed at Jahangirpuri station was 55 $\mu g/m^3$ BFL which was considered to be in the range of 'Satisfactory' NO_2 levels and 24 $\mu g/m^3$ DFL that comes under the range of 'Good' NO_2 levels. All the other monitoring stations have also recorded significant improvement in air quality during both the covid wave lockdowns approximately between 30 and 60%.

Analysis of $\mathbf{NO_2}$ in Uttar Pradesh. Most of the monitoring stations in UP had shown huge reduction in NO_2 levels, approximately around 50–80%. Central School Lucknow air monitoring station has recorded 114 $\mu g/m^3$ BFL that lies just in the 'Moderate' NO_2 range and 13 $\mu g/m^3$ DFL that moved under the 'Good' range of NO_2 levels. Other stations had recorded NO_2 levels in the range of 40–80 $\mu g/m^3$ which is clearly in the range of just 'Satisfactory' during BFL and moved under the range of 'Good' during DFL. During DSL, Jai Bhim Nagar Meerut, Indirapuram Ghaziabad and Sector 1 Noida had benefitted around 50% to 80%. Overall, the state of Uttar Pradesh had reduction in NO_2 levels by 58.5 and 23.2% during DFL and DSL, respectively.

Analysis of $\mathbf{NO_2}$ in Madhya Pradesh. Most of the stations registered huge reduction in NO_2 levels around 50% to 80% and 30% to 60% during DFL and DSL, respectively. Indore and Singrauli had registered maximum NO_2 levels of 64 $\mu g/m^3$ and 68 $\mu g/m^3$ respectively during BFL in the range of 'Satisfactory' NO_2 levels and 25 $\mu g/m^3$ and 32 $\mu g/m^3$ respectively during DFL which are in the range of

'Good' NO_2 levels. During AFL, Singrauli had registered 87 μg/m^3 in the range of 'Moderate' NO_2 levels and 28 μg/m^3 DSL in the range of 'Good' NO_2 levels. Overall, the state of Madhya Pradesh had been highly benefitted by 52.7% and 32.7% in their NO_2 levels during DFL and DSL, respectively.

Analysis of NO_2 in Haryana. Gobind Pura station had registered reduction in the NO_2 levels during DFL and DSL by 32.41% and 20.68%, respectively. During AFL, Gobind Pura and Sector 11 Faridabad had registered NO_2 levels of 43 μg/m^3 and 44 μg/m^3 in the range of 'Satisfactory' NO_2 level and 21 μg/m^3 and 11 μg/m^3 in the range of 'Good' NO_2 level, respectively during DSL. Overall, the state of Haryana had been highly benefitted by 32.4% and 24.6% during DFL and DSL, respectively.

Analysis of NO_2 in Punjab. During DFL, the NO_2 levels registered at Jalandhar, Khanna and Mandi Gobindga had seen huge reduction by 57.8%, 49.1% and 39.2% respectively. During AFL, Amritsar and Mandi Gobindga monitoring stations had registered 114 μg/m^3 and 100 μg/m^3 in the range of 'Moderate' and 9 μg/m^3 and 12 μg/m^3 respectively during DSL in the range of 'Good' NO_2 levels. Overall, the state of Punjab had significantly been benefitted in their NO_2 levels by 23.9% during DFL.

Analysis of NO_2 in Rajasthan. Jaipur had registered reduction of NO_2 levels by 66.4% and 44.6%, Jodhpur by 35.8% and 36.3%, Kota by 60% and 7.4%, Moti Doongri by 26% and 31.1%, Udaipur by 69% and 25.8%, respectively during DFL and DSL. During BFL, Jaipur and Jodhpur had registered NO_2 levels of maximum 43 μg/m^3 and 46 μg/m^3 in the range of 'Satisfactory' NO_2 levels, 12 μg/m^3 and 19 μg/m^3 in the range of 'Good' NO_2 levels during DFL. Overall, the state of Rajasthan had been a huge beneficiary in their NO_2 levels by 51.5% and 29.1% during DFL and DSL, respectively.

3.2 *Analysis of $PM_{2.5}$*

The desired range for $PM_{2.5}$ to be considered as 'Good' should be between 0 and 30 μg/m^3, 31 and 60 μg/m^3 to be considered 'Satisfactory', 61 and 90 μg/m^3 to be considered 'Moderate', will be considered as 'Poor' if the range is between 91 and 120 μg/m^3, 'Very poor' if the range is between 121 and 250 μg/m^3 and 'Severe' if the range exceeds 250 μg/m^3.

Analysis of $PM_{2.5}$ in Delhi. During DFL, stations at Delhi had shown high decline in their $PM_{2.5}$ levels around 30–50% and around 15–30% approximately during DSL. Most of the air monitoring stations in Delhi had registered $PM_{2.5}$ levels in the range of 'Very Poor' category (121–250 μg/m^3) and 'Poor' category (91–120 μg/m^3), during BFL and DFL, respectively. Bagh, Jahangirpuri, Mundka, Narela and Wazirpur had significant improvement in their $PM_{2.5}$ levels by 28.7%, 41.7%, 41.4%, 29.6% and 42.8%, respectively. Overall, the state of Delhi had highly been benefitted in their $PM_{2.5}$ levels by 38.2% and 19.9% during DFL and DSL, respectively.

Analysis of $PM_{2.5}$ in Uttar Pradesh. Stations at Uttar Pradesh had shown huge decline in their $PM_{2.5}$ levels over 25% to 60% and around 15% to 30% during

DFL and DSL, respectively. 7 out of 13 air monitoring stations had recorded $PM_{2.5}$ levels in the range of 'Very Poor' category (121–250 $\mu g/m^3$) and in the range of 'Poor' category (91–120 $\mu g/m^3$), during BFL and DFL, respectively. Lajpat Nagar Moradabad station moved from 'Very poor' category to 'Moderate' category as the average values dropped from 182.24 $\mu g/m^3$ to 72.33 $\mu g/m^3$. Even during DSL, all the stations had shown significant improvements around 15–30%. Overall, the state of Uttar Pradesh had been highly benefitted around 41.5 and 21.2% during DFL and DSL, respectively.

Analysis of $PM_{2.5}$ in Madhya Pradesh. During DFL and DSL, the decline shown in the concentration levels of $PM_{2.5}$ were around 20% to 40% and around 10% to 40%, respectively. 4 out of 7 air monitoring stations had $PM_{2.5}$ levels that moved from 'Very Poor' category (121–250 $\mu g/m^3$) to 'Moderate' category (61–90 $\mu g/m^3$) during DFL. During DSL, majority of the stations had jumped from 'Poor' category (91–120 $\mu g/m^3$) to 'Moderate' category (61–90 $\mu g/m^3$). Overall, state of Madhya showed very significant decline in the concentration levels of $PM_{2.5}$ of around 27.9% and 18.7% during DFL and DSL, respectively.

Analysis of $PM_{2.5}$ in Haryana. Gobind Pura and Sector 16a Faridabad stations had recorded decline in their $PM_{2.5}$ concentrations by 28.8% and 39.6%, respectively during DFL and by 6.8% and 24.2%, respectively during DSL. They had recorded improvement in their $PM_{2.5}$ levels from 'Very Poor' to 'Poor' category during DFL. Overall, the state of Haryana had got benefitted in their PM2.5 concentrations by 34.2% and 10.9% during DFL and DSL, respectively.

Analysis of $PM_{2.5}$ in Punjab. All the 5 air monitoring stations had registered decline in their $PM_{2.5}$ concentrations around 25–40% approximately DFL and around 5% approximately during DSL. Amritsar, Jalandhar and Mandi Gobindga air monitoring station's average $PM_{2.5}$ concentration data were in range of 'Very Poor' category during BFL and 'Poor' category during DFL. Overall, the state of Punjab had benefitted highly in their $PM_{2.5}$ levels during DFL, by 29.6%. Figure 2 reports the reduction in the concentration level of the pollutants from the analysis.

Analysis of $PM_{2.5}$ in Rajasthan. The $PM_{2.5}$ concentrations recorded in stations at Rajasthan saw significant drop in their values around 15% to 40% and 5% to 25% approximately during DFL and DSL, respectively. The average $PM_{2.5}$ concentration recorded at Jaipur and Moti Doongri were in the range of 'Very Poor' category during BFL and moved to 'Poor' category during DFL. During DSL also, Jaipur and Moti Doongri stations had shown similar improvements in their $PM_{2.5}$ concentrations. Overall, the state of Rajasthan was highly benefitted in their $PM_{2.5}$ levels by 24.3% and 7.4% during DFL and DSL, respectively.

3.3 Analysis of PM_{10}

The desired range for PM_{10} to be considered as 'Good' should be between 0 and 50 $\mu g/m^3$, 51 and 100 $\mu g/m^3$ to be considered 'Satisfactory', 101 and 250 $\mu g/m^3$ to be considered 'Moderate', will be considered as 'Poor' if the range is between

State	City	$PM_{2.5}$			PM_{10}			O_3			NO_2			SO_2			CO		
		DFL	AFL	DSL	DFL	AFL	DSL	DFL	AFL	DSL	DFL	AFL	DSL	DFL	AFL	DSL	DFL	AFL	DSL
Delhi	Alipur	-34.6%	52.6%	-18.5%	-21.1%	71.1%	-0.0%	88.9%	-25.5%	58.5%	-43.3%	23.9%	-14.9%	16.2%	-6.3%	25.1%	-34.8%	49.1%	38.9%
	Bagh	-28.7%	35.5%	-17.3%	-13.6%	38.3%	-23.4%	94.4%	8.2%	85.6%	-39.3%	81.0%	-49.8%	240.8%	-63.8%	33.6%	-49.5%	129.4%	-48.1%
	Jahangirpuri	-41.7%	66.5%	-25.5%	-45.2%	110.3%	-25.7%	284.4%	119.4%	94.3%	-60.5%	-9.4%	-30.9%	14.1%	28.8%	91.5%	-44.8%	73.0%	-39.4%
	Mandka	-41.4%	59.1%	-21.4%	-41.7%	99.7%	-25.4%	25.5%	-41.1%	46.3%	-19.1%	7.9%	-37.7%	0.9%	-12.0%	30.1%	-50.5%	104.7%	-36.9%
	Narela	-29.6%	40.2%	-15.8%	-21.5%	62.1%	-11.7%	59.7%	-62.6%	101.7%	-26.8%	31.3%	-31.3%	-5.8%	-31.4%	75.4%	-4.7%	32.0%	-44.7%
	Pusa	-42.7%	57.7%	-23.1%	-42.8%	104.6%	-22.9%	83.7%	-45.8%	67.8%	-43.6%	54.5%	-25.4%	-63.4%	123.1%	33.9%	-70.0%	191.0%	-60.0%
	Rk Puram	-39.0%	73.4%	-19.4%	-49.5%	102.6%	-10.6%	116.9%	-23.6%	114.4%	-50.1%	100.5%	-29.6%	31.0%	11.9%	44.9%	-41.0%	182.9%	0.9%
	Satyawati College	-42.1%	67.9%	-23.2%	-60.5%	131.2%	-23.7%	46.6%	6.5%	84.8%	-60.5%	168.6%	-23.9%	19.9%	29.4%	36.5%	-38.6%	144.1%	-11.6%
	Sonia Vihar	-38.9%	69.8%	-19.5%	-35.8%	104.6%	-25.5%	-8.0%	1.8%	36.3%	-40.1%	60.7%	-28.6%	2.9%	27.9%	24.7%	17.4%	13.3%	-12.4%
	Wazirpur	-42.8%	40.9%	-14.6%	-61.4%	119.7%	-20.1%	403.4%	-39.0%	66.2%	-24.2%	60.0%	-52.9%	73.2%	36.5%	-73.1%	-23.9%	20.9%	-64.2%
Haryana	Faridabad	NA	NA	-7.8%	NA	116.6%	-30.9%	NA	-11.4%	155.6%	NA	-64.4%	-63.6%	NA	-50.9%	-41.6%	NA	-28.7%	-40.6%
	Gobind Pura	-28.8%	42.3%	-6.8%	-23.6%	53.9%	-10.3%	11.0%	5.2%	88.4%	-32.4%	46.2%	-20.7%	17.2%	-52.0%	46.8%	-53.0%	159.4%	-26.1%
	Sector 11 Faridabad	NA	39.9%	-5.2%	NA	117.4%	-31.7%	NA	24.5%	7.8%	NA	191.6%	-4.5%	NA	117.5%	-49.4%	NA	13.4%	-54.3%
	Sector 16a Faridabad	-30.6%	66.8%	-24.2%	NA	NA	NA	26.1%	-33.0%	146.0%	NA	NA	-27.9%	-81.8%	316.8%	-7.7%	-64.0%	182.0%	-35.7%
	Sector 30 Faridabad	NA	68.9%	-10.7%	NA	66.5%	-14.3%	NA	26.6%	58.1%	NA	-13.0%	-6.4%	NA	-60.3%	-37.5%	NA	114.0%	-60.8%
MP	Bhopal	-31.6%	23.3%	-23.2%	-25.9%	15.9%	-14.7%	5.5%	-6.4%	22.8%	-78.4%	219.2%	-38.1%	-20.0%	21.3%	59.5%	-55.2%	122.5%	-44.5%
	Dewas	-19.0%	9.5%	-18.2%	-5.4%	-4.0%	-1.4%	52.8%	-29.4%	-8.4%	-37.0%	90.3%	-46.2%	51.4%	4.1%	6.3%	-14.2%	-29.3%	-30.1%
	Gwalior	NA	50.4%	-36.0%	NA	30.9%	-24.6%	NA	-34.2%	25.5%	NA	183.3%	-39.6%	NA	48.3%	-33.3%	NA	90.6%	-51.1%
	Indore	-36.3%	36.5%	-19.4%	-20.9%	9.9%	-2.7%	94.3%	3.7%	19.4%	-51.8%	14.2%	-55.9%	-39.1%	80.0%	13.2%	-65.8%	205.7%	-38.0%
	Jabalpur	-34.4%	27.8%	-22.2%	-34.5%	44.1%	-26.6%	34.5%	-23.3%	24.3%	-66.9%	87.5%	-24.4%	-24.0%	136.0%	22.1%	-57.9%	147.6%	-28.9%
	Singrauli	-17.4%	12.0%	-0.8%	0.7%	-4.8%	3.9%	37.0%	-68.6%	23.0%	-26.2%	-16.8%	26.0%	14.6%	-37.6%	11.9%	-21.7%	65.1%	24.6%
	Ujjain	-28.7%	25.8%	-11.4%	-13.5%	17.4%	-3.7%	0.7%	12.8%	2.7%	-65.9%	130.0%	-60.4%	-12.5%	-13.2%	-16.1%	-41.6%	26.1%	-33.7%
Punjab	Amritsar	-36.5%	52.7%	1.0%	-27.4%	63.6%	-18.3%	-37.9%	22.5%	31.2%	24.6%	65.0%	-63.1%	-0.3%	-0.6%	22.5%	9.5%	-25.0%	18.1%
	Jalandhar	-28.1%	27.9%	0.2%	-30.5%	59.6%	-0.3%	-0.4%	10.3%	-22.9%	-57.8%	123.1%	18.7%	-25.0%	-72.7%	-56.5%	-49.4%	80.8%	-16.4%
	Khanna	-25.0%	35.1%	-1.8%	-18.6%	44.2%	-6.3%	9.7%	14.9%	-17.0%	-49.1%	205.7%	38.0%	-29.2%	8.7%	64.8%	-100.0%	NA	-19.6%
	Mandi Gobindga	-27.7%	37.4%	-1.1%	-26.2%	39.4%	-0.2%	41.2%	-34.9%	16.0%	-39.2%	91.6%	63.1%	-54.2%	-13.6%	13.5%	26.3%	-9.4%	2.6%
	Patiala	-31.8%	52.5%	-3.2%	-22.8%	53.7%	0.3%	87.1%	33.6%	34.6%	2.0%	165.6%	15.4%	-18.7%	19.3%	-10.0%	-63.2%	140.5%	-32.0%
Rajasthan	Jaipur	-19.1%	19.0%	-7.0%	-13.6%	25.3%	-15.1%	-0.9%	-10.6%	2.7%	-66.4%	117.3%	-44.6%	-4.1%	-9.8%	-4.5%	-47.7%	71.2%	-36.7%
	Jodhpur	-15.0%	15.4%	-2.8%	-17.9%	25.6%	6.1%	34.4%	-39.7%	2.5%	-35.8%	39.5%	-36.3%	-40.9%	95.4%	-22.9%	-34.4%	47.6%	-18.9%
	Kota	-23.2%	37.7%	0.7%	-20.8%	41.0%	5.9%	15.3%	-16.4%	29.6%	-60.0%	78.0%	-7.4%	0.1%	7.3%	10.5%	-21.8%	38.6%	-4.9%
	Moti Doongri	-37.0%	58.8%	-22.3%	-29.6%	47.6%	-14.4%	-0.1%	-13.4%	-23.7%	-26.0%	7.4%	-31.1%	-31.1%	-9.2%	-4.3%	4.1%	42.6%	-25.9%
	Udaipur	-27.3%	37.7%	-6.7%	-25.3%	36.5%	4.9%	-4.4%	-12.8%	-17.7%	-69.0%	257.1%	-25.8%	-30.2%	93.6%	-31.8%	-67.7%	144.1%	-46.1%
UP	Ardhali Bazar Varanasi	-34.8%	28.4%	-7.8%	NA	NA	-27.9%	35.3%	-19.8%	-4.7%	-14.2%	66.4%	79.4%	112.2%	-24.5%	-13.4%	NA	NA	NA
	Central School Lucknow	-34.0%	49.3%	-22.3%	NA	NA	-45.3%	1.4%	87.5%	-62.5%	-86.5%	377.9%	-6.1%	45.0%	23.8%	-41.4%	-27.1%	11.0%	-10.2%
	Ganga Nagar Meerut	-25.4%	32.8%	-20.8%	-9.5%	37.7%	-6.8%	37.5%	2.6%	-47.4%	-26.0%	43.8%	-34.2%	-26.1%	4.2%	-2.8%	1.2%	-11.7%	-100.0%
	Gomti Nagar Lucknow	-25.6%	41.6%	-26.0%	NA	NA	NA	-3.3%	6.1%	37.5%	-67.1%	159.0%	-46.8%	-28.4%	3.9%	24.2%	NA	NA	NA
	Indirapuram Ghaziabad	-45.3%	71.3%	-22.8%	-43.6%	111.1%	-19.4%	116.4%	-19.8%	37.0%	-62.7%	151.7%	-60.3%	17.2%	2.3%	-18.4%	-47.4%	108.2%	-32.0%
	Jai Bhim Nagar Meerut	-100.0%	NA	-19.9%	-100.0%	NA	4.1%	-100.0%	NA	1.2%	-100.0%	NA	-84.3%	-100.0%	NA	27.9%	-100.0%	NA	-35.4%
	Lajpat Nagar Moradabad	-60.3%	153.1%	-23.8%	-62.6%	223.6%	-28.0%	36.4%	549.9%	-9.6%	-67.8%	20.7%	-37.0%	-62.5%	80.3%	4.4%	-13.9%	12.5%	-33.5%
	Loni Ghaziabad	-37.5%	65.0%	-21.9%	-23.2%	78.0%	-10.9%	53.3%	-17.1%	57.7%	-54.2%	114.5%	-35.2%	34.4%	-4.4%	12.4%	-1.5%	31.0%	-100.0%
	Pallavpuram	-36.9%	39.4%	-28.4%	-19.6%	59.3%	-12.4%	8.8%	-47.7%	-41.4%	-54.6%	467.9%	35.4%	2.2%	23.6%	74.7%	32.1%	-45.8%	-10.5%
	Sanjay Nagar Ghaziabad	-34.7%	48.7%	-22.5%	-30.6%	94.8%	-16.4%	28.3%	-8.0%	79.5%	-38.2%	108.9%	-34.1%	37.3%	-27.5%	89.6%	-19.7%	57.4%	-23.0%
	sector 1 Noida	-45.9%	59.9%	-27.3%	-35.2%	68.8%	-16.1%	21.7%	-61.7%	-14.4%	-71.9%	312.6%	-64.8%	8.7%	-26.7%	90.7%	-15.4%	3.0%	-63.7%
	Sector116 Noida	-39.4%	49.2%	-19.4%	-35.0%	88.3%	-11.6%	66.4%	-26.7%	-100.0%	-65.9%	141.0%	-9.6%	-27.1%	-10.3%	133.4%	-26.5%	36.7%	-44.4%
	Talkatora Lucknow	-29.6%	39.7%	-12.7%	NA	NA	-48.4%	NA	NA	17.8%	-61.5%	215.1%	-24.6%	-73.5%	261.3%	-13.7%	-32.5%	31.9%	13.9%

Fig. 2 Comparison Table describes the change in percentage (%) of six pollutants namely Particulate Matter ($PM_{2.5}$ and PM_{10}), ground-level ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO) during the 4 phases BFL, DFL, AFL, DSL. (NA: Data Not Available; DFL: During First Wave Lockdown; AFL: After First Wave Lockdown; DSL: During Second Wave Lockdown.)

251 and 350 µg/m^3, 'Very poor' if the range is between 351 and 430 µg/m^3 and 'Severe' if the range exceeds 430 µg/m^3.

Analysis of PM_{10} in Delhi. Delhi was highly benefitted around 20% to 55% and around 10% to 30% approximately in their PM_{10} levels during DFL and DSL, respectively. During BFL, the PM_{10} levels were in the range of 'Moderate' (101–250 µg/m^3) and moved to 'Satisfactory' category (51–100 µg/m^3) during DFL. Overall, the state of Delhi was highly benefitted in their PM_{10} levels by 37.3% and 19.8% during DFL and DSL, respectively.

Analysis of PM_{10} in Uttar Pradesh. PM_{10} concentrations in UP had highly declined around 10% to 65% and 10% to 50% approximately during DFL and DSL, respectively. During DFL, the average PM_{10} concentrations moved from 'Moderate' to 'Satisfactory' category. Even during DSL, Ardhali Bazar Varanasi and Central School Lucknow stations, their average PM_{10} concentrations moved from 'Moderate' to ‘Satisfactory’ category. Overall, the state of Uttar Pradesh had highly benefitted in their PM_{10} levels by 39.9% and 19.9% during DFL and DSL, respectively.

Analysis of PM_{10} in Madhya Pradesh. PM_{10} concentrations in MP had highly declined around 10% to 35% and 5% to 30% during DFL and DSL, respectively. Overall, the average improvement in the PM_{10} levels recorded was around 17.1% and 10% during DFL and DSL, respectively.

Analysis of PM_{10} in Haryana. The PM_{10} concentration in the state of Haryana had highly declined around 25% and around 10% to 35% during DFL and DSL, respectively. Gobind Pura station’s PM_{10} concentration data had recorded huge improvement as their average PM_{10} levels moved from ‘Moderate’ to ‘Satisfactory’ category during DSL. Overall, the state of Haryana was significantly benefitted in their PM_{10} levels by 23.6% and 21.8% during DFL and DSL, respectively.

Analysis of PM_{10} in Punjab. Punjab’s PM_{10} concentration had highly declined around 15% to 30% and 5% to 20% approximately during DFL and DSL, respectively. Overall, the state of Punjab had significantly been benefitted in their PM_{10} levels by 25.1% and 4.7% during DFL and DSL, respectively.

Analysis of PM_{10} in Rajasthan. Rajasthan’s PM_{10} concentration had shown huge decline around 15% to 30% and 15% during DFL and DSL, respectively. Jodhpur station had recorded massive improvement in the PM_{10} levels as they moved from ‘Moderate’ to ‘Satisfactory’ category. Overall, the state of Rajasthan was highly benefitted in their PM_{10} levels by 21.4% and 2.5% during DFL and DSL, respectively.

3.4 Analysis of CO

The desired range for CO to be considered as ‘Good’ should be between 0 and 1.0 mg/m^3, 1.1 and 2.0 ^{mg}/m^3 to be considered ‘Satisfactory’, 2.1 and 10 ^{mg}/m^3 to be considered ‘Moderate’, will be considered as ‘Poor’ if the range is between 10.1 and 17 ^{mg}/m^3, ‘Very poor’ if the range is between 17.1 and 34 ^{mg}/m^3 and ‘Severe’ if the range exceeds 34 ^{mg}/ m^3.

Analysis of CO in Delhi. Delhi’s CO concentration had significant decline around 20% to 70% and 10% to 60% during DFL and DSL, respectively. The average CO concentration recorded at Wazirpur monitoring station was in the range of ‘Very Poor’ category during BFL and moved to ‘Poor’ category during DFL. Almost all the other stations had recorded average CO concentrations in the range of ‘Poor’ category and moved to ‘Moderate’ category during both DFL and DSL periods. Overall, the state of Delhi had highly been benefitted in their CO levels by 34% and 26.8% during DFL and DSL, respectively.

Analysis of CO in Uttar Pradesh. The decline in UP’s CO concentrations were around 5% to 50% and 10% to 55% during DFL and DSL, respectively. The average CO concentration recorded in Sector 1 Noida was in the range of ‘Very Poor’ category during BFL and ‘Poor’ category during DFL. Overall, the state of UP was highly benefitted in their CO levels by 22.8% and 39% during DFL and DSL, respectively.

Analysis of CO in Madhya Pradesh. The decline in MP's CO concentrations were around 15% to 65% and 30% to 55% during DFL and DSL, respectively. They moved from higher ranges of 'Moderate' category to lower ranges of 'Moderate' category during both DFL and DSL. Overall, the state of Madhya Pradesh was highly benefitted in their CO levels 42.7% and 28.8% during DFL and DSL, respectively.

Analysis of CO in Haryana. The CO concentration of Haryana hugely declined around 50% to 65% and 25% to 55% during DFL and DSL, respectively. The range of CO concentration levels moved from 'Poor' to 'Moderate' category during both DFL and DSL. Overall, the state of Haryana was highly benefitted their CO concentration levels by 58.5% and 41.5% during DFL and DSL, respectively.

Analysis of CO in Punjab. CO concentration in Punjab had reduced significantly around 45% to 55% and around 15% to 35% during DFL and DSL, respectively. The CO concentration levels moved from higher ranges of 'Moderate' to lower ranges of 'Moderate' category during DFL and DSL. Overall, the state of Punjab was highly benefitted around 33.3% and 9.5% during DFL and DSL, respectively.

Analysis of CO in Rajasthan. In the state of Rajasthan, 5 air monitoring stations helped in recording levels of CO concentration and the analysis had shown that the CO concentrations suspended in air had highly been reduced around 20% to 70% and around 5% to 45% during lockdown periods due to 1st and 2nd wave of COVID-19, respectively.

3.5 Change in Air Quality Index

The Lockdown strategies contributed to the improvement of air quality and it provides a chance to identify the source which affects the quality of air. Based on the guidelines of CPCB, the AQI was calculated. In Fig. 3, it shows the change in percentage of Air Quality Index for the six states during the first and second wave of COVID-19 pandemic period and also shows the percentage reduction for each monitoring station for the 4 phases BFL, DFL, AFL and DSL.

4 Discussion

4.1 Impact of Air Quality

Air Quality is the most important factor, it has been affected by the day-to-day activities that emerge from the primary and secondary pollutants present in the air. Pollutants like NO_2, SO_2 and CO are considered as the primary pollutant [15]. NO_2 pollutant is mainly occurred due to transportation. In developing countries, the urban areas are facing 90% of air pollution that are occurred due to vehicle emissions [19]. SO_2 primary pollutant is mainly released from the industrial, manufacturing yard

State	City	DFL	AFL	DSL
Delhi	Alipur	-0.22	0.25	-0.05
	Bagh	-0.21	0.21	-0.09
	Jahangirpuri	-0.23	0.26	-0.10
	Mundka	-0.26	0.28	-0.11
	Narela	-0.19	0.18	-0.06
	Pusa	-0.39	0.47	-0.10
	Rk Puram	-0.32	0.42	-0.03
	Satyawati College	-0.25	0.24	-0.09
	Sonia Vihar	-0.24	0.30	-0.07
	Wazirpur	-0.22	0.17	-0.09
Haryana	Faridabad	NA	0.14	-0.06
	Gobind Pura	-0.21	0.23	0.42160%
	Sector 11 Faridabad	NA	0.33	-0.02
	Sector 16a Faridabad	-0.30	0.36	-0.08
	Sector 30 Faridabad	NA	0.54	0.00
Madhya Pradesh	Bhopal	-0.33	0.27	-0.23
	Dewas	-0.24	0.09	-0.17
	Gwalior	NA	0.37	-0.32
	Indore	-0.42	0.49	-0.19
	Jabalpur	-0.34	0.27	-0.18
	Singrauli	-0.09	-0.04	0.08
	Ujjain	-0.31	0.24	-0.04
Punjab	Amritsar	-0.30	0.43	-0.01
	Jalandhar	-0.29	0.28	0.01
	Khanna	-0.30	0.41	0.00
	Mandi Gobindga	-0.23	0.31	-0.03
	Patiala	-0.35	0.57	-0.05
Rajasthan	Jaipur	-0.20	0.19	-0.05
	Jodhpur	-0.10	0.08	-0.01
	Kota	-0.28	0.36	0.08
	Moti Doongri	-0.42	0.70	-0.25
	Udaipur	-0.33	0.38	0.02
UP	Ardhali Bazar Varanasi	-0.25	0.08	0.05
	Central School Lucknov	-0.27	0.28	-0.09
	Ganga Nagar Meerut	-0.16	0.11	-0.07
	Gomti Nagar Lucknow	-0.20	0.18	-0.14
	Indirapuram Ghaziabad	-0.32	0.36	-0.09
	Jai Bhim Nagar Meerut	-1.00	NA	-0.12
	Lajpat Nagar Moradaba	-0.37	0.61	-0.13
	Loni Ghaziabad	-0.20	0.25	-0.08
	Pallavpuram	-0.12	0.05	-0.40
	Sanjay Nagar Ghaziaba	-0.23	0.25	-0.07
	sector 1 Noida	-0.22	0.18	-0.16
	Sector116 Noida	-0.35	0.45	-0.07
	Talkatora Lucknow	-0.15	0.17	-0.05

Fig. 3 Comparison Table describes the change in AQI percentage (%) of six states namely Delhi, Haryana, Madhya Pradesh, Punjab, Rajasthan and UP from 45 monitoring stations. (NA: Data Not Available, DFL- During First Wave Lockdown, AFL-After First Wave Lockdown, DSL-During Second Wave Lockdown).

and also by burning of coal in power plant. Secondary pollutant namely Ground Level/Tropospheric ozone is formed due to the photo-chemical reaction between the nitrogen oxide and volatile organic compounds (VOC) which has been indirectly added to the pure air and reduces the quality of air. Even Infection and Fatality caused by covid-19 is an exposure of air pollution especially by the pollutants NO_2 and $PM_{2.5}$ [20].

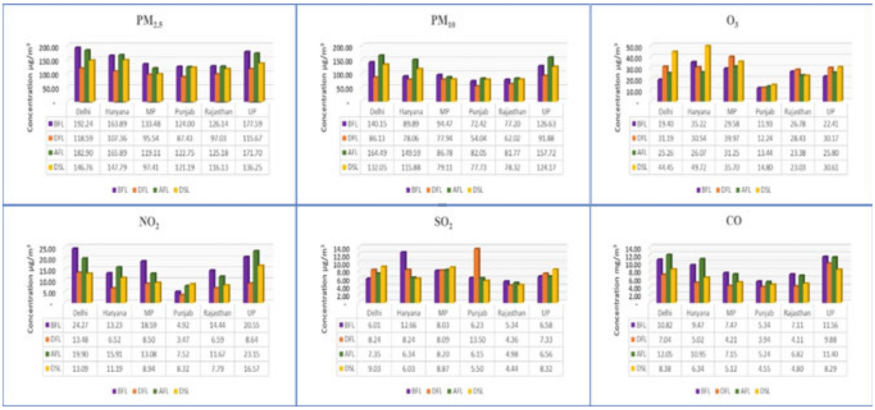


Fig. 4 Comparison chart describes the average pollutants emission during the four phases as BFL (Before First Wave Lockdown), DFL (During First Wave Lockdown), AFL (After First Wave Lockdown),
DSL (During Second Wave Lockdown)

4.2 Variation in Air Quality

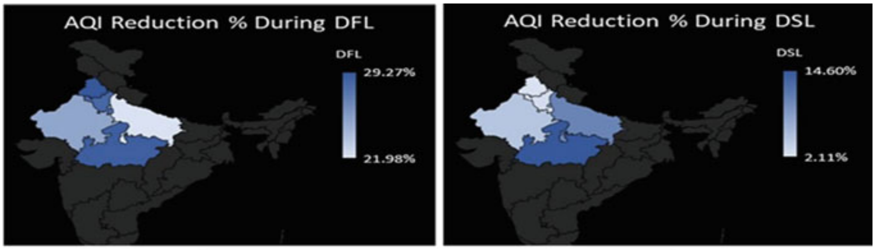
It is observed that there is a significant improvement in the concentration level of pollutants during the first and second waves of the COVID-19 pandemic. Figure 4 describes the average concentration level of pollutants like $PM_{2.5}$ and PM_{10}, O_3, NO_2, SO_2, CO. $PM_{2.5}$ emission rate was as high as 192.24 $\mu g/m^3$ and as low as 87.43 $\mu g/m^3$, PM_{10} emission rate was as high as 140.15 $\mu g/m^3$ and as low as 54.04 $\mu g/m^3$, NO_2 emission rate was as high as 24.27 $\mu g/m^3$ and as low as 3.47 $\mu g/m^3$ during BFL and DFL period, respectively.

4.3 Variation in AQI

The declined emission of pollutants like PM $_{2.5}$, PM $_{10}$, O_3, NO_2, SO_2, CO leads to the improvement of quality in the air. Figure 5 describes the overall change in percentage of reduction in Air Quality Index for the states namely Delhi, Haryana, Madhya Pradesh, Punjab, Rajasthan and UP. Figure 6 describes the analysis of AQI average during the 4 phases BFL, DFL, AFL and DSL.

5 Conclusion

From this analysis, it was observed that there was significant reduction in the pollutant's concentration level for $PM_{2.5}$ by approximately 42% and 21%, PM_{10} by 40%



State	DFL	DSL
Delhi	24.93%	7.65%
Haryana	28.37%	2.53%
Madhya Pradesh	28.89%	14.60%
Punjab	29.27%	2.11%
Rajasthan	25.11%	4.73%
UP	21.98%	10.38%

Fig. 5 The overall change in percentage of AQI during the lockdown period

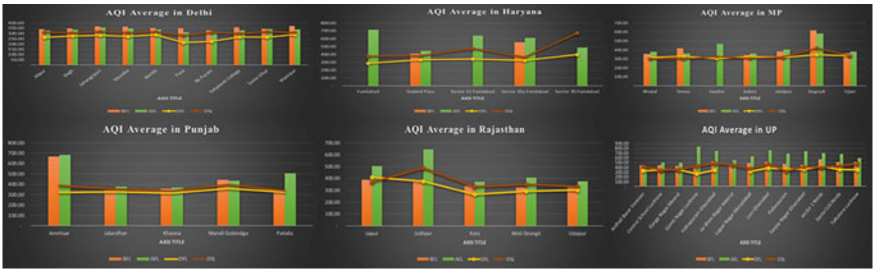


Fig. 6 Analysis of AQI Average for all the states during the 4 phases BFL, DFL, AFL and DSL

and 22%, NO_2 by 60% and 33%, CO by 59% and 42% during DFL and DSL, respectively. And also, there were significant improvement in AQI ~30% and ~14% during DFL and DSL. Also, the outcome shows the highly significant reduction in the concentration level of NO_2 compared to the other pollutants $PM_{2.5}$, PM_{10}, O_3, SO_2 and CO.

These analyses were done from the data sets taken from the stationary atmospheric air pollution monitoring stations which have limitations in the form of deploying a large infrastructure, choosing area, meticulous set of rules for operation, extreme costs for maintenance and standardization, sparsely located in urban areas. To provide a cost-efficient solution, potential IoT-based sensor devices could be deployed in various locations even in remote areas to monitor the pollutant's concentration level in the atmosphere accurately.

To monitor the hazardous air pollutants, the values retrieved from various IoT-based sensors could be stored in cloud platform to continuously monitor the ambient air quality. IoT-based sensors that would act as atmospheric air pollutant detection system would identify various sectors that cross the standard limits defined by CPCB based on WHO guidelines. The most important take-away of this analysis is that it identifies the sources of the atmospheric air pollution and introduces the efficient technology to monitor the air quality and to take necessary actions against the sectors that purposefully cross the ideal limits to cause the ambient air pollution that affects the human survival.

References

1. National Air Quality Status and Trends by Central Pollution Control Board 23-sep-2020: Online. Available: https://cpcb.nic.in/upload/NAAQS_2019.pdf
2. Fu F, Purvis-Roberts KL, Williams B (2020) Impact of the covid-19 pandemic lockdown on air pollution in 20 major cities around the world. Atmosphere 11(11):1189
3. World Health Organization (2018) State of global air: a special report on global exposure to air pollution and its disease Burden: Online Available: https://www.stateofglobalair.org/sites/default/files/soga-2018-report.pdf
4. Chatterjee P (2016) India takes steps to curb air pollution. World Health Organization. Bullet World Health Org 94(7):487
5. World Health Organization, Ambient air pollution: pollutants: Online. Available: https://www.who.int/airpollution/ambient/pollutants/en/
6. Krzyzanowski M, Kuna-Dibbert B, Schneider J (eds) (2005) Health effects of transport-related air pollution. WHO Regional Office Europe
7. World Health Organization (2010) WHO guidelines for indoor air quality: selected pollutants
8. World Health Organization Coronavirus disease 2019 (COVID-19) Situation Report – 94 online. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf
9. Numbers SI, Assessment WR (2020) Coronavirus disease 2019 (COVID-19). Americas 10(2):1
10. World Health Organization, Ambient (outdoor) air pollution online. Available: https://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health
11. Andrews MA, Areekal B, Rajesh KR , Krishnan J, Suryakala R, Krishnan B, Muraly CP, Santhosh PV (2020) First confirmed case of COVID-19 infection in India: a case report. Ind J Med Res 151(5):490
12. MHA (Ministry of Home Affairs), Govt. of India. Circulars For Covid-19 https://www.mha.gov.in/sites/default/files/PR_NationalLockdown_26032020_0.pdf
13. Nigam R, Pandya K, Luis AJ, Sengupta R, Kotha M (2021) Positive effects of COVID-19 lockdown on air quality of industrial cities (Ankleshwar and Vapi) of Western India. Sci Rep 11(1):1–12
14. MHA (Ministry of Home Affairs), Govt. of India. Circulars For Covid-19 https://www.mha.gov.in/sites/default/files/Guidelines_0.pdf
15. Wang H, Tan J, Li X (2021) Global NO_2 dynamics During the COVID-19 pandemic: a comparison between two waves of the coronavirus. IEEE J Select Top Appl Earth Observat Remote Sens 14:4310–4320
16. Website Reference: https://indianexpress.com/article/india/covid-19-second-wave-heres-a-list-of-states-that-have-imposed-lockdowns-7306634/
17. Mo Z, Huang J, Chen Z, Zhou B, Zhu K, Liu H, Mu Y, Zhang D, Wang S (2021) Cause analysis of PM 2.5 pollution during the COVID-19 lockdown in Nanning, China. Sci Rep 11(1):1–13
18. Balakumar S, Kavitha AR (2021) Quorum-based blockchain network with IPFS to improve data security in IoT network. Stud Informat Control 30(3):85–98. ISSN: 1220-1766. https://doi.org/10.24846/v30i3y202108.

19. Shrivastava M, Ghosh A, Bhattacharyya R, Singh SD (2018) Urban pollution in India. John Wiley & Sons, Ltd.: Chichester, UK, pp 341–356
20. Ali N, Islam F (2020) The effect of air pollution on COVID-19 infection and mortality—a review on recent evidence. Front Public Health 8. Article 580057. https://doi.org/10.3389/fpubh.2020.580057

Arduino UNO Controller and RTC-Based Medication Reminder and Monitoring System

M. Kasiselvanathan, J. Prasad, and G. Sekar

Abstract This study could result in a system that aids in the management and monitoring of medications. There's a potential that patients will make sure to take their medications at periodic intervals but fail to take the one that needs to be taken at that specific moment. The project focuses on those people who are taking medications; it reminds them to take medications on time. This system consists of an ARDUINO UNO microcontroller and a real-time circuit and a GSM. We are building this program to save time and provide a user-friendly system by utilising current technological breakthroughs. This process is powered by an embedded programme that accepts predetermined parameters and processes them using system parameters entered via a keypad or other interface device. It's equipped with a GSM network to warn both the patients and the physician at the appropriate time, making it even more cutting-edge. By using GSM, the system also sends SMS to the patient for the drugs reminder. The analysing logic is built into the embedded programme, which then triggers the alert via an auditory alarm. which not only has an alarm, but also has an LCD that shows the pills that need to be given at the time of the reminder. Using a nurse specifically for one patient is not a straightforward or easily accessible option in most clinics. To overcome such issues, we've devised a project that may notify patients of their medication intake at periodic times.

Keywords GSM · Patient reminder · Arduino UNO · LCD · Voice alarm · SMS alert

---

M. Kasiselvanathan (✉)
Department of Electronics and Communication Engineering, Sri Ramakrishna Engineering College, Coimbatore, India
e-mail: kasiselvanathan@gmail.com

J. Prasad
Department of Electronics and Communication Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

G. Sekar
Department of Electronics and Communication Engineering, Sri Ramakrishna Institute of Technology, Coimbatore, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_17

1 Introduction

It is a 24/7 healthcare and medication Reminder Monitoring System. In Taiwan, the population over 65 exceeded 12.51% of the entire population in 2015. Now that population ageing is a worldwide phenomenon, hence the need to create knowledge for how to improve the wellness and well-being of the elderly. We can't always put our bodies at risk by not taking medications at the correct time, delayed or skipping doses, or indeed picking the wrong dosage [1]. Due to less immunity and improper diet a person grows old he gets affected by a lot of diseases. Elderly people fail to take their prescribed medication at the proper time because they often forget due to aging, which can lead to a decrease in their health drastically, causing a further increase in their illness [2]. In most of hospitals, employing a nurse to a single patient is not easy and available. The amount of medications that each individual takes has grown rapidly, making it difficult for us to consume all at the specified times [3]. For this purpose, there needs to be some facility for us which monitors patients and takes care. Nowadays we are all living technology-based life. We can use this technology in a way that will be beneficial for us [4]. This system deals with time taken for patients to take their pill on a particular time. The schedule for medication is framed as per the patient's requirement and this schedule can be changed if it is necessary [5]. Using GSM, which is connected to a cellular telephone or indeed any handheld device, we can remember to take our medicine at a specified time. The soul gadget is a portable device [6]. The latest fad in medicine has continued with electronic drugs, and the employment of new technologies in the medical difficulties has been evaluated and addressed [7]. When the time arrives to take medicine it automatically displays in the LCD screen, along with Led blinking on the side [8]. When it's time to take medicine, an automated medication recall is sent to the patient's cell phone. The study suggested for a medication reminder and monitoring system that uses GSM and includes monitoring and warning setting, as well as reminding the user to take medicine three times per day and without having to replenish the box every week [9].

2 Block Diagram

The flow chart of our medication reminder is shown in Fig. 1. The Arduino UNO is the main controller in this project. RTC is used to determine when medications will be delivered to patients. Voice Board is used to produce the sound whilst it reaches the medicine time. With the help of a voice board, we are able to make alert voice commands like "medicine time" or whatever else we want to say. GSM is used for alert messages if the medicine is not given to the patient on time. The message was sent to the corresponding staff of the patient. These details are continuously monitored in the LCD display.

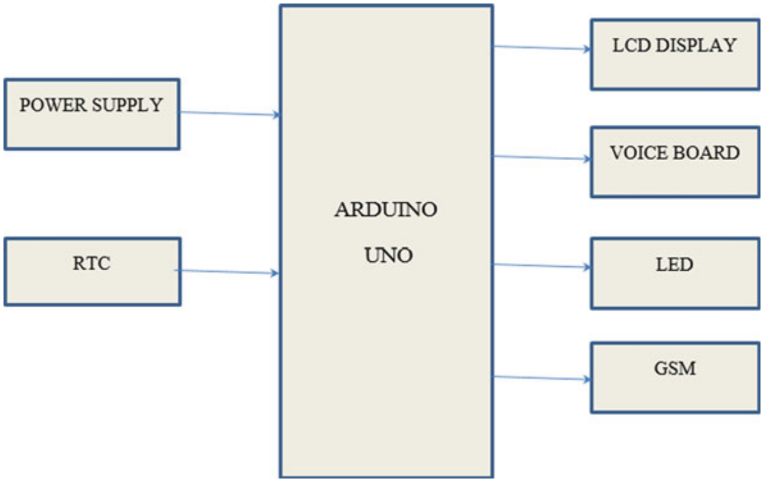


Fig. 1 Block diagram of medication reminder

2.1 Arduino UNO

Figure 2 shows the Arduino UNO microcontroller. The ATmega328P microprocessor is used in the Micro Controller board. Six analogue inputs, a 16 MHz quartz, a Universal serial bus, an influence jack, an ICSP header, and a push button are amongst the 14 digital i/p and o/p pins (six of which are usually used as PWM outputs). It includes everything you 'll need to begin started with the mcu, including a USB cable to connect it to a computer and an AC-to-DC converter or batteries to charge it. In Italian, the term "UNO" signifies "1," and it was chosen to mark the beginning of Arduino Programming (IDE) 1.0. The Uno panel and ARDUINO Software (IDE) version 1.0 were the default versions of ARDUINO, that have presently expanded towards future versions.

Fig. 2 Arduino UNO ATmega328P microcontroller

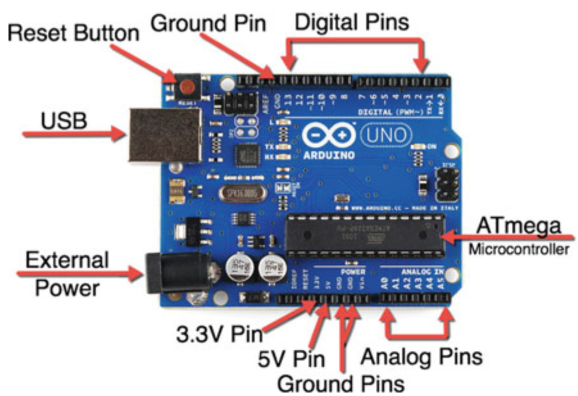


Fig. 3 GSM Modem

2.2 GSM SIM800A

Figure 3 shows the GSM Modem. GSM stands for GSM technology and can be a wireless transmission device (GSM). Bell Laboratories started the concept for GSM in 1970. It represents the more widely used cellular communication system in the planet. GSM (Global System for Mobile Communications) would be an independent and digitised mobile platform that contains the 850, 900, 1800, and 1900 MHz frequency bands to provide mobile connectivity for sending data. The GSM method was designed as a computer circuit using the time division multiple access (TDMA) technology for data transmission. A GSM converts and compression data prior delivering it through a network with two client data streams, within every slots. Data rates range from 64 kbps to 120 Mbps can be managed by the digital system.

2.3 RTC DS1307

The RTC Pin Configuration is shown in Fig. 4. As the names suggest, real-time clocks (RTC) are time components. The DS1307 real-time clock (RTC) IC is an 8-pin module with an I2C interface. The DS1307 could have been a low-power clock/calendar with a 56-byte SRAM backup batteries. Secs, mins, hrs, days, dates, months, and years are displayed just on timer. The much more important advantage of RTC is that it mandates the installation of a battery backup, which maintains the clock/calendar running whilst the power goes off.

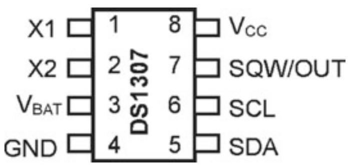


Fig. 4 Pin configuration of RTC

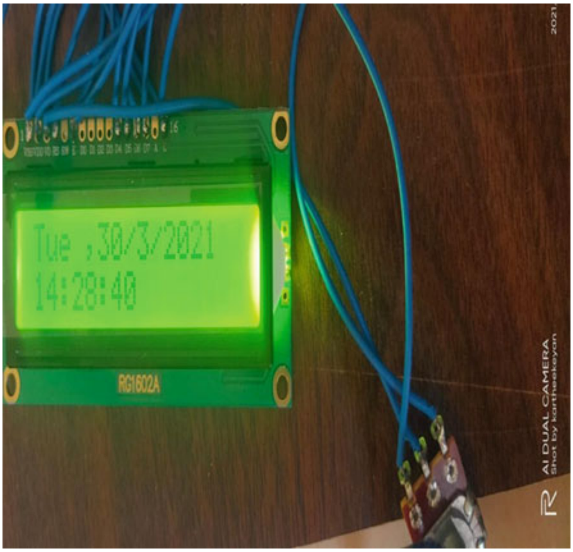


Fig. 5 LCD Display

3 Results and Discussion

Figure 5 shows the LCD display of Day, Date, Month, Year, and Time. With the assistance of GSM, we will receive messages as shown in Fig. 6. It's a notification to require our medication on Time, and also, we used a voice board through which when the time involves take medicine It says "Medicine Time" through Speaker.

4 Conclusion

With basic electronics techniques, a low-cost, functional model for an automatic pill reminder was created in this research. A bell and an LCD display have been fitted for detection purposes and alarm because then the person in question takes his pill on correctly and in a timely amount without personalised monitoring. The device also keeps track of when and where you take your pills, which can be beneficial for future medical consultations. If medicines are not given on time, relatives are notified. This simple method can be a useful choice for families with work-hour impulses or who are obligated to maintain a mistress for a family member who is suffering from medical issues.

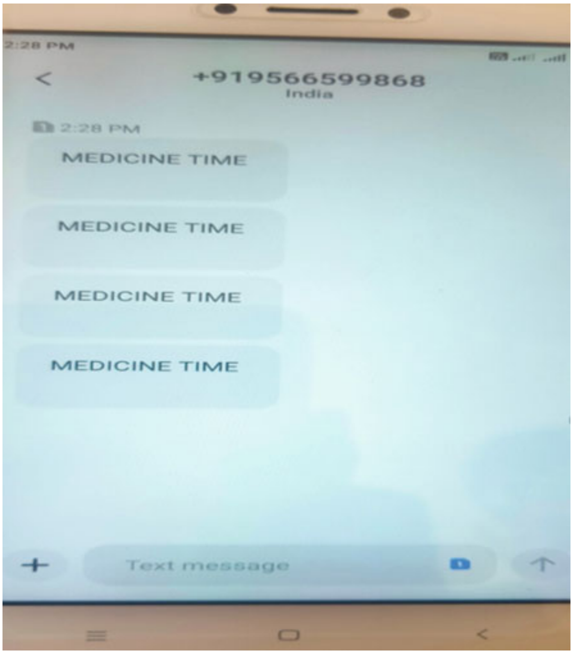


Fig. 6 Mobile phone SMS

References

1. Billingsley L, Carruth A (2015) Set of technology to promote effective medication adherence. J. Cont. Educ. Nurs. 46(8):340–342
2. Kale S, Bhadane A, Pawar P, Achaliya PN (2017) An android based medicine reminder system based on OCR using ANN. Int. J. Comput. Appl. 5:1–5
3. Jayashri I, Zope PH (2018) Aadhar enabled ration distribution and monitoring using smart card. Int J Sci Spiritual Bus Tech 6(1):73–78
4. Evgeny S, Ilya P, Ivan T (2010) Mobile phone sensors in health applications. In: Proceeding of the 12th conference of fruct association, pp 136–141
5. Wissam A, Ali A, Suleiman A, Abdallah K, Mustapha H, Chady E (2018) Smart Medicine Dispenser (SMD). In: IEEE 4th Middle East Conference on Biomedical Engineering (MECBME)
6. Shaw, Thomas J (1997) Automatic pill dispensing apparatus. U.S. Patent No. 5,609, 268.11
7. Fang, Kerry Y, Anthony JM, Heidi B (2016) Current trends in electronic medication reminders for self care. In: The promise of new technologies in an age of new health challenges: selected papers from 5th global telehealth conference, New Zealand, pp 1–2
8. Patel S et al (2013) Mobilizing your medications: an automated medication reminder application for mobile phones and hypertension medication adherence in a high-risk urban population. J Diabetes Sci Tech 630–639
9. Mahtani KR, Heneghan CJ, Glasziou PP, Perera R (2011) Reminder packaging for improving adherence to self-administered long-termmedications. Cochrane Database Syst Rev 7(9):1–5

Corrosion Detection and Prediction Approach Using IoT and Machine Learning Techniques

Vaibhav A. Parjane and Mohit Gangwar

Abstract Underwater pipelines widely used to supply the oil and gases by the entire world; in recent developments, various countries are using underwater pipelines and aquatic transportation. In the oil & gas companies, corrosion around asbestos is a serious issue. It is not only the natural gas, power and processing sectors. That has impacted, but it can also occur anywhere air conditioning built on pipelines. On steel pipelines that have fitted with good conductivity, this oxidation happens. It is therefore quite challenging to detect since the existence of corrosion hidden by the inner insulation. Numerous machine learning and IoT systems have done a few research on that. Still, it generates a high error rate due to some intangible parameters not considered by those systems. The speed of substance, temperature, Ph values, and pipe thickness are the most influenceable parameters for generating corrosion. In this system, we proposed corrosion detection and prevention using IoT and machine learning. In the first stage, the system deals with the IoT environment, which generates event data like Ph values, Temperature, Speed, Thickness, etc. That parameter has extracted every six hours. Various machine learning and deep learning has used to evaluate the proposed system. The experiment analysis has done around 100 days of data to identify the system's performance evaluation.

Keywords IoT · Deep Learning algorithms · Recurrent Neural Network · Corrosion detection · Prediction system · Classification

1 Introduction

The mares restrain unknown beasts and enormous energy reserves, performing an imperative role in preserving survival on earth. From the twentieth century, marine

V. A. Parjane
Department of Computer Science and Engineering, Bhabha University, Bhopal, Madhya Pradesh, India

M. Gangwar (✉)
Department of IT, IET, Dr. Rammanohar Lohia Avadh University, Ayodhya, Uttar Pradesh, India
e-mail: mohitgangwar@rmlau.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_18

investigation comprehensive has been actively interested in high-tech enterprises. Vision technology has brought critical awareness for its ability to carry high knowledge frequency. Researchers aim to achieve high-quality undersea data for a variation of sunken statements, including robotics, salvation sorties, human-made constructions investigation, environmental monitoring, sea plants pursuing, and real-time exploration. Nevertheless, the quality of submarine thoughts incredibly impressed by the appropriate ecological and synthetic characteristics of underwater situations, supporting arguments that more clearly succeeded in temporal imaging. As a result, we require deep learning-based detection and testing systems that warn us to water's early access to surface images and insulation. To create rust estimations and improve viewpoints, deep learning is used. Create various corrosion image characteristics of the pipeline under various tests and estimations—the execution of trains and testing with variable fold assumptions—which the utility does not count. Internal pipeline oxidation causes outside sunlight, ending in corrosion attack. Homogeneous rust, cracking, fracture rust, and microbiological rust are all caused by thin inner walls. Insulation repair is the most prevalent issue (ICC) in the oil and gas business, for example. The intended research study will build and construct a system for identifying and forecasting corrosion on the underwater pipeline picture data set using a deep learning algorithm. The system classifies methods and predicts the likelihood of detecting corrosion using various deep learning algorithms.

2 Literature Survey

The regression coefficients used for VNA validation theoretically randomize testing data for computer learning following [1]. The density of duplex adhesives used for the deterioration safety of carbon fibre substances is exhibited positive by the automated method after learning. Without any need for repetitive reconfiguration, the precision in terms of standard variability compared to the full scale is greater than 13%. The approach is simple and well-suited for automation use. Furthermore, for mechanical testing, crack identification, and surface area calculations, the process evaluated. A new approach that can identify corrosion using object identification has defined in [2]. The device proposes a detection model to carry out such IoT-based systems relocation and its handful detection. YOLO is the method of identification of state-of-the-art artefacts that centred on computer vision. YOLO adopts controlled methods of learning. It takes the labelled dataset as an input. So on the insulator dataset of 2000 objects, the device trained YOLOv3 with appropriate precautions for two groups, the outer layer and the defect. These images use multiple convolutional to remove features. For the performance of posterior probability and mark projections, regression analysis then used. YOLO's superiority over other image recognition methods based on speed; therefore, it is fast and can handle 45 frames a second.

A novel identification mechanism presented via an effective way of subtracting objects [3]. This technique, in particular, involves four steps. The first move is to use cameras and GPS systems mounted in the investigation train to capture the speedway infrastructure and their associated mileage along the same route multiple times. Then the GPS data is again accomplished by adding an enhanced SVM classification algorithm. Second, the device identifies one passing of the object information recorded along with Known coordinates as a reference. The first frame alignment process operated via GPS and visual features for the remaining passes that need to identify. The SURF system applied for image matching in the fourth part. Then the simple arithmetic operation is performed around each matching pair of images. Finally, the system uses analytical filtering methods to exclude false positives found by the system.

According to [4], the building corrosion picture of the power transmission validates the suspicious detection process. This system proposes an adaptive approach incorporating canny regulator and Hough transformation, aiming at the lack of a significant amount of unwanted noise generated by the conventional Otsu's optimal thresholding packaging material. The advanced method can help extract the edge and the irregular portion of the target imager compared to traditional methods, filters out unnecessary ambient sound points, and achieves efficient tower deterioration detection. In [5], this device based on processor statistical measure. The core of both the upper computer software program is constructing a series—intelligent detection system of a grounding grid degradation state. Firstly, the technology relies on the standard of length, with the Mobile software system's centre. It has activities such as data collection, data saving, and statistical analysis at about the same time. It can calculate the power network's electromagnetic parameters accurately and effectively and determine the corrosion rate and the depth of corrosion. Secondly, this device's hardware sorting unit and computer processing algorithm resolve the issue that the previous machinery cannot get reliable, grounded grid data due to the magnetic contamination.

A working prototype and technology for CUI identification using low-cost equipment will include in [6]. Detecting CUIs, 600–1000 Hz frequency domain signals produced by a sinusoidal waveform power supply were applied. The route optimization program showed the results for CUI. The experiments showed that the photocopied core probe take coil can easily prevent those CUIs. The amplitude and frequency eddy current device developed with the suggested CUI spam detection can track and detect the CUIs with a wall temperature of up to 3 mm.

According to [7] system is effective in assessing the surface humidity conditions of pavement drainage pipes via an electrical resistivity dependent sensor environment. The device configured to perform measured data inside a residential sewer pipe in Sydney City, Australia. The comment analysis revealed the survival under aggressive sewer environments of the sensing device. It illustrated its suitability for lengthy surveillance inside sewage pipes—a data analysis approach designed for outlier detection in addition to sensor growth. The model integrates a prediction model for anomaly detection using a periodic autoregressive (methodology). The visual data, the reliability evaluated, and the result showed its successful efficiency. Overall, the proposed sensor suite can impact the way energy providers control the primary water corrosion.

An innovative semi test method leading to low induction of eddy-current was built in [8] and evaluated on a microstructural system, including an insulating gap. The methodology allows calculations of it through a diameter between 4 and 16 mm, with 1 mm estimation errors. The method is quantitative, semi and enables the transmitter coil to locate outside of the reinforced steel framework. It is helpful when there is no connection to the exterior of automotive containers or pipelines. The device allows the detection over time of surface roughness. It offers a quantitative method for the integrity assessment of concrete sectors and recognizing commercial containers or pipelines preliminary stage corrosion. The process based on the Faculty of Medicine of a bespoke dual-coil device modelled and installed.

In system [9], a process of Classification and tracking of deep neural network-based welded joints:

1. Higher operation data are drawn for weld beads with many samples, together with image analysis and Generative Adversarial Network (GAN).
2. The test dataset updating framework is developed to ensure that the deep neural network model can cover all samples.
3. The convolutional neural network understands the detection and recognition of material layers, preventing the handcrafted features of traditional machine learning methods.

The basis of model training is an adequate and cost-effective sample selection method of welding beads. Tiny amounts of weld beads, the computer vision and GAN network used to expand the data collection realize the back-propagation algorithm.

The Grey Wolf algorithm in [10] supports the vector machine to maximize the least squares. Simultaneously, to correct the forecast data, an error correction model is implemented, thus increasing the predictive performance and checking the sampling frequency's quality. The findings demonstrate that the combined generalized least back-propagation proposed methodology enhanced by the grey wolf algorithms. The error detection used to estimate the ground network corrosion rate, which is stronger than the least square support vector. It enhanced by the system of the least square classification algorithm or the Grey Wolf method. The computer model validation loss is smaller. The accuracy is higher, and the rate of grounding grid corrosion prediction is more accurate. The Grey Wolf Optimizer influenced by the predation behaviour of the grey wolf. It optimizes it by simulating the grey wolf's centralized inequality, tracking, engulfing, and assaulting prey in nature.

The remarkable capacity to identify and monitor problems beneath valve coating owing to the intrusion of water that might ultimately dissolve that pipeline is demonstrated in [11]. The antenna is made up of two identically cross-polarized printed monopole dipole antennas. The writer antenna is placed directly above the pipeline to remotely gather the tag ID's resonances in real-time for early identification and monitoring of potential pipeline moisture damage. The proposed approach provides a low-cost, real-time solution. The sensing device, as well as the delivering antennas of both sensors that extend beyond the pipe's surface and the reader antennas, make up the device. This method enables for the prediction of out-of-sight coating flaws in real time. A sensor array of scattered spiral resonators was utilised, each creating its own spectral frequency signature.

Using complicated Bayesian networks, a Long Operational Life hybrid physics-based and information changing circumstances of structural systems assess the influence of numerous causes, similar to [12]. (DBNs). The hypothesis is supported by either empirical complex geometry, the framework constructed, and a variable model of auto-encoders for the thermal deterioration caused by a single component, therefore resolving an issue of insufficient evidence. RUL performance and expectations were formed as a result of incorporating these degrading performance indicators. The RUL value is determined using the output failure threshold and the transit time between the point of identification and the anticipated item of evidence. Monitoring technologies and expert knowledge may be included into the estimating model to update the RUL value as needed. Subsea pipes are being used in offshore oil and gas subsea production systems to demonstrate the suggested technology. Modeling degradation methods including fatigue, corrosion, sand erosion, and internal waves is done using DBNs. The DBN-based RUL approach is used to calculate the RUL.

In [13], ultrasonic detection of SCC was achieved utilising 60 bar pressured gas as an acoustic coupling medium. A custom test processor with 32 transmit channels was constructed to make circumferential scans of 425 mm 60 from the interior of pipe samples with diameters approaching 36 inches. The tests have to be carried out in a pressure tank housing both the test scanner and pipe samples due to the 60 bar hydrogen environment. Diverse pulses were produced, and signals were received at all scan places. The extracted parameters, such as spectral power and total length estimations, were measured. With processed variables and conserved image data, 2D plots revealed the capacity to identify and size genuine fractures, fracture fields, and other loose material.

Convolutional layered neural networks (CNN) are used to train effective algorithms from PEC knowledge in an efficient end-to-end model [14]. The framework, in particular, builds a general inter-model based on 1D CNN to estimate both the class and the depth of defects at the same time. Extensive testing has shown that this model is capable of performing both identification and correlation operations on PEC data. This designed approach achieves more accuracy and fewer mistake when compared to other comparable applications.

In an outdoors environmental setting in Chengdu. A galvanic ACM detector consisting of carbon metal anodes but stainless steel copper coils exposed for 34 days [15]. Due to their low amounts in Qingdao, contaminants such as SO_2, NO_2, O_3, CO, PM2.5 and PM10 were not the critical factors in the IACM. The impacts of ambient temperature, temperature and precipitation on the initial atmospheric corrosion established to be greater than that of the particles in the air, sulphur dioxide, hydrogen sulfide, hydrogen sulfide including ozone, and use a naïve Bayes (RF)-based machine learning method. The RF model showed greater accuracy in predicting immediate corrosion attack than the artificial neural network (ANN) and linear regression models. After considering the massive effect of rust formation on the detector, the accuracy level further is enhanced.

Method	Methodology	Strength	Gap analysis
Acoustic emission	Collect the IoT based signals to detection of leakage of pipelines	This system can provide early detection of corrosion before event generation	The noise data is hard to filtration and detect the actual objects from input data
Fibre optics sensing	Identification of leaks by using temperature variations based on optical cable quality according to the presence of leaking	The optical communication is impervious to electrical interference and may assist as both a sensor and a transmitting data intermediate	Very high implementation cost required for set up the hardware resource and monitoring 24*7
Vapour sampling	To detect trace amounts of certain hydrocarbon molecules, employ hydrocarbon vapor distributed into the sensor tube	Small amounts of dispersed gas may be detected using this device	The time it takings to identify a leak is identifiable, making underwater pipes ineffective
Infrared thermography	Infrared imaging methods for detecting temperature differences in the pipeline surroundings are used to identify leaks	Powerful for converting detected items into visual pictures, simple to use, and quick reaction time	It's difficult to measure leak perforations small than 1.0 mm utilizing IRT-based devices
Ground penetration radar signals	By dragging an antennae lengthways a material, electromagnetic radiation are forward to controller	Low time required for detection of leaks in large environment	In a clay soil context, GPR signals can be easily twisted, exclusive, and need a extremely accomplished operator
Fluorescence	The quantity of fluid released and the rate of light reflected at various signals are proportionate	High detection accuracy and low computation time for detection	The detection medium must be inherently brilliant
Electromechanical impedance	Use mechanical susceptibility variations discovered as a result of a pipeline failure	A piezoelectric material may function as both a sensor and also an actuator	It's exclusively for metal pipes, and it has operating restrictions in high-temperature situations
Acoustic emission	Collect the IoT based signals to detection of leakage of pipelines	This system can provide early detection of corrosion before event generation	The noide data is hard to filteration and detect the actual obejcts from input data

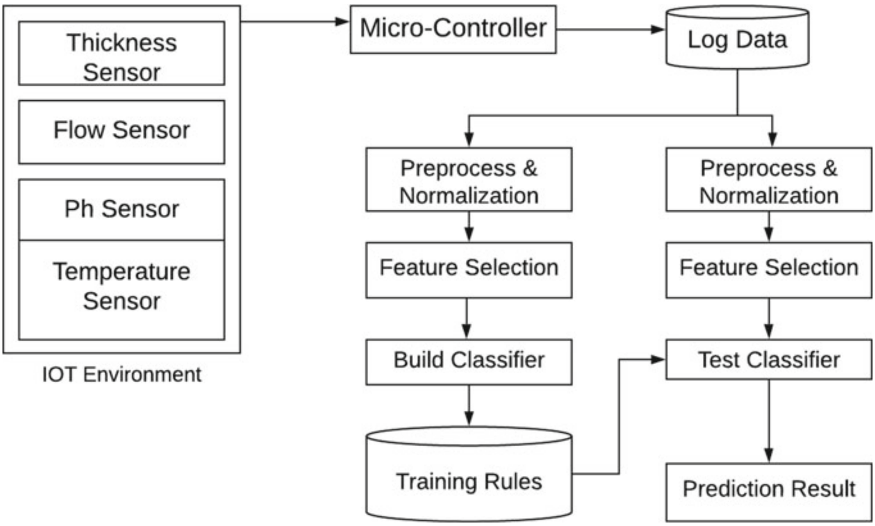


Fig. 1 Proposed system architecture

3 Materials and Methods

This system collects data from the IoT module that clusters various sensors and connected to the microcontroller. Our fundamental aim is to detect the extent of degradation in the pipelines. Different gases containing many impure molecules are available in the combustion process—corrosion caused by these adulterated contaminants and infrastructure vulnerability to toxic conditions. We have to know the core gas parameters in the pipeline, including temperature, applied load, and CO2 pressure or ph levels, to measure corrosion rate. We can ultimately measure the corrosion process in gas pipes and use these variables and place them to choose to follow the corrosion measurement method using ladder logic. The corrosion rate is detected, and the value is much higher than the volume range, the flow stopped as all valves are closed (Fig. 1).

The background knowledge is training rules generated by train classifies, which basically validate the test input data which generates from real time IoT dataset. According to the suggested algorithm, each event awards a reward or a consequence, and each event's weighted state changes as a result. As a result, the system produces BK rules throughout execution.

4 Algorithm Design

Input: Test DB List as test dataset, Train db List as training dataset and desired threshold as th

Output: Return highest probability with class label and weight
Step 1: Read data from the dataset using the below formula

$$test\_Feature(m) = \sum_{m=1}^{n}(.feature\_Set[A[i]\ldots\ldots..A[n] \leftarrow \text{TestDBLits})$$

Step 2: Extract the attribute values from each event of $testFeature$ (m) using the below function.

Extracted_Feature_Setx[t...–...n] = $\sum_{x=1}^{n}$(t) $\leftarrow$ $test\_Feature$ (m)

Extracted_Feature_Setx [t] having a data of respective domains

Step 3: Read all training dataset to generate the matrix

$$trainFeature(m) = \sum_{m=1}^{n}(.feature\_Set[A[i]\ldots\ldots..A[n] \leftarrow \text{TrainDBList})$$

Step 4: Extract the attribute values from each event of $trainFeature$ (m) using the below function.

Extracted_Feature_Set_y[t......n] = $\sum_{x=1}^{n}$(t) $\leftarrow$ $test\_Feature$ (m)

Extracted_Feature_Sety [t] having a data of respective domains.

Step 5: Now validate each feature with a testing set

$$weight = calcSim(\text{Feature\_Set\_x} \mid\mid \sum_{i=1}^{n}\text{Feature\_Sety}\left[\text{y}\right])$$

Step 6: Return as class lable with weighted score <object_id, weight_val>

5 Results and Discussion

This section we examine the whole execution of the proposed system in two separate open-source platforms for experiment analysis. To produce sensor nodes, the system first generates a software simulation environment. The whole log has been utilized as an IoT communication log created by numerous analogue sensors in simulation. The approaches provided in the foundation approach were used to compute the various parameters between sensors. The accuracy obtained with different thresholds is discussed in Table 1.

Table 1 demonstrates corrosion detection accuracy with various threshold values using Recurrent Neural Network. The different thresholds can impact detection accuracy, while 0.70 gives higher accuracy with a low error rate during the execution. In another experiment, we have evaluated our deep learning methodologies with the Recurrent Neural Network algorithm. Figure 2, demonstrates in detail the experiment.

Table 1 Performance evaluation with various threshold using RNN

Threshold	0.50	0.55	0.60	0.70
Accuracy	91.92	97.25	98.50	99.10
Precision	90.67	96.97	97.80	98.60
Recall	89.33	95.63	96.10	99.90
F-Score	91.99	94.29	95.11	98.50

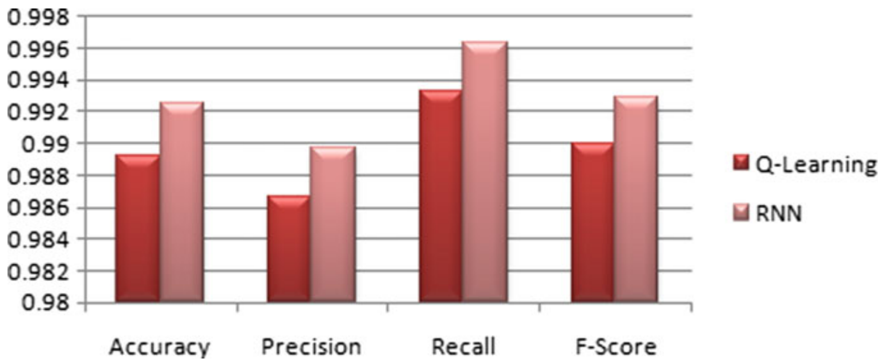


Fig. 2 Evaluation of performance analysis using RNN and Q-Learning

Figure 2 illustrates RNN provides better detection accuracy over Q-Learning based semi-supervised learning algorithm.

6 Observations

According to above literature we analyse numerous underwater image quality analysis and quality enhancement methods using various segmentation techniques. Below are the findings we identified after this investigation.

- Using soft computing methods, the system achieves 99 percent accuracy in a real-time IoT context.
- The major goal was to develop an underwater image processing approach that could be applied to pipeline networks, resulting in comparable faults in each file.
- Prior to corrosion estimate, picture restoration and enhancement were evaluated on both self-collected and publicly accessible degraded underwater image datasets, with encouraging results in both cases.

The corrosion issues raise concerns about the tubes' remaining • safe life. Some probabilistic techniques may be used to examine corrosion-related cracking and pitting processes, and others are being developed.

7 Conclusion

The system demonstrates that various sectors have started to improve the explosion of IoT products. Creating life form efficiency in different tasks while introducing individuals to high-risk conditions is achievable by exploiting Artificial Intelligence (AI) technologies with IoT. This system, using machine learning and Recurrent Neural Network. We presented a method for predicting oil drilling breaches in isolated oil and gas areas with minimal last-mile connectivity. Within the system of surface oil exploration, our solution framework required. Sub-sea pipelines needing cameras designed for underwater object activities are beyond our reach. In this poster's framework, reservoirs located beneath were also not protected. They would require any robust visual capturing of the piping section and an approach to function. To implement the system with large IoT dataset will deep learning algorithms for better detection of corrosion will be an interesting task in the future direction.

References

1. Husby K, Myrvoll TA, Knudsen OO (2019) Eddy Current duplex coating thickness Non-Destructive Evaluation augmented by VNA scattering parameter theory and Machine Learning. In: 2019 IEEE Sensors Applications Symposium (SAS), IEEE
2. Adou MW, Xu H, Chen G (2019) Insulator faults detection based on deep learning. In: 2019 IEEE 13th international conference on Anti-counterfeiting, Security, and Identification (ASID), IEEE
3. Gao L et al (2020) Anomaly detection of trackside equipment based on GPS and image matching. IEEE Access 8: 17346–17355
4. Ye L, Li W, He C (2020) Research on detection method of tower corrosion based on hough transform. In: 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), IEEE
5. Gao Y et al (2020) Design and implementation of intelligent detection equipment for corrosion status of grounding grid. In: 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), IEEE
6. Sodsai K, Noipitak M, Sae-Tang W (2019) Detection of corrosion under coated surface by Eddy current testing method. In: 2019 7th international Electrical Engineering Congress (iEECON), IEEE
7. Thiyagarajan K et al (2020) Robust sensor suite combined with predictive analytics enabled anomaly detection model for smart monitoring of concrete sewer pipe surface moisture conditions. IEEE Sensors J
8. Guilizzoni R, Finch G, Harmon S (2019) Subsurface corrosion detection in industrial steel structures. IEEE Magn Lett 10:1–5
9. Yang L et al (2020) Automatic detection and location of weld beads with deep convolutional neural networks. IEEE Trans Instrumentation Measurement
10. Hongbo S et al (2020) Corrosion rate prediction of grounding network based on improved least square support vector machine. In: 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), IEEE
11. Deif S, Daneshmand M (2019) Multi-resonant chipless RFID array system for coating defect detection and corrosion prediction. IEEE Trans Industrial Electronics

12. Cai B et al (2019) Remaining useful life estimation of structural systems under the influence of multiple causes: subsea pipelines as a case study. IEEE Trans Industrial Electronics 67(7): 5737–5747
13. Norli P et al (2019) Ultrasonic detection of stress corrosion cracks in pipe samples in a gaseous atmosphere. In: 2019 IEEE International Ultrasonics Symposium (IUS), IEEE
14. Fu X et al (2019) Towards end-to-end pulsed eddy current classification and regression with CNN. In: 2019 IEEE international instrumentation and Measurement Technology Conference (I2MTC), IEEE
15. Pei Z et al (2020) Towards understanding and prediction of atmospheric corrosion of a Fe/Cu corrosion sensor via machine learning. In: Corrosion science, 108697

Low-cost ECG and Heart Monitoring System Using Ubidot Platform

J. Lakshmi Prasanna, S. V. Aswin Kumer, M. Ravi Kumar, M. Sangeetha Lakshmi, P. Srilaxmi, and Chella Santhosh

Abstract The sector of health care is the most promising use of IoT. The application of IoT in the healthcare arena is investigated in this research, and a system is developed to monitor the patient's electrocardiographic (ECG). ESP32 Module, AD8232 Electrocardiographic (ECG) Module, Electrocardiographic (ECG) Electrodes, Electrocardiographic (ECG) Electrode Connector, Breadboard, and Connecting Wires make up this system. On the Ubidots dashboard or stage, there is an electrocardiogram. Ubidots, the most promising IoT application is in the field of health and wellness. The implementation of IoT in the healthcare arena is discussed in this research, and a framework for screening the Electrocardiographic (ECG) of the quiet is proposed.

Keywords Electrocardiographic (ECG) · ESP32 module · ECG module AD8232 · ECG electrodes · ECG electrode connector

J. Lakshmi Prasanna · S. V. Aswin Kumer (✉) · M. Ravi Kumar · M. Sangeetha Lakshmi · P. Srilaxmi · C. Santhosh
Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
e-mail: svaswin@gmail.com

J. Lakshmi Prasanna
e-mail: lakshmiprasannanewmail@kluniversity.in

M. Ravi Kumar
e-mail: ravikumar@kluniversity.in

M. Sangeetha Lakshmi
e-mail: 180040346@kluniversity.in

P. Srilaxmi
e-mail: 180040710@kluniversity.in

C. Santhosh
e-mail: csanthosh@kluniversity.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_19

1 Introduction

Till date you will monitor the ECG flag as it were once you are shown at that framework. So, taking into account the current scenario it would be a far better estimation generated by movement or situations with a longer cathode, it is designed to harvest, enhance, and channel tiny bio potential signals. An ultralow-power analog-to-digital converter (ADC) or an inserted microcontroller can simply secure the yield flag with this design. So, in this project, we will connect the AD8232 ECG Sensor to the ESP32 Module and keep an eye on the ECG flag on the Ubidots stage [1].

1.1 ESP 32 Module

The ESP32 is a low-cost, low-power system-on-chip microcontroller series with built-in Wi-Fi and dual-mode Bluetooth. Espressif Systems, a Shanghai-based Chinese company, designed and manufactured the ESP32, which was manufactured by TSMC using their 40 nm technology. It could be the ESP8266 microcontroller's successor. ESP32 can be used as an independent device. High-speed caches allow ESP32 to access the external QSPI streak and SRAM. The CPU code space can be memory-mapped with up to 16 M bytes of external streak, allowing 8, 16, and 32-bit operations [2]. The sample module is shown in Fig. 1 and pin description is shown for the same in Fig. 2.

Fig. 1 ESP32 module

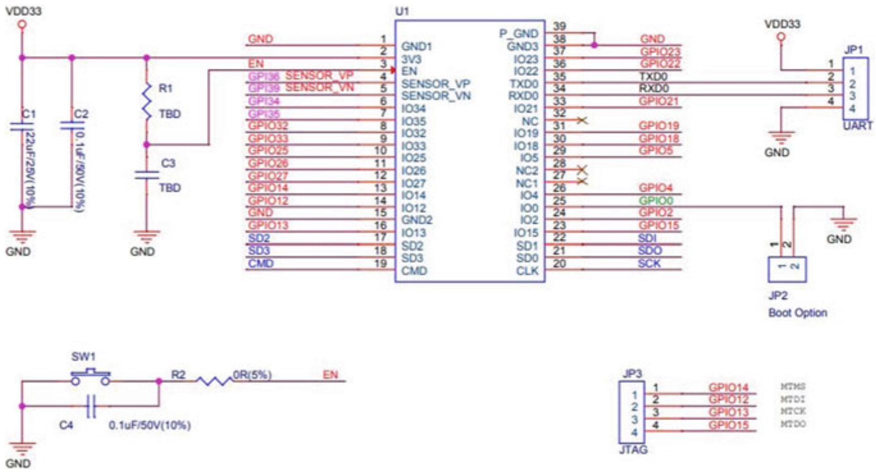


Fig. 2 Pin description of ESP32 module

1.2 ECG Module (AD8232)

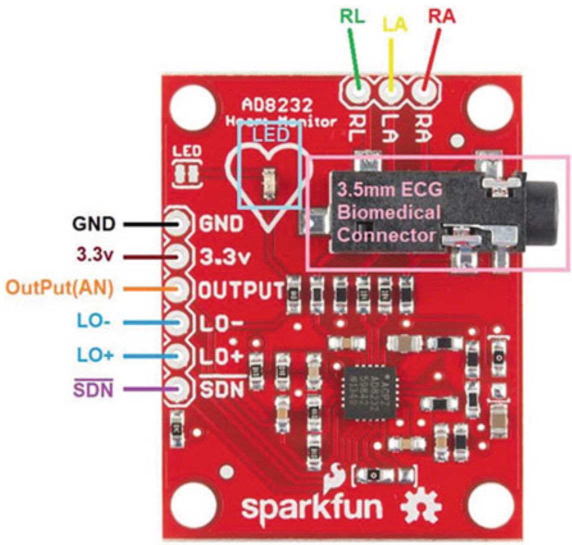
The AD8232 could be used as a configured hail conditioning element. It's designed to eliminate, increase, and channel tiny bio potential signals in the presence of agitating conditions, such as those caused by development or a blocked off anode arrangement. For apportioning with development artefacts and the terminal half-cell potential, the AD8232 may execute a two-pole high-pass channel. This channel is tightly associated with the enhancer's instrument plan to allow both broad choices up and high-pass filtering in a single organizer, saving both space and money [3].

To reduce commotion, the AD8232 uses an uncommitted operational intensifier to create a three-pole low-pass channel. To fit a variety of applications, the client can set the repeat cut off all channels. The AD8232 is part of a rapid re-establish task that reduces the term of something else's long settling high-pass filter tails. The AD8232 generally switches to the next channel cut off after an unforeseen hail modification that rails the enhancer (such as a leadoff condition). This feature allows the AD8232 to recover rapidly, necessitating significant estimations some time lately after tampering with the terminals to the subject. Figure 3 shows the pin description of AD8232 [4].

1.3 ECG Electrodes

Anodes (small, skin-colored plastic patches) are placed on the chest, arms, and legs in precise places. An ECG machine is connected to the anodes via lead cables. The

Fig. 3 Pin description of ECG module AD8232



heart's electrical improvement is measured, evaluated, and printed out at that time. There is no command transmitted to the body.

There are three types of ECGs: While lying down in a comfortable position, a resting ECG is conducted. An exercise bike or treadmill is used to perform an amplify or work out ECG. The cathodes are attached to a small supporting machine worn at your midsection during a walking ECG (also known as a Holter screen), and your heart can be monitored at home for 1 or more days [5] (Fig. 4).

2 Literature Survey

This inquiry about work proposes a toll convenient ECG remote framework and has extraction and clutter location calculation. The framework plan comprises a transportable ECG flag generator circuit, an information exchange gadget, and a sensible gadget. The plausibility of any heart infection utilizing this framework can effortlessly be checked by somebody. The points of interest of this method can be valuable some time recently, amid, and after a systole for genuine time checking of a quiet at anyplace. It might moreover diminish passing much appreciated to assault and other cardiovascular illnesses and more particularly giving wellbeing benefit by specialized specialists, to provincial ranges. This proposed inquiry is more useful for wellbeing security with moo taken a toll.

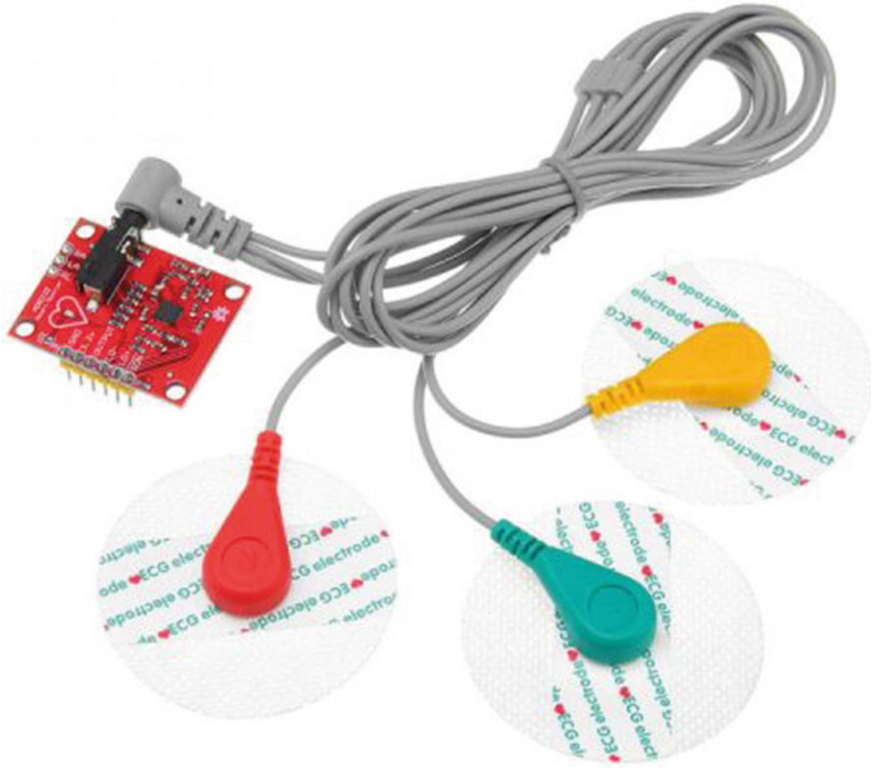


Fig. 4 ECG electrodes

3 Design and Analysis of Proposed Work

This proposed system utilizes a live track program to control and track the person's pulse. The proposed contraption is included inside the individual. It comprises ESP32 Module, Electrocardiographic (ECG) Module AD8232, Electrocardiographic (ECG) Cathodes, Electrocardiographic (ECG) Anode Connector, Breadboard, Interfacing Wires. The AD8232 could be a sensor utilized to determine the action of a heart. Where bio potentials are created by volume conduction of streams given by electrogenic cell collections. Amid the nearness of loud conditions, such as those caused by movement or farther terminal arrangement, it is outlined to extricate, increase, and transmit little bio potential signals. This permits the ultra low power digitizer to provide efficient data to the microcontroller [6]. Figure 5 shows the result obtained from the proposed methodology.

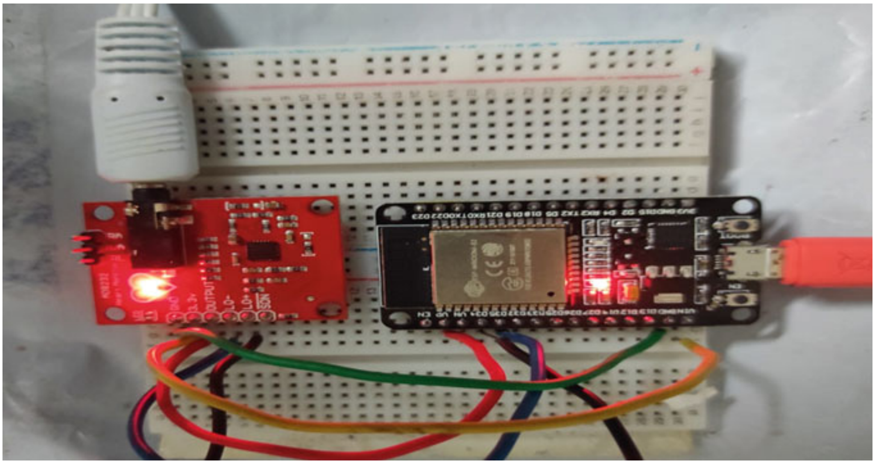


Fig. 5 Hardware implementation of the proposed system

4 Libraries Used

4.1 WiFi.h

This library permits an Arduino board to connect to the internet. It can work either a server tolerating approaching associations or a client making active ones. WEP and WPA2 Individual encryption are given by the library, but not WPA2 Venture. To note, on the off chance that the SSID isn't broadcast, the shield cannot connect [7].

4.2 WiFiUDP.h

The WiFiUDP course underpins the sending and getting of STA convention multicast parcels. Supplant UDP when sending a multicast parcel. StartPacket (addr, harbour), udp. Udp ought to be utilized. DestinationIP() to appear in the event that the obtained bundle has been sent to a multicast or unicast address [7].

4.3 PubSubClient.h

MQTT may be a convention for lightweight communications, appropriate for versatile gadgets. This library makes a difference for you to send MQTT messages and get them. It bolsters all consistent equipment for the Arduino Ethernet Client [8].

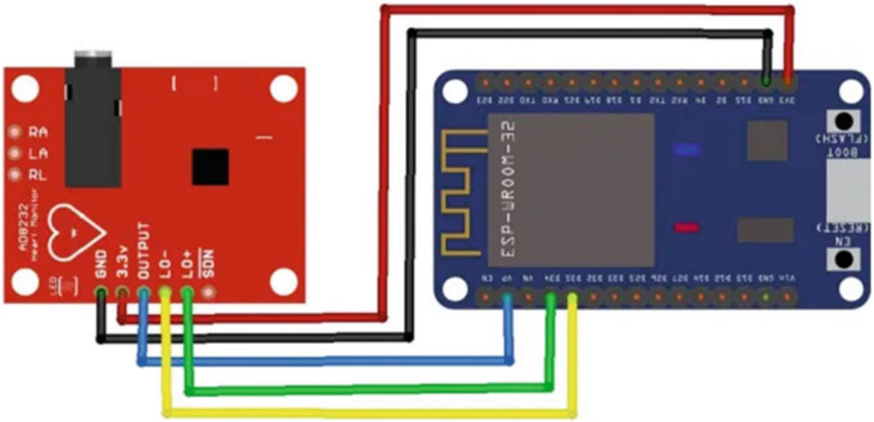


Fig. 6 Circuit diagram of proposed methodology

4.4 NTPClient.h

The Arrange Time Convention (NTP) is an application for clients/servers. In order to synchronize the clock to the organized time server, a workstation, modem, or server must be fitted with an NTP client program. Client programs as of now exist within the working framework of each computer in most occurrences [9].

5 Circuit Diagram

The AD8232 is a coordinate flag conditioning square. It is intended to isolate, enhance, and channel small bio potential signals in the presence of noisy environments, such as movement or inaccessible anode settings. This design makes it simple to acquire the yield flag using an ultralow-power analog-to-digital converter (ADC) or an added microcontroller. To that end, we'll connect the AD8232 ECG Sensor to the ESP32 Module and keep an eye on the ECG flag on the Ubidots stage [6–9]. The circuit diagram for the proposed system is shown in Fig. 6.

6 Simulation Results

The ECG sensor data gathered from the sensor with timestamp is displayed in the serial monitor of the Arduino IDE which is shown in Fig. 7. The same data will be posted in the UBIDOTS cloud platform along with the graphical representation is shown in Figs. 8 and 9.

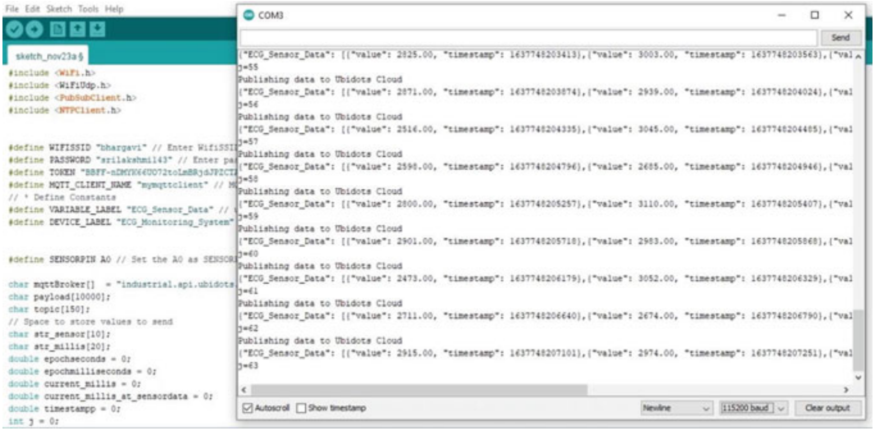


Fig. 7 Result in serial monitor

DATE	VALUE	CONTEXT	ACTIONS
New data available			
2021-11-24 15:33:56 +05:30	3201.00	{}	
2021-11-24 15:33:56 +05:30	2997.00	{}	
2021-11-24 15:33:56 +05:30	3135.00	{}	
2021-11-24 15:33:56 +05:30	3095.00	{}	
2021-11-24 15:33:56 +05:30	3125.00	{}	
2021-11-24 15:33:55 +05:30	3327.00	{}	
2021-11-24 15:33:55 +05:30	3371.00	{}	
2021-11-24 15:33:55 +05:30	2846.00	{}	
2021-11-24 15:33:54 +05:30	1689.00	{}	

Fig. 8 Data representation

7 Conclusion

After effective transfer of the open serial show, we can see our ESP32 effectively connected to the modified wifi and relegated to the IP address on the serial screen. MQTT reference to Ubidots Cloud was at that point started and connected to it. At that point, in conjunction with a timestamp, it began distributing ECG data. A vector stamp and 4 Adjust time-stamp information focuses are utilized with any information packet. Go and get to it once more at Ubidots. You'll see the live Electrocardiogram on the dashboard of Ubidots. Ubidots get information from our module ESP32 and outline the vector on the chart with the given Age Timestamp. Since we are sending

Fig. 9 Graphical representation of data

information with Timestamp, due to server inactivity on the chart, there will be no stop. This makes a difference to plot a correct electrocardiogram for Ubidots.

References

1. Nouira I, Said MH (2020) Smart ECG monitoring through IoT. In: Smart medical data sensing and IoT systems design in healthcare. IGI Global, pp 224–246
2. Wai, K, Aung NP, LwinHtay L (2019) Internet of Things (IoT) based healthcare monitoring system using NodeMCU and Arduino UNO. Publ Int J Trend Sci Res Dev (ijtsrd) 3(5):755–759
3. Rahman MA et al (2021) Remote monitoring of heart rate and ECG signal using ESP32. In: 2021 4th international conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), IEEE
4. Rahman MA, Li Y, Nabeed T, Rahman MT (2021) Remote monitoring of heart rate and ECG signal using ESP32. In: 2021 4th international conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), IEEE, pp 604–610
5. Joseph S, Ferlin Deva Shahila D, Patnaik S (2019) IOT based remote heartbeat monitoring. In: 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), IEEE, pp 1–5
6. Cabra J, Castro D, Colorado J, Mendez D, Trujillo L (2017) An IoT approach for wireless sensor networks applied to e-health environmental monitoring. In: 2017 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), IEEE, pp 578–583
7. Manoj G, Divya PS, Raj Barath S, Justin Santhiyagu I (2021) Mathematical modeling of IOT-based health monitoring system. In the Internet of Medical Things. Springer, Cham, pp 211–233
8. Rahma MM, Salman AD (2021) A wearable medical monitoring and alert system of COVID-19 patients. Iraqi J Comput Informatics 47(1):12–17
9. Atta RM (2021) Cost-effective vital signs monitoring system for COVID-19 patients in smart hospitals. In: Health and technology, pp 1–15

Machine Learning

Comparative Performance Evaluation of TFC-SVM Approach for Regression Test Case Prioritization

Omdev Dahiya, Kamna Solanki, Rahul Rishi, Sandeep Dalal, Amita Dhankhar, and Janpreet Singh

Abstract Software testing is carried out for uncovering the errors in the software through different testing methods. Regression testing is one of the testing techniques that chooses the test cases, emphasizing the adapted section of the software. Test Case Prioritization (TCP), a technique for regression testing, includes organizing test cases in a sequence, which improves the efficiency in attaining desired objectives. One of the objectives is to enhance the fault detection rate. Test cases must be executed in such a way that enhances the probability of fault detection and recognizes faults in minimum time in the testing process. In this research work, a comparative performance evaluation of the TFC-SVM technique for regression test case prioritization is performed with another existing cuckoo search and firefly algorithm. The simulation outcomes indicate that the TFC-SVM method performs better in APFD with 90.032% and other optimization techniques such as Cuckoo search with a 72.9085% and Firefly algorithm with 84.868% also with better results with various other parameters.

Keywords Regression Testing · Test Case Prioritization (TCP) · Cuckoo Search (CS) · Firefly Algorithm (FFA) · Term Frequency Cuckoo-Support Vector Machine (TFC-SVM)

O. Dahiya (✉) · J. Singh
School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India
e-mail: Omdev.26990@lpu.co.in

J. Singh
e-mail: janpreet.singh@lpu.co.in

K. Solanki · R. Rishi · A. Dhankhar
Department of Computer Science and Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, India
e-mail: Rahulrishi@mdurohtak.ac.in

S. Dalal
Department of Computer Science and Applications, Maharshi Dayanand University, Rohtak, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_20

1 Introduction

Software testing deals with examining the software for checking the presence of faults. It is either performed by manual evaluation if the developed application is small in size or by automation tools if the developed application is large, which generally happens. Two sorts of analyses are carried out in the software development process. Behavioral or physical assessment means identifying faults in application software without utilizing any device and is carried out by human beings. In this scenario, test cases are created to check that the testing is complete. In another case, the analyst writes down code in automatic testing, and robotic equipment is used, which enhances correctness, test range, and is less expensive than human testing [1]. Regression testing is executed for all releases in both kinds of testing, which checks whether no new faults have entered in the system due to changes done in it and whether the faults generated are removed, or they still exist, thus ensuring that the software is of good quality.

Test Case Prioritization (TCP) explained how fault detection is enhanced when test cases are arranged in a priority sequence. If a method computes the number of defects by following a certain sequence of test cases, they are then arranged accordingly by some defined criteria for the priority order. When the Average Percentage of Fault Detected (APFD) value was calculated for both prioritized and non-prioritized as well, it proved that more excellent value is achieved for prioritized test cases [2]. This showed that the TCP method maximizes the error identifying rate. It has also been proved that prioritized test cases attained more coverage in the initial implementation stage than non-prioritized [3–6].

1.1 Machine Learning

It is the process of carrying out automated analysis so that systems can be automated and the load on human beings can be minimized. Its use is increasing now because of its increasing need. Machine learning has three different categories: (i). Supervised learning- It comprises the target, speculated from a stated set of inputs such as decision tree, (ii). Unsupervised learning- In this, there is no objective and is used for clustering like K-means, (iii). Reinforcement learning- The device is assembled to acquire skill by revealing it to surroundings like the Markov decision process [7].

1.2 Problem Statement

The present prioritization methods were mainly coverage-based (statement, branch, or other coverage) and may be costly to employ and assure that complete errors are equally critical. Generally, fewer errors or faults may be severe (high-risk failure), a significant loss of economic clients, delaying product consignment, or security implications. The existing test case prioritization method can be improved by integrating extra knowledge gained by requirement engineering work [8]. A high complexity requirement can have higher chances of faults. Requirement unpredictability that results in re-design adds or deletes the additional requirement, which may increase the project risk, error density, or maybe increase the project failure rate. Generally, 20% of the system is accountable for 80% of errors. Thus, there is a requirement to increase the efficiency of testing resources that is;—Efficient that may not add essential overhead to teamwork, enhances testing resource efficiency, and recognizes more severe faults.

This proposed concept is mainly focused on enhancing the procedure of TCP by improved Term Frequency Cuckoo-Support Vector Machine (TFC-SVM) Method that can effectively give good test cases. The research work has different phases, such as (i) designing the TFC-SVM method using machine learning and soft computing-based methods [6]. (ii) Evaluating and comparing the proposed parameters such as Max, Min, Mean, Median, SD, APFD, and Execution Time.

The paper is arranged as follows. Section 2 provides explanations of related work in the TCP domain. In Sect. 3, an outline of the TFC-SVM approach. Section 4 consists of dataset description, evaluation parameters, and comparative result analysis. In the end, Sect. 5 concludes the paper as well as provides future scope.

2 Related Work

This section discusses the existing work of researchers in the test case prioritization domain. Various methods are used for test case prioritization, which helps in enhancing software's standard and fault identification rate, which eventually leads to customer gratification. The process consists of different parts such as user priority, alterations in demands, execution difficulty, convenience, application flow, and the effect of any fault. Bagherzadeh et al. (2021), have proposed a TCP technique by "reinforcement learning" in an environment of "Continuous Integration" (CI) so that test cases can be ranked with enhanced accuracy. This accuracy will boost the discovery of regression faults within a short time and with fewer resources; as in the context of CI, changes are taking place frequently [9]. Huang et al. (2021), have proposed a TCP technique by utilizing a "random forest algorithm" in combination with different metaheuristic algorithms. They have used the APFD metric to evaluate the performance of their technique. The results obtained show the enhanced APFD values in comparison with other existing techniques [10]. Geetha et al. (2021), have proposed a TCP technique which is the combination of the genetic algorithm with a "multi-level random walk reduction technique". Interesting results were obtained by them which suggests that a combination of the prioritization and reduction techniques along with genetic algorithm is a fruitful approach in this domain [11]. Gokilavani and Bharathi (2021), for prioritizing the test cases used "feature selection", "clustering" and "ranking" techniques. In the comparative performance evaluation, the authors have shown that the proposed technique has achieved efficient results on different parameters such as enhanced rate of fault detection, clusters ranking and diverse faults discovered corresponding to the test case number. They have compared their technique with different existing techniques to show the efficiency of their proposed method [12]. Lachmann (2018), has proposed a TCP technique that incorporates machine learning approaches to improve the effectiveness of the regression testing. The author used three diverse industrial-level datasets which are real-life datasets and has analyzed the result of their technique on the basis of potential fault discovery [13]. Lachmann et al. (2016), have used the machine learning technique for proposing a TCP technique. They have used the history of test cases and the description of test cases for prioritization. They have shown that their proposed technique outperforms random order prioritization and prioritization order given by a testing expert in terms of enhanced fault detection rate [14].

3 Proposed Methodology

One of the software engineering regions having relevant and practical usage of Artificial Intelligence (AI) methods is software evaluation, and these methods are known as Meta-Heuristic (MH) methods. MH methods are primarily utilized for resolving nearly all optimization issues [15–22]. The primary objective of MH methods is to effectively investigate the search space and order to discover an optimal output for any issue undergoing examination. This section provides a description of the TFC-SVM technique for regression test case prioritization (Fig. 1).

3.1 TFC-SVM Technique

The research work algorithm originality is concerned with the non-linear model's processing, which leads to a more fabulous Classification Rate (CR). Cuckoo, as well as Support Vector Machine (SVM), is also utilized along with term frequency, which helps in refining requirements based upon TCP. This is processed in the proposed technique [6]. It indicates the originality of suggested work as it attains more fabulous CR regarding high mean, median, and low standard deviations.

Proposed steps are described as:

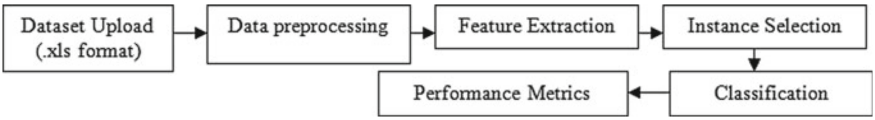


Fig. 1 Proposed methodology flowchart

Step I: First of all, uploading the data set retrieved from [23] is done through Graphic User Interface (GUI). GUI is an important concept that is helpful in the human–machine association. GUI utilized in the end-user association is constructed in MATLAB for beneficial data collection and simple visualization of information processing.

Step II: After that, Data Mining (DM) executes pre-processing so that valuable data can be collected. DM consists of regularization of data, which indicates the importance of the need and the task's priority to be carried out in the lowest implementation.

Step III: In the next step, extrication of term frequency attributes is carried out, which helps find Feature Vector (FV). FV will remove features in terms of frequencies of the request to be implemented in a test case to attain more significant priorities of the refined demands from the details.

Step IV: The process of choosing an instance is executed by cuckoo search optimization. The data exists in instances, so picking information is an essential step of processed requirement-based TCP. Instance picking executes optimization in terms of optimization heaviness and distance. The weight of the processed information, which is carried out sequentially, is determined, which lessens the repetition of data.

Step V: After this, training and testing are carried out. In these steps, classification is performed, which checks the priority level of test cases. This generates priorities for test cases. The outcome will create a trained model for the higher and lower priorities for the test cases to be executed.

Step VI: In the end, performance assessment is done based on various specifications, which are calculated by utilizing Computation Time (CT), APFD, Mean, Standard Deviation (SD), and minimum and maximum values, which helps in attaining performance evaluations for the system.

4 Comparative Performance Analysis

This section documents various parameters that are taken into consideration for evaluating and comparing the performance. In the comparative analysis, performance parameters are the following such as (i) APFD, (ii) Execution Time (iii) Min and Max (iv) Median (v) Standard Deviation, etc.

- Comparative Result Analysis

In the experimental setup, the machine with 8.00 GB RAM and Intel (R) Core ™ i7-4600 U CPU @ 2.10 GHz 2.70 GHz Intel processor is used. The simulation is done in MATLAB with GUI and has calculated the experimental results.

This section showed the comparative result with the Firefly optimization algorithm, cuckoo search algorithm, and TFC-SVM technique for TCP. The novel approach has improved the system performance and execution time rate as compared with the existing methods. In this section, APFD outcomes for an individual benchmark program are demonstrated using bar-graph-plots as defined in Fig. 2. The

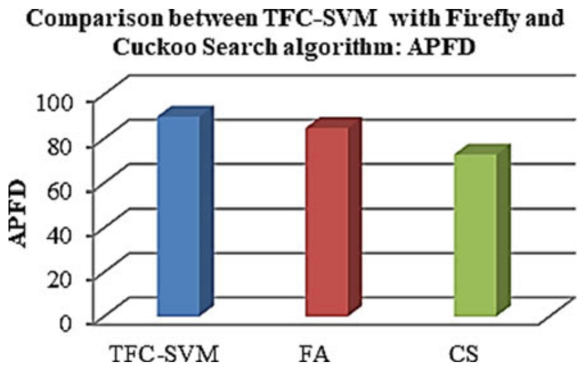


Fig. 2 Comparison with different algorithms: APFD value

Table 1 Comparison analysis of TFC-SVM model and existing (Firefly and Cuckoo Search) optimization models

Parameters	TFC-SVM	Firefly algorithm	Cuckoo search algorithm
APFD (%)	90.0232	84.868	72.9085
Execution time (ms)	108.445	142.666	71.0158
Max value Min value Mean	0.99436 0.18726 0.81071	0.96493 0.33159 0.27521	0.83623 0.64886 0.21527
Median	0.99952	0.77385	0.66357
SD	0.05731	0.72478	0.53266

complete metrics valuation for an individual benchmark program is also tabulated as defined in Table 1.

The analysis of values in Table 1, shows that the research work can attain high system performance for classification rates and more efficiency to attain minimum loss methods for TCP levels with existing ones.

Figure 2 presents the fault detection rate comparison, as it is one of the main parameters which depicts the TFC-SVM proposed method detection of average faults in the form of performance metrics. It should be maximum for the high-performance calculation to detect more faults in the given data set. Figure 2 shows the comparison between different TFC-SVM, FA, and CS algorithms. The Firefly and Cuckoo Search have achieved the APFD rate values to 84.8% and 72.9–73% and TFC-SVM with 90.0232%.

Figure 3 presents the execution time comparison. It defines that the TFC-SVM proposed method attained a standard range of execution time. It defines the comparison between different methods such as TFC-SVM, FA, and CS algorithms. With the smaller difference between existing cuckoo searches, the proposed TFC-SVM algorithm provides a high-performance classification rate and all other metrics. The existing firefly algorithm (FA) consumed a high range of execution time in all the executions with fewer performance metrics than the proposed (TFC-SVM) algorithm.

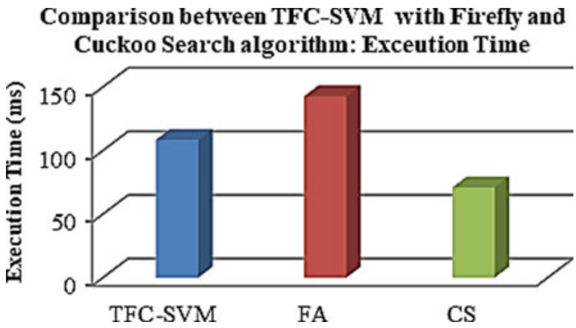


Fig. 3 Comparison with different algorithms: execution time

Figure 4 defines the comparison analysis with the proposed model, cuckoo search, and firefly algorithm approach in the form of maximum value rate, which must be high when compared to other techniques.

Figure 5 represents the comparative analysis of the TFC-SVM, Firefly optimization, and Cuckoo Search algorithms calculated in the form of the minimum value, which must be low.

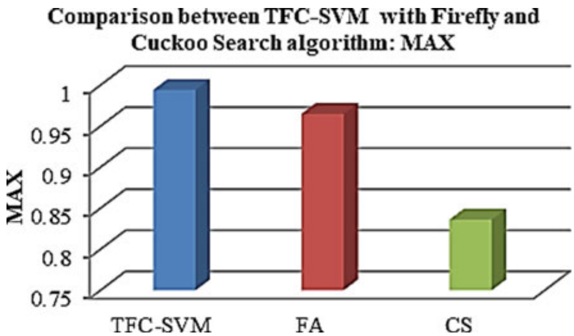


Fig. 4 Comparison with different algorithms: max value

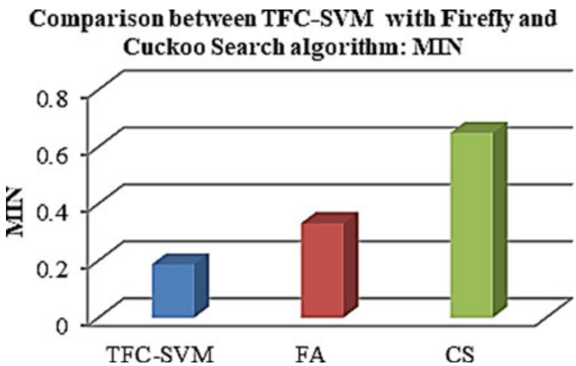


Fig. 5 Comparison with different algorithms: min value

Fig. 6 Comparison with different algorithms: SD (standard deviation) value

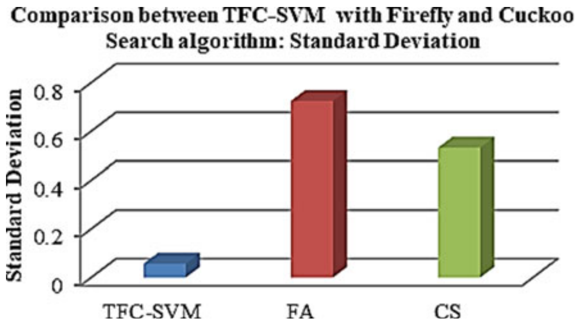


Fig. 7 Comparison with different algorithms: median

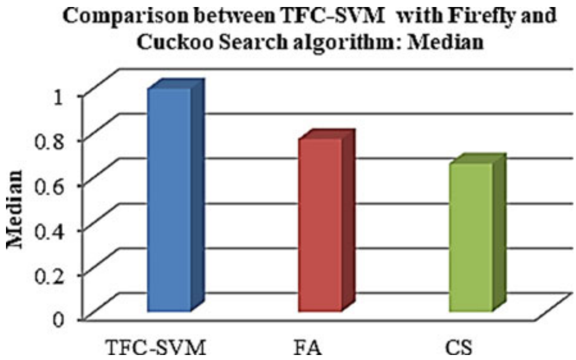


Figure 6 represents the SD performance metric comparison between TFC-SVM, FA, and Cuckoo Search Optimization algorithms. The TFC-SVM research method is well efficient to get minimum deviations from the mean distribution.

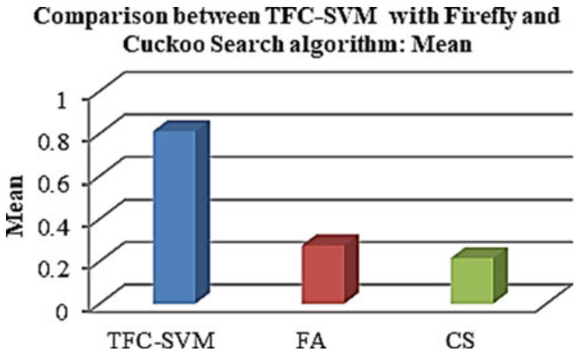
Figure 7 defines the comparison between the proposed TFC-SVM, Firefly optimization, and Cuckoo Search optimization algorithm in the form of MEDIAN values, which is also one of the essential performance metrics. The MEDIAN value is also supposed to be maximum.

Figure 8 represents the mean between the proposed TFC-SVM and existing algorithms (Firefly and Cuckoo Search Optimization), a normal distribution procedure and this must be high for low error rates.

5 Conclusion and Future Scope

In this research work, we compared the performance of different metaheuristic algorithms for TCP. The research's main motive was to classify the rates, optimize the execution time and cost of testing. The outcomes of the analysis suggest if the fault is detected speedily at the current phases of development, then the cost and execution time of software testing has been less. The proposed TFC-SVM algorithm has

Fig. 8 Comparison with different algorithms: mean value



optimized the feature sets, classifies the data rate, and improves the system performance. This proposed model has compared with the techniques like Cuckoo Search and Firefly Optimization algorithms. The proposed model has enhanced the min, max, median, SD (standard deviation), execution time, and APFD rate performance metrics. This comparison algorithm showed that it improves the APFD, execution time; mitigates the error rates and existing problems.

Further improvements can implement a cost-effective and deep learning method for enhancing TCP and regression testing procedures by prioritizing the test cases.

References

1. Joshi SA, Tiple BS (2014) Literature review of model based test case prioritization. Int J Comput Sci Inf Technol 5(5)
2. Arafeen MJ, Do H (2013) Test case prioritization using requirements-based clustering. In: 2013 IEEE sixth international conference on software testing, verification and validation, IEEE, pp 312–321
3. Salem YI, Hassan R (2010) Requirement-based test case generation and prioritization. In: 2010 International Computer Engineering Conference (ICENCO), IEEE, pp 152–157
4. Dahiya O, Solanki K (2021) An efficient APHT technique for requirement-based test case prioritization. Int J Eng Trends Technol 69(4):215–227
5. Dahiya O, Solanki K (2018) A systematic literature study of regression test case prioritization approaches. Int J Eng Technol 7(4):2184–2191
6. Dahiya O, Solanki K (2021) An efficient requirement-based test case prioritization technique using optimized TFC-SVM approach. Int J Eng Trends Technol 69(1)
7. Meçe EK, Paci H, Binjaku K (2020) The application of machine learning ın test case prioritization—a review. Eur J Electrical Eng Comput Sci 4(1)
8. Srikanth H, Williams L (2002) Requirements-based test case prioritization. In: Doctoral symposium in ınternational conference of software engineering
9. Bagherzadeh M, Kahani N, Briand L (2021) Reinforcement learning for test case prioritization. IEEE Trans Softw Eng
10. Huang Y, Shu T, Ding Z (2021) A learn-to-rank method for model-based regression test case prioritization. IEEE Access 9:16365–16382
11. Geetha U, Sankar S, Sandhya M (2021) Acceptance testing based test case prioritization. Cogent Eng 8(1):1907013

12. Gokilavani N, Bharathi B (2021) Test case prioritization to examine software for fault detection using PCA extraction and K-means clustering with ranking. Soft Comput 25(7):5163–5172
13. Lachmann R (2018) Machine learning-driven test case prioritization approaches for black-box software testing. In: The European test and telemetry conference, Nuremberg, Germany
14. Lachmann R, Schulze S, Nieke M, Seidl C, Schaefer I (2016) System-level test case prioritization using machine learning. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, pp 361–368
15. Yang XS, Deb S (2010) Engineering optimisation by cuckoo search. Int J Math Modelling Numerical Optimisation 1(4):330–343
16. Payne RB, Sorensen MD (2005) The cuckoos, vol 15. Oxford University Press
17. Yang XS, Deb S (2014) Cuckoo search: recent advances and applications. Neural Comput Appl 24(1):169–174
18. Yang XS, Deb S (2009) Cuckoo search via Lévy flights. In: 2009 World congress on nature & biologically inspired computing (NaBIC), IEEE, pp 210–214
19. Barthelemy P, Bertolotti J, Wiersma DS (2008) A Lévy flight for light. Nature 453(7194):495–498
20. Nagar R, Kumar A, Singh GP, Kumar S (2015) Test case selection and prioritization using cuckoos search algorithm. In: 2015 international conference on futuristic trends on computational analysis and knowledge management (ABLAZE), IEEE, pp 283–288
21. Kruse R, Borgelt C, Braune C, Mostaghim S, Steinbrecher M, Klawonn F, Moewes C (2011) Computational intelligence. Vieweg+ Teubner Verlag
22. Singh A, Vandana, Singh R (2018) Regression test case minimization with firefly based algorithm. Int J Comput Sci Eng 6:336–341
23. Mondal S (2021) Data files for Mahtab: phase-wise acceleration of regression testing for C. Retrieved 15 January 2021, from https://data.mendeley.com/datasets/7fvwj88jvm/2

Future of Work: How Artificial Intelligence Will Change the Dynamics of Work Culture and Influence Employees Work Satisfaction Post-covid-19

Rashmi Singh and Preeti Tarkar

Abstract This is evident that covid-19 has a devastating effect on businesses and employees worldwide. As we are slowly moving toward retrieval, it is very clear that the business and the way of doing work will change forever in the new normal phase. The experience of Covid-19 has drastically fastened the organization's digital transformation. Thus, the purpose of our study is to explore the impact of AI-enabled work culture on employees' satisfaction level post-pandemic. Companies now focus more on data-driven strategies and decision-making, which seems critical today for their business survival and success tomorrow. Few people may argue that technology has always been a part of an organization's work culture then; what is new? Moreover, the answer will be the overnight shift of work on-site to work from home, leading to a drastic change in the work culture. The situation caused due to covid-19 has forced the employees and organizations to rely fully on technology, which had made the organizations adopt the use of technology as it was never. Organizations today are investing more in digitalization and automation to secure their business ventures in these shifting dynamics. However, how will AI change the dynamics of future work? As the digitalization and use of technology have evolved, AI development has increased the fear of losing their jobs in people's minds influencing their work satisfaction. Therefore, our paper aims to study the impact of artificial intelligence on the work culture and its effect on employees' work satisfaction. Moreover, the paper also discusses the related gap found in the skillset of employees and managers to work on AI and other related technologies during the pandemic and the probable solutions. The responses of 150 employees were collected, and the results have concluded that AI-enabled work culture will significantly impact the employees' work satisfaction post-covid-19.

Keywords Artificial intelligence (AI) · Automation · Work culture · Work satisfaction · Covid-19 · Work from home · Information technology (IT)

R. Singh (✉) · P. Tarkar
Department of Management, GLA University, Mathura, India
e-mail: rashmisinghmathura@gmail.com

P. Tarkar
e-mail: preeti.tarkar@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_21

1 Introduction

The sudden hit of Covid-19 has created an alarming situation worldwide, and as an urgent need to respond to the pandemic and continue the business activities, AI seems to be the only way out for the companies. As per [1], the Covid-19 pandemic has fastened the digital transformation up to 59% in the organizations surveyed. Before the advent of Covid-19, the application of AI in organizations was rare. A study [2] highlights that 80% of managers consider AI a key competitive factor, but the actual application of AI is less. Only one-fifth of the companies have implemented chatbots, and numerous other AI-enabled applications are still in the initial phase of planning [3]. However, today, the changing workplace dynamics due to Covid-19 and the shift of work culture into the work from home paradigm are now forcing the executives to prioritize their competencies.

The ongoing pandemic has altered the work culture forever by changing the dynamics of how organizations operate. For example, during pandemics to reimburse the inaccessibility of employees, AI has supported the organizations even for their content monitoring on YouTube, and chatbot has served the purpose of customer service providers [4]. In addition, semi-autonomous robots have helped with cleaning and sanitization and even delivering medicines and food [5]. Therefore, this has increased the executive's confidence in what technology can ensure, making them push the digital transformation ahead. As a result, companies are now investing in digitalization and automation [6].

The organizations focus more on changing their strategies and policies, integrating AI in their operations. As per the survey [1], organizations have shifted their operations to more cloud-based activities, automation processes, change management, and digital transformation.

Adopting AI-enabled culture has made the organizations automate their business process, enhance their efficiency and speed, and focus on higher-value work. However, there are chances that this will lead to employee dissatisfaction as it has created a fear in employees' minds to lose their jobs. As per the estimates of Oxford Martin, in 2016, 37% of the jobs in the UK, 47% of the US, whereas 77% of China jobs are at risk due to automation. This is because the organizations focus on employees' health, safety, and flexibility but have deprioritized their satisfaction. In addition to this, the increasing priority of cost management also hampers the employee's satisfaction and support. However, some experts argue that the scenario will not remain the same. According to the world economic forum experts, technology may take away 75 million jobs in 2020, creating 133 million new jobs. Therefore, AI will generate more jobs than it terminates by 2025, changing employees' current jobs [7].

During the pandemic, digital technologies like web conferencing, video calls are being used to sustain interaction and social communications. Moreover, work from home culture has even made the individuals more acquainted with the comfort of using technology, increasing their trust, and shaping opportunities for the technology-enabled workplace culture. Culture plays an important role in understanding how individuals gain familiarity and interact with technology through cultural learning [8]. The cultural theory used by the information system suggests an idea of the IT culture linked to the values adopted by the people [9]. This idea will help understand the people's social practices when they interrelate with the technology and help to explain their preferences to use it [10]. Therefore, our study focuses on the future of work and will highlight the changing dynamics of work culture due to pandemics.

1.1 Research Gap

Various studies have been conducted that study the impact of AI on work satisfaction. However, the Covid-19 pandemic has brought a huge change in the workplace resulting in the increased application of technology worldwide. As a result, they are now investing more in AI and automation, which will surely change the future of work. Therefore, our research focuses on the changing dynamics of work culture due to AI-enabled workplaces. Moreover, in our research, with the help of quantitative data, we have analyzed the impact of AI-enabled work culture on employee satisfaction post-covid-19.

2 Role of AI in Various Sectors During Covid-19

Since the introduction of AI, it has always been a topic of discussion and has attracted much hype in the past few years because of its rapidly growing contribution to the various sectors. AI is a self-regulating system that examines many possibilities while finding the optimal solution. By the passage of time, the current role of AI has undergone drastic changes due to the widespread contribution of research and development in AI. Many developing countries like China prioritize technology usage by increasing their investments in IT industries [11]. Moreover, during the covid-19, AI became the most critical and powerful device to reduce the virus spread [12] and revive the economy from the losses through business continuity. The tremendous contribution of AI during the global pandemic situation has helped various sectors to revive, and it has played a vital role in different sectors, mentioned below:

2.1 Healthcare Sector

Considering the seriousness of the disease, substantial attention has been given to its contribution to the health sector. Various AI-enabled solutions are being used to stop the spread of the virus and even deliver medicines and take consultation sessions from doctors through video calls [13]. AI has helped develop IT-enabled solutions, which have proved to be very effective and efficient in forecasting, identifying, and handling chronic and minor diseases [14]. IT is important for diagnosis and treatment and contributes to developing treatment conventions and patient monitoring without being physically available [15, 16]. Artificial neuron network to stimulate the nerve cell network to the brain is one of the best examples of the role of AI in the healthcare sector [17].

Similarly, robots are yet another major area where AI has contributed significantly to the Covid-19. Robots have compensated well for humans' need to clean and sanitize the hospitals [18] and even for the patient's treatment. The robots equipped with cameras helped the doctors to monitor the patient's condition and check their reading on screen any time they wanted. Therefore, it also helped protect many doctors and nurses from the spread of disease by avoiding contact. Moreover, smartphone applications have helped people in risk assessment at home and provide real-time information about the infected patients to healthcare experts [19].

2.2 Business and Manufacturing Sector

As the covid-19 cases started to increase worldwide, the government had implemented restrictions on the travel and movement of people. This has increased the challenges of businesses to continue and grow. Considering the situation, the companies support AI for their retrieval. One of the immediate effects of coronavirus was the shift of on-site work to work from home. AI-based virtual care solutions have helped executives address employees' mental stress. AI chatbots helped answer the employee's queries effectively, either related to return-to-work plans or payments and holiday privileges [20]. AI not only helped the business manage the problems of employees but also helped manage their customers. Interactive voice response solutions helped provide quick answers to the customer's query without the help of an agent. Websites and messages through SMS, Facebook, or WhatsApp messenger are yet another effective way to answer the customer's query appropriately. AI-enabled chatbots help provide quick solutions to many business issues [20].

2.3 Banking Sector

Even before the advent of covid-19, banks were the major users of AI. The machine learning system of AI helps banks detect fraud, helps them stay in contact with their customers, and helps them meet the governing necessities [21]. In addition to this, the pandemic has accelerated the application of digital transformation to meet the new data processing requirements. The first wave of covid-19 has resulted in a 10–20% increase in the use of mobile and online banking across Europe [22]. In addition to these digital solutions have helped the banks increase their operational efficiency through data screening and by prominent use of incoming mails scanned through optical character recognition, which allows in smart cataloging along with metadata and transfers to the digital mailbox. This service helped customers to avoid physical obligations. AI also helps the banks serve their customers by providing advanced portfolios, offerings, and advisory services [23].

2.4 Agriculture Sector

AI turned out to be a powerful tool for all the sectors during the covid-19 pandemic. It has presented new ways of creating and managing the business and linking them through innovation and sustainability. Agriculture is the sector that requires AI and deep machine learning, blockchain, cloud computing, IoT (internet of things) to keep track of the supply chain [24]. During the pandemic, AI appeared to be the most appropriate way of handling customers' requirements and improving the quality of agricultural products.

The use of AI has paved the way for increasing the growth of the agriculture sector through numerous IT solutions and advancing the interactive software program. After covid-19, the anticipated market size of AI in agriculture is predicted to reach USD 186 billion by 2025 [25]. In addition to this, AI is also helpful in assessing food products within a few seconds and helps reduce waste and labor costs.

2.5 Hospitality Sector

The hospitality sector is one of the majorly affected sectors due to covid-19. The tourism sector is very much prevalent to AI as many technologies highly drive it. Experts have predicted that in the next 75 years, the whole hospitality sector will go through a huge transformation of digitalization [26]. Several hotels have already been using the technology for many years for all their operations ranging from office work to maintaining interaction with customers [27]. However, the recent hit of the pandemic has widened the scope of technology in the hospitality sector even more. The sector is now investing more in automation and digitalization to ensure their customers' safety and gain their trust. They are using no contact applications which customers can operate from their smartphones. They are using AI and robots for cleaning, sanitizing, and even for the delivery of food and other services. Webster and Ivanov [28] has specified the essential role of AI and robots in the hospitality industry. They have appealed that AI will have a major impact on trade, income, work, and social, political, and economic perspectives in the future.

Moreover, the hospitality sector uses AI for real-time, building correlations, and enhancing customers' experience. Today, AI is becoming crucial for satisfying the hospitality sector's key necessities, enhancing their operations in various ways [29]. Adopting AI through adequate understanding will also help this sector revive the pandemic loss, eliminating uncertainty and risk.

3 Impact of AI on Work Culture

Work culture plays a very important role in enhancing the organization's performance. An ineffective work culture harms the organization's effectiveness and reduces the shareholder's return. An organization's work culture helps distinguish one organization based on beliefs, values, and behavior patterns that the organization possesses [30]. An organization with a strong work culture influences employees' perception of the organization and develops a strong working attitude to motivate them to improve performance [31]. A strong work culture indicates a positive relationship between employers and employees by establishing transparent and consistent communication. At the same time, a weak work culture lacks a uniform direction and poses a risk for the organization's existence [32].

Innovation has always been an effective tool for improving performance [33]. Hence, the development of AI and technology progress leads to drastic changes in the work culture. A study conducted by [34] acknowledged that AI is likely to influence the future of work and, therefore, implications work culture and HR functions. Moreover, the recent hit of covid-19 has led to the frequent involvement of AI in organizations. Furthermore, the shift of work to remote working and flexible working hours has enhanced the involvement of AI in work culture even more. According to a report by walking and talking (global culture organization), AI will bring considerable changes in the organizations to survive in the cut-throat competition and take the best possible advantage of the technology; firms must build a culture of innovation. In today's, time the researchers have been directed about the capability of machines to read the emotions of humans, where it studies the human emotions in a particular scenario and then adapts its behavior to give precise responses.

According to walking and talk, the AI will influence the work culture in two major ways: 1: AI helps manage the work culture; for instance, it helps conduct an existing cultural assessment by interpreting the behavioral data sets of employees. 2: Helps to identify the target culture: Internal and external data helps predict the best target culture that monitors the behavioral changes and calculates the business results. Moreover, AI also helps develop a cultural plan for the firms by identifying and eliminating the current risks related to a specific problem and their significances [35].

3.1 Ways in Which AI is Impacting the Work Culture

Fostering Environment of Innovation and Creativity

Covid-19 has made a philosophical shift of work culture where the routine tasks get automated, and employees' focus shifted to the strategic and creative ways of problem-solving. In addition, the organization's performance metrics also moved toward collecting information on how workers can create value for the business through the most creative ways. Therefore, along with the routine work, the employees can work on new technologies that will positively impact the organizations' work culture [36].

Motivate Employees to Experiment

AI-enabled work culture helps organizations to do experiments. Companies try to formulate and work through new customer interaction methods, which help enhance customer experience and improve employee engagement. AI also helps companies eliminate the risk by letting them work on their innovative ideas without anxiety [36].

Upskilling of Employees

With the help of predictive analysis, AI helps organizations find the gap between the existing skills and the required skills of an employee. Therefore, it helps to know that the prospective candidate will fit the existing work culture. Moreover, AI also helps develop learning coursework for employees to get trained for the appropriate skills required [37].

3.2 AI and Employees Engagement Impacting Work Culture

Employee engagement differs from company to company and employee to employee. Engaged employees always find meaning and purpose in their work. These employees are self-motivated, enthusiastic, and are focused on growth. Organizations do performance assessments and annual reviews to analyze the amount of employee engagement in their organization. Employees working in an organization are willing to allocate their time and effort only when they feel motivated. A system based on AI promotes employee engagement by guiding, analyzing, rewarding, and punishing their activities [38]. AI-enabled employees control system helps organizations to control the performance of the workers. The two major examples of this kind of control system are 1. Behavior outcome, which directly monitors the employees' activities, and 2. Outcome control measures the results of employees' activities [39–41]. Uber application is one example of companies using AI control systems to ensure their employees' engagement. AI tells the driver where to go, whom to pick, how to reach the destination, and even how to drop the client to their destination [42]. Chatbots, real-time feedback is some of the examples which are helping companies in their employee engagement through AI.

3.3 AI and Employee's Perception Impacting Work Culture

The emergence of covid-19 has made the companies trust the technology and made the organizations' executives understand the necessity of implementing the technology [43]. As a result, companies focus more on adopting technology and investing in automation and digitalization. It is important to know employees' perceptions of this major work culture shift. From the past few years, it is evident that employees fear AI will replace their jobs [44]. Most employees, including managers, are fearful that if AI is part of organizational culture, less educated and white-collar employees will be impacted the most [45]. However, situations like the current pandemic suggest making humans and machines work together and highlight a need to reskill or hire employees with AI skills [46].

In a recent study [47], they have recommended that for the implementation of AI, the organization will have to go through the drastic changes of redesigning and reshaping and will require a change in the skillset of employees to complete a particular activity or task. Employees' struggle to learn and work on smart technology like AI will subsequently impact an employee's work satisfaction and commitment, leading to increased employee turnover intention and depression. Contradicting this, a study conducted by [48] emphasized that employees need to enhance their skills and knowledge to deal with the technology. Humans are irreplaceable and will always remain an essential part of organizations. Their study has also emphasized that machines and humans will complement each other as much as they interact. Therefore, the key to remaining employed is upgrading oneself and modifying oneself for the best use of technology. In addition, AI assisting employees to work will increase their work satisfaction [48]. In the study of [49], they have suggested that organizations should effectively communicate with the employees, create awareness among the employees, and encourage them to adopt an open mind for the smooth working of the organization.

4 Employee's Skills Related to AI

The use of AI in organizations has been rapidly increasing since the day the world has experienced the massive hit of a life-threatening disease, covid-19. This leads to digitation and automation of various activities which the employees earlier carried out. However, the rapid technology advancement may have major consequences on laborers' and employees' futures [50]. Furthermore, as AI and automation are transforming businesses, they are also reshaping the present jobs and are giving rise to new job roles. Hence, there is a need to fill the skill gaps in the procedures, and upskilling the employees is one of the critical tasks of the organizations.

4.1 Why is There a Need to Upskill Employees Now?

Jobs related to AI increased by 23% in 2019. According to Canadian data analysis, 43,000 new jobs were created by the market in the same year. However, few skilled employees are available to fill the posts, and this scenario is predicted to be widened by 2031. In Canada, 63% of the managers feel that their companies will face the maximum skill gaps in the next two years. In addition, the lack of digital literacy is creating problems in the effort to invest in AI and other technologies. As per a study conducted by Deloitte, only about 16% of the companies have used AI to continue their work in 2019. Upskilling the employees and managers will create an understanding among them about the usage of AI, the skills they lack, and the skillset they need to upgrade [51].

There can be various levels of employees are working in an organization. The companies have to deal with vendors, workers, retailers, etc. hence, hiring solely the already skilled staff to use AI is difficult. Therefore, they should adopt a multi-prolonged method to reduce this skillset gap. The staff should be provided with the right combination of training, learning, and on-the-job experience to help the organization build a business-future-ready workforce. Training the existing staff will lead to even more benefits for the organization as the existing staff is already well aware of the work culture and current technological infrastructure. At the same time, the organizational commitment toward the existing staff to train and upgrade their conceptual skills will help retain and attract the top talent.

Demand for technology advancement and application in organizations is increasing rapidly. Therefore, the skills required to process and work on AI will increase simultaneously. Emotional, social, and other cognitive skills that need complex and creative thinking also have excessive demand. Due to the current pandemic, the demand for digital skills is also increasing and is expected to continue in the future.

According to [52], the change in working hours in Europe and the United States, 2016 versus 2030, will bring a shift of 203 to 174 for physical and manual skills, reducing it to 14%; it will shift basic cognitive skills from 115 to 97 reducing it to 15%, higher and cognitive skills it will result from shifting 140 to 151, resulting in 8% increase for the same, social and emotional skills will shift from 119 to 148, with an increment of 24 percent. Technological skills will shift from 73 to 113, with an increase of 55%, the highest among all the skills. Therefore, the existing digital trend may reduce the manual and physical skills, but it will remain the major group of personnel skills by 2030. However, it will increase the challenges of current workforce skills, create a demand for a new credentialing system, and require innovative solutions.

5 Use of AI Positively Strengthens the Workplace Culture

AI offers much potential to the organization [53]. AI will change the future of work and is predicted to influence business strategies, business models, sales processes, customer service options, and customer behavior [54]. AI technologies can be productively functional to many domains, including education, research, businesses, and others [55]. Systems based on AI technologies are less expensive, easy to use, and more effective in solving organizations sought-of problems [56]. AI-enabled technologies help create opportunities for businesses, support them by ensuring the appropriate availability of all resources, and improve the quality of products and services they offer to their customers [57]. AI helps to enhance the performance of the organizations at both functional level and process level and hence results in improving the value of the businesses by transforming projects.

5.1 Improve Operational Flexibility

According to IDC estimates, 2019 will improve digitalization by 40% in the service sector, and 75% of businesses will use AI in their processes by 2021. AI helps the organizations improve operational flexibility, which helps the employees access support to operate various systems and adapt to their changing dynamics and business strategy. A flexible AI structure allows employees to use AI resources effectively, prevent organizational issues through self-optimization and a self-configuring structure, promote innovation in business strategies, and will help in enhancing the performance of the organization through available resources [58].

5.2 Enhance Supply Chain Performance

Organizations are required to develop critical competencies to enhance the performance of their supply chain, hence improving the companies' overall information competencies. The technologies of AI such as advance tracking, industry 4.0, blockchain help in minimizing the risks associated with the supply chain and improve the supply chain performance. Moreover, AI also helps eliminate the supply chain risks associated with privacy, hacking, political interventions, and financial instability. AI provides innovative solutions to managers to improve the supply chain and enhance their profits by developing innovative ways to design new products, solve customers' problems, and create new ways to tackle risk [59].

5.3 Upskilling and Self-development

AI helps the employees in maintaining the existing skills along with helping them to learn and adopt new conceptual skills. By doing so, organizations encourage their employees to engage in long-term learning programs and to enhance their motivation and satisfaction at the same time. In a study conducted on the occupation of pilots, David Mindell stressed that the pilots master the major skills required for flying a plane on their own, but the higher-level performance of the Autoland system for a few functions of pilots helps to motivate them to enhance their skills. Therefore, for handling AI, traditional skills are essential, along with acquiring new conceptual skills [60].

5.4 Connectivity

The use of AI technology scales up the learning ability of the whole system rather than a single system. Hence, working on AI brings out pooled results in terms of collective intelligence. IoT (the internet of things) is the best example of such cooperation. This suggests that if an individual fulfills a goal, a more conceptual intelligence arises from it. For example, if a self-driving car makes an error, then with the help of collective intelligence, all the other self-driving cars will learn from that and eliminate the error. Therefore, AI helps to build a network in the workplace that enhances the performance of the whole network [61].

6 Adopting AI

AI is grabbing the attention of organizations and individuals very frequently nowadays. Viewing the popularity of AI-enabled solutions and robot organizations shows their high interest in adopting and making AI an essential part of the organization to automate their processes. AI-enabled solutions eliminate the chances of human error through smart decisions and enhance the reliability and accuracy of operations. According to experts, by 2030, 25% of the daily tasks will be completed by robots.

As per a survey conducted by BCG and MIT, 84% of the companies adopt AI because of competitive advantage, 75% because of a shift in business, 75% because of the shift in markets, and 63 percent to reduce the cost. This specifies the major strategic role of AI and provides various advantages to the company, like reducing costs, ensuring operations efficiency, and increasing profit [62]. AI can estimate the various results of businesses accurately [63]. Adopting AI in the organization will lead to the implementation of AI in the business process and the acceptance of various workers and employees. However, adopting AI is a complex task as it gets influenced by many constraints. A wide range of understanding is required to adopt AI, which depends on the adopting behaviors of the employees and businesses.

6.1 Framework for Adopting AI

The framework for adopting AI is shown in Table 1:

Table 1 The framework of Adopting AI

Technology-organization-environment framework (TOE) explains the TOE factors that hamper the adoption of AI and innovation
Diffusion of Innovations (DOI), this concept describes the range to which digitalization diffuse in firms and why and how they will spread. It is also used to study the adoption of AI at both individual and business levels
Dynamic Stackelberg Games (DSG) describes how inadequate data influences the new technology's planned decisions
Best Worst Modeling and Binary Choice Model, this survey method is used to examine people's behavior toward digitalization and innovation. It uses the most appealing and the least appealing constraints to measure people's behavior toward adopting AI
Uncertainty Reduction Theory (URT) helps reduce the chances of uncertainty by determining the actions of others with the help of collecting information
Agent-Based Simulation Modeling, this concept helps in classifying the individuals whose behavior is different but are put together in a group. It simulates the activities and inerrability of independent agents to measure the effects on the whole system
Theories of Risk Perception and Trust include constraints such as perceived profit and risk, predicted perception, and trust while perusing adoption
Cognitive Dissonance Theory helps know the people's uniformity in their opinions and behavior. If there is a difference in agreement, then the change in attitude will be adopted to resolve the dissonance
The multivariate Probity Model examines the interlinked decisions by assessing numerous associated binary results together
Watching the Eyes Effect, to maintain the reputation of oneself, individuals behave selflessly before watching the eye
Bandura's self-efficacy theory describes the confidence in oneself that one can influence any event
The Media Richness Theory (MRT) defines the relationship between excellence and features of the channel of communication and communication activities
Uses and Gratification Theory is used to describe why individuals adopt particular innovations. It works on the assumptions of customers not being passive. It helps generate an understanding of the adoption of technologies like Siri and other related technologies

Source [64]

7 Impact of AI-Enabled Work Culture on Employees Work Satisfaction Post-pandemic

In the above sections, we have seen that the first wave of covid-19 turned into a huge wave of technology that has affected future organizations' work. This work shift has made the organizations develop and adopt the technology and declare that their companies will be fully digitalized in the coming years. However, despite being a major contributor to innovation and creativity in organizations, AI is a big threat to human service jobs [61]. Various studies indicate that AI positively impacts employees' work satisfaction. Bhargava et al., 2020 findings indicate that employees perceived AI and automation as opportunities, not threats. Donepudi et al., 2020 in his study highlighted that the use of AI and automation is not meant to replace humans but to accolade them. Additionally, studies demonstrate that AI replaces two workers in the companies on an average while two jobs have been formed outside.

Moreover, few studies indicate that AI has a negative impact on employees' work satisfaction. A study conducted by Schwabe & Castellacci, 2020 confirms that the use of AI in industrial firms has persuaded 40% of the employees to fear that their jobs will be replaced by AI and other smart machines in the future, which is negatively impacting their present job satisfaction. In their study, Franken and Wattenberg, 2019 emphasized that the influence of AI should lead to cognitive and physical relief, securing and increasing jobs, and the enhancement of work-life balance but, there are more employee concerns related to disqualification, fear of losing jobs, and rising self-sufficiency of IT systems. The job satisfaction level of low-skilled workers tends to have more negative effects on them in terms of high-skilled workers as they have fewer opportunities to find a new job in the labor market if they get replaced by AI.

There are various studies conducted in the past that focus on the impact of AI on wages and demand in different industries, and there are also few studies that contribute to the impact of AI on job satisfaction. However, a sudden shift of work toward technology due to covid-19 has shifted the focus, and therefore, the present paper specifically analyzes the relevant changes held in work culture due to the increased use of AI and its impact on employee's work satisfaction. For conducting the study, 150 employees' responses were collected based on which the data was analyzed. Therefore, the hypothesis framed is:

H1: AI-enabled work culture significantly impacts employees' work satisfaction post-covid-19.

8 Research Methodology

The study is quantitative, and the data is collected through convenience sampling, which is one of the methods of non-probability sampling used by the researchers to gather the data from the respondents that they feel are convenient to them for collecting the data. This method helps the researchers to get responses easily and quickly. The sample size of the study is the respondents working in both private and public enterprises as the emergence of covid-19 has changed the way of working, and the use of technology has majorly increased swiftly in both these organizations. The data was collected from the employees working in different states of India. The questionnaire was prepared with the help of a google form and was distributed among the respondents through email and LinkedIn, for which a total of 150 responses were received. The survey measures were self-administrated, consisting of 12 questions, and divided into two categories. In the first category, the items were placed to know the organization's AI-enabled work culture, and in the second category, the items were placed to know the satisfaction level of employees due to increased use of AI during covid-19 and the reliability of the scale was checked. The measurements of the survey

Table 2 Correlation

	Mean	Std. Deviation		AI-enabled work culture	Work satisfaction
AI-enabled work culture Work satisfaction	41.59	6.472	Pearson Correlation	1	0.607[a]
			Sig. (2-tailed)		0.000
	28.66	5.645	Pearson Correlation	0.607**	1
			Sig. (2-tailed)	0.000	

[a, **] Correlation is significant at the 0.01 level (2tailed)

item in this study are constructed on a seven-point Likert scale ranging from 1–7 in which one was for "highly disagree," 2 for "disagree," 3 for "somewhat disagree," 4 for "neutral," 5 for "somewhat agree," 6 for "agree," and 7 for "highly disagree." The total number of male respondents was 69%, whereas female respondents were 30%. The data collected with the help of a questionnaire was analyzed with the help of SPSS.

9 Data Analysis and Results

9.1 Descriptive Statistics and Correlation Analysis

Table 2 depicts the descriptive statistics of the measured items, which include the mean and standard deviation among the factors. The respondents have demonstrated positive concern about both the factors items. On average, the mean value of the first seven items under factor AI-enabled work culture is $M = 41.59$, and the mean value of the rest of the five items under factor Work satisfaction is $M = 28.66$. The Pearson correlation between the factors reveals that all the hypothesized relationships developed were statistically significance at level p, 0.01, which means AI-enabled work culture ($r = 0.607$, p, 0.01) positively correlates with work satisfaction significantly. As shown in the table, both the factor values are 0.607, which shows that these values are correlated. According to [65], the correlation values should not go beyond 0.8 for the persistence of evading multicollinearity. In our study, the largest value of the correlation coefficient is 0.607, which means it is less than 0.8, which depicts that the data does have a problem of multicollinearityF.

Table 3 Factor analysis and Scale reliability

Variables	Scale items	Factor loading	Eigen values	KMO	% Of cumulative variance	Cronbach's Alpha
AI-enabled work culture (How does your organization make use of AI during Covid-19?)	Give you work from home	0.742	3.377	0.813	48.241	0.815
	Does your work involve the use of advanced technology daily?	0.734				
	My organization still need to enhance the use of AI (technology) in business operations	0.733				
	Does your leader inspire you for the use of AI?	0.691				
	My organization Foster an Environment of Creativity and Innovation	0.672				
	Does your company enthusiastically push AI-enabled work Culture beyond the traditional work culture?	0.665				
	Rate the idea of translating AI into Work culture for gaining competitive advantage	0.615				
Work satisfaction	I fear losing a job due to the increased use of AI in my organization	0.844	3.161	0.750	63.219	0.844
	Has AI increased my task complexities?	0.835				

(continued)

Table 3 (continued)

Variables	Scale items	Factor loading	Eigen values	KMO	% Of cumulative variance	Cronbach's Alpha
	My stress increased due to the implementation of AI in the organization	0.802				
	Your perception of AI impacting your current and future job satisfaction	0.790				
	I feel demotivated and effortless due to the increased use of AI	0.695				

10 Factor Analysis and Scale Reliability

Table 3 depicts the validity of the factors is assessed with the help of factor analysis, and in order to assess the validity of both the factors separately, the KMO value is examined, and the values of both the factors are more than the threshold value of 0.5. The Kaiser–Meyer–Olkin (KMO) value of all 12 items together is more than the threshold value of 0.5, which is 0.850, which means that our sample size is adequate to conduct the study [65]. In the study, Bartlett's test is significant as the value of all the items together and for both the factors AI-enabled work culture and work satisfaction separately, the p-value is 0.000, which is less than the alpha value 0.05 (p-value < 0.05). Therefore, Bartlett's KMO value and significance highlights the relevance of factor analysis for the data.

Table 3 highlights the factor loading of every item in the scale, sorted by size. According to [66], "factors with eigenvalue below one should be omitted and factor loading less than 0.5 will be eliminated" [67]. In the given data, the eigenvalue of each factor is greater than 1.0, and the percentage of cumulative variance for both the factors is almost more than 50%, and the percentage of cumulative variance for all 12 items together is 65% which is also more than 50%.

Table 5 shows that the total number of respondents was 150, out of which all 150 responses were valid. For testing the reliability of the scale, the Cronbach's Alpha value is being analyzed, and the Table 4 depicts that for all the 12 items within the scale, the overall score of Cronbach's alpha value was 0.878, which is higher than the preferable score value, which is 0.70 highlighting that the scale is reliable and

Table 4 Reliability statistics

Cronbach's Alpha	N of Items
0.878	12

Table 5 Case processing summary

		N	%
Cases	Valid	150	100.0
	Exclude[a]	0	0.0
	Total	150	100.0

[a]Listwise deletion based on all variables in the procedure

consistent enough to measure the constructs [68]. The individual score of each factor is AI-enabled work culture 0.815 and for Work satisfaction 0.844.

11 Linear Regression

Simple linear regression is used to test the hypothesis as the study has one independent variable, AI-enabled work culture, and one dependent variable, work satisfaction. Table 6 shows that variable A has a significant and positive influence on the dependent variable as the significance value of our independent variable is 0.000, which is less than the alpha value 0.05 (p-value < 0.05), which means that our null hypothesis is rejected. Therefore, the research concludes that AI-enabled work culture significantly impacts employees' work satisfaction post-covid-19. The following simple linear regression equation will be formed based on SPSS output:

$$y = A + Bx$$

$$\text{Work satisfaction} = 6.642 + 0.529\ (\text{AI - enabled work culture})$$

Table 6 Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig	95.0% Confidence Interval for B	
		Work satisfaction	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	6.642	2.398		2.770	0.006	1.904	11.381
	AI-enabled work culture	0.529	0.057	0.607	9.292	0.000	0.417	0.642

Dependent Variable: Work satisfaction

Table 7 ANOVA

Model		Sum of squares	df	Mean square	F	Sig
1	Regression	1749.236	1	1749.236	86.341	0.000[b]
	Residual	2998.424	148	20.260		
	Total	4747.660	149			

[a]Dependent Variable: B
[b]Dependent Variable: B
A is AI enabled work culture
B is Work satisfaction

Table 8 Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the estimate
1	0.607[a]	0.368	0.364	4.501

[a]Predictors: (Constant), A
[b]Dependent Variable: B

Table 7 depicts that the significance value of the ANOVA table is 0.000, which is less than the alpha value of 0.05, which tells that there is a simultaneously significant relationship between both the dependent and independent variables.

Table 8 of the model summary shows the relationship between the dependent and independent variables. The R square is 0.368, which is moderate for the study and tells that a moderate variance relationship lies between the dependent and independent variables.

12 Conclusion and Implications of the Study

The future of work will not remain the same as it was before covid-19. There is a requirement for quick skill up-gradation for the employees to cope with this rapidly changing business environment because the companies are now focusing more on adopting AI in their work culture to sustain their businesses in this competitive market. However, this rapidly changing work culture is disturbing the employees due to the fear of getting replaced by AI and other smart machines. Moreover, the results of our study have concluded that AI-enabled work culture will significantly impact the employees' work satisfaction post-covid-19 which means that the increased use of AI in the organizations due to covid-19 influences the work satisfaction level of the employees.

The study will be relevant for the organizations who are thinking of turning their traditional work culture into a fully automated and AI-based work culture as it will have a direct implication on their employees, who are the essential assets of the organizations.

13 Limitations and Future Scope

The study can include more items in the scale to have a more accurate idea about the topic as the present study considers only 12 questions. Moreover, the study is a small-scale study with only 150 responses, and for the more accurate result on the given topic, the study can be conducted taking a larger sample size.

References

1. IBM (2020) COVID-19 and the future of business, IBM, Sep. 21, 2020. https://www.ibm.com/thought-leadership/institute-business-value/report/covid-19-future-business. Accessed 15 Jun 2021
2. Adesso (2019) Künstliche Intelligenz (KI) verändert den Blickwinkel, Adesso. https://ki.adesso.de/de/. Accessed 22 Oct 2021
3. Franken S, Wattenberg M (2019) The impact of AI on employment and organisation in the industrial working environment of the future
4. Howard A, Borenstein J (2021) AI, robots, and ethics in the age of COVID-19. In: MIT sloan management review. https://sloanreview.mit.edu/article/ai-robots-and-ethics-in-the-age-of-covid-19/. Accessed 22 Oct 2021
5. OECD (2021) Using artificial intelligence to help combat COVID-19. https://www.oecd.org/coronavirus/policy-responses/using-artificial-intelligence-to-help-combat-covid-19-ae4c5c21/. Accessed 22 Oct 2021
6. Coombs C (2020) Will COVID-19 be the tipping point for the intelligent automation of work? A review of the debate and implications for research. Int J Inf Manag 55:102182. https://doi.org/10.1016/j.ijinfomgt.2020.102182
7. Smith A, Janna, Anderson (2014) AI, robotics, and the future of jobs. Pew Research Center: Internet, Science & Tech,06 Aug 2014. https://www.pewresearch.org/internet/2014/08/06/future-of-jobs/. Accessed 22 Oct 2021
8. Walsh I, Kefi H, Baskerville R (2010) Managing culture creep: toward a strategic model of user IT culture. J Strateg Inf Syst 19(4):257–280. https://doi.org/10.1016/j.jsis.2010.09.002
9. Walsh I (2014) A strategic path to study IT use through users' IT culture and IT needs: a mixed-method grounded theory. J Strateg Inf Syst 23(2):146–173. https://doi.org/10.1016/j.jsis.2013.06.001
10. Abubakre M, Ravishankar MN, Coombs C (2017) Revisiting the trajectory of IT implementation in organizations: an IT culture perspective. Inf Technol People 30(3):562–579. https://doi.org/10.1108/ITP-09-2015-0217
11. Cyranoski D (2018) China enters the battle for AI talent. Nature 553(7688):260–261. https://doi.org/10.1038/d41586-018-00604-6
12. Bullock J, Luccioni A, Pham KH, Lam CSN, Luengo-Oroz M (2020) Mapping the landscape of artificial intelligence applications against COVID-19. J Artif Intell Res 69:807–845. https://doi.org/10.1613/jair.1.12162
13. Gambhir S, Malik SK, Kumar Y (2016) Role of soft computing approaches in healthcare domain: a mini-review. J Med Syst 40(12):287. https://doi.org/10.1007/s10916-016-0651-x
14. Piccialli F, di Cola VS, Giampaolo F, Cuomo S (2021) The role of artificial intelligence in fighting the COVID-19 pandemic. Inf Syst Front. https://doi.org/10.1007/s10796-021-10131-x
15. Razzaque A, Hamdan A (2021) Artificial Intelligence based multinational corporate model for EHR interoperability on an E-Health platform. In: Hassanien AE, Bhatnagar R, Darwish A (eds) Artificial Intelligence for sustainable development: theory, practice and future applications. Springer International Publishing, Cham, pp 71–81. https://doi.org/10.1007/978-3-030-51920-9_5

16. Ramesh AN, Kambhampati C, Monson JRT, Drew PJ (2004) Artificial intelligence in medicine. Ann R Coll Surg Engl 86(5):334–338. https://doi.org/10.1308/147870804290
17. Priddy KL, Keller PE (2005) Artificial neural networks. 1000 20th Street, Bellingham, WA 98227-0010 USA: SPIE, https://doi.org/10.1117/3.633187
18. Mullane MA, Bischofberger C (2020) Robots in the frontline of the fight against COVID-19. In: IEC e-tech, 15 Mar 2020. https://etech.iec.ch/issue/2020-02/robots-in-the-frontline-of-the-fight-against-covid-19. Accessed 27 Dec 2021
19. AL-Hashimi M, Hamdan A (2021) Artificial Intelligence and coronavirus COVID-19: applications, impact and future implications. Importance New Technol Entrep Bus Dev Context Econ Divers Dev Ctries 194:830. https://doi.org/10.1007/978-3-030-69221-6_64
20. Dizes K (2020) 3 ways Artificial Intelligence can strengthen your business (during COVID-19 and beyond), Engage Hub, 23 Nov 2020. https://engagehub.com/3-ways-artificial-intelligence-can-strengthen-your-business-during-covid-19-and-beyond/. Accessed 16 Jun 2021
21. Kaur N, Sahdev SL, Sharma M (2020) (PDF) Banking 4.0: the influence of artificial intelligence on the banking industry & how AI is changing the face of modern day Banks‖. ResearchGate, Jun 2020. https://doi.org/10.34218/IJM.11.6.2020.049
22. Anderson J, Bholat D, Gharbawi M, Thew O (2021) The impact of COVID-19 on artificial intelligence in banking | Bruegel. The impact of COVID-19 on artificial intelligence in banking, 15 Apr 2021. https://www.bruegel.org/2021/04/the-impact-of-covid-19-on-artificial-intelligence-in-banking/. Accessed 16 Jun 2021
23. World Economic Forum (2020) How digital innovations helped banks adapt during COVID-19. World Economic Forum, 20 Aug 2020. https://www.weforum.org/agenda/2020/08/how-digital-innovations-helped-banks-adapt-during-covid-19/. Accessed 16 Jun 2021
24. Sun Q, Zhang M, Mujumdar A (2018) Recent developments of artificial intelligence in drying of fresh food: a review. Crit Rev Food Sci Nutr 59:1–65. https://doi.org/10.1080/10408398.2018.1446900
25. Di Vaio A, Boccia F, Landriani L, Palladino R (2020) Artificial Intelligence in the agri-food system: rethinking sustainable business models in the COVID-19 scenario. Sustainability 12(12):4851. https://doi.org/10.3390/su12124851
26. Buhalis D (2019) Technology in tourism-from information communication technologies to eTourism and smart tourism towards ambient intelligence tourism: a perspective article. Tour Rev 75(1):267–272. https://doi.org/10.1108/TR-06-2019-0258
27. Revfine.com (2021) How to use Artificial Intelligence in the hospitality industry. Revfine.com, 21 Oct 2020. https://www.revfine.com/artificial-intelligence-hospitality-industry/. Accessed 22 Oct 2021
28. Webster C, Ivanov S (2019) Future tourism in a robot-based economy: a perspective article. Tour Rev, vol ahead-of-print, Sept 2019. https://doi.org/10.1108/TR-05-2019-0172
29. Zhou W (2021) The impact of AI on the hospitality industry. Medium, 09 Jul 2021. https://towardsdatascience.com/the-impact-of-ai-on-the-hospitality-industry-8ba7adb66be. Accessed 27 Dec 2021
30. Ortega-Parra A, Sastre Castillo M (2013) Impact of perceived corporate culture on organizational commitment. Manag Decis 51. https://doi.org/10.1108/MD-08-2012-0599
31. Flamholtz E, Randle Y (2012) Corporate culture, business models, competitive advantage, strategic assets and the bottom line: theoretical and measurement issues. J Hum Resour Costing Account 16. https://doi.org/10.1108/14013381211284227
32. Tedla T (2016) The impact of organizational culture on corporate performance. Walden Diss Dr Stud, Jan 2016 (Online). Available: https://scholarworks.waldenu.edu/dissertations/2509
33. Soni N, Sharma E, Singh N, Kapoor A (2020) Artificial Intelligence in business: from research and innovation to market deployment. Proc Comput Sci 167:2200–2210. https://doi.org/10.1016/j.procs.2020.03.272
34. Parry E, Battista V (2019) The impact of emerging technologies on work: a review of the evidence and implications for the human resource function. Emerald Open Res 1:5. https://doi.org/10.12688/emeraldopenres.12907.1

35. Hunt Scanlon Media (2021) How Artificial Intelligence is changing the culture of work. Hunt Scanlon Media, 22 Jun 2018. https://huntscanlon.com/how-artificial-intelligence-is-changing-the-culture-of-work/. Accessed 18 Jun 2021
36. Chiradeep (2019) 3 Ways AI is impacting company culture and ideas for AI-based hiring. HR Technologist, 29 Mar 2019. https://www.hrtechnologist.com/articles/culture/3-ways-ai-is-impacting-company-culture-and-ideas-for-ai-based-hiring/. Accessed 18 Jun 2021
37. Vantage C (2020) 6 ways AI in HR is changing today's work culture. Nurture an Engaged and Satisfied Workforce | Vantage Circle HR Blog, 29 Jan 2020. https://blog.vantagecircle.com/ai-in-hr/. Accessed 18 Jun 2021
38. Marquis EB, Kim S, Alahmad R, Pierce CS, Robert Jr LP (2018) Impacts of perceived behavior control and emotional labor on gig workers. In: Companion of the 2018 ACM conference on computer supported cooperative work and social computing, Jersey City NJ USA, Oct 2018, pp 241–244. https://doi.org/10.1145/3272973.3274065
39. Cardinal LB, Kreutzer M, Miller CC (2017) An aspirational view of organizational control research: re-invigorating empirical work to better meet the challenges of 21st century organizations. Acad Manag Ann 11(2):559–592. https://doi.org/10.5465/annals.2014.0086
40. Dennis A, Robert L, Curtis AM, Kowalczyk ST, Hasty BK (2012) Trust is in the eye of the beholder: a vignette study of post-event behavioral controls' effects on individual trust in virtual teams. Inf Syst Res ISR 23(2):546–558
41. Ouchi WG, Maguire MA (1975) Organizational control: two functions. Adm Sci Q 20(4):559–569. https://doi.org/10.2307/2392023
42. Hughes C, Robert L, Frady K, Arroyos A (2019) Artificial Intelligence, employee engagement, fairness, and job outcomes, pp 61–68. https://doi.org/10.1108/978-1-78973-077-720191005
43. Dabbous A, Aoun Barakat K, Merhej Sayegh M (2021) Enabling organizational use of artificial intelligence: an employee perspective. J Asia Bus Stud, vol Ahead-of-print, no. ahead-of-print, Jan 2021. https://doi.org/10.1108/JABS-09-2020-0372
44. Abdullah R, Fakieh B (2020) Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 22(5):e17620. https://doi.org/10.2196/17620
45. Dekker F, Salomons A, Waal J (2017) Fear of robots at work: the role of economic self-interest. Socio-Econ Rev 15. https://doi.org/10.1093/ser/mwx005
46. Plastino E, Purdy M (2018) Game-changing value from Artificial Intelligence: eight strategies. Strategy Leadersh 46(1):16–22. https://doi.org/10.1108/SL-11-2017-0106
47. Raj M, Seamans R (2019) Primer on artificial intelligence and robotics. J Organ Des 8. https://doi.org/10.1186/s41469-019-0050-0
48. Bhargava A, Bester M, Bolton L (2021) Employees' perceptions of the implementation of Robotics, Artificial Intelligence, and Automation (RAIA) on job satisfaction, job security, and employability. J Technol Behav Sci 6(1):106–113. https://doi.org/10.1007/s41347-020-00153-8
49. Webster C, Ivanov S (2019) Robotics, Artificial Intelligence, and the evolving nature of work, pp 127–143. https://doi.org/10.1007/978-3-030-08277-2_8
50. Schwabe H, Castellacci F (2020) Automation, workers' skills and job satisfaction. PLoS ONE 15(11):e0242929. https://doi.org/10.1371/journal.pone.0242929
51. Deloitte (2020) COVID-19—the upskilling imperative: building a future-ready workforce for the AI age, p 22
52. Manyika J, Sneader K (2018) AI, automation, and the future of work: ten things to solve for (Tech4Good) | McKinsey, 01 Jun 2018. https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for. Accessed 19 Jun 2021
53. Jöhnk J, Weißert M, Wyrtki K (2021) Ready or not, AI comes—an interview study of organizational AI readiness factors. Bus Inf Syst Eng 63(1):5–20. https://doi.org/10.1007/s12599-020-00676-7
54. Davenport T, Guha A, Grewal D, Bressgott T (2020) How artificial intelligence will change the future of marketing. J Acad Mark Sci 48(1):24–42. https://doi.org/10.1007/s11747-019-00696-0

55. ÓhÉigeartaigh SS, Whittlestone J, Liu Y, Zeng Y, Liu Z (2020) Overcoming barriers to cross-cultural cooperation in AI ethics and governance. Philos Technol 33(4):571–593. https://doi.org/10.1007/s13347-020-00402-x
56. Keding C (2021) Understanding the interplay of artificial intelligence and strategic management: four decades of research in review. Manag Rev Q 71(1):91–134. https://doi.org/10.1007/s11301-020-00181-x
57. Abbass HA (2019) Social integration of Artificial Intelligence: functions, automation allocation logic and human-autonomy trust. Cogn Comput 11(2):159–171. https://doi.org/10.1007/s12559-018-9619-0
58. Wamba-Taguimdje S-L, Fosso Wamba S, Kala Kamdjoug JR, Tchatchouang Wanko CE (2020) Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. Bus Process Manag J 26(7):1893–1924. https://doi.org/10.1108/BPMJ-10-2019-0411
59. Belhadi A, Mani V, Kamble SS, Khan SAR, Verma S (2021) Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism: an empirical investigation. Ann Oper Res. https://doi.org/10.1007/s10479-021-03956-x
60. Smids J, Nyholm S, Berkers H (2020) Robots in the workplace: a threat to—or opportunity for—meaningful work? Philos Technol 33(3):503–522. https://doi.org/10.1007/s13347-019-00377-4
61. Huang M-H, Rust RT (2018) Artificial Intelligence in service. J Serv Res 21(2):155–172. https://doi.org/10.1177/1094670517752459
62. Nam K, Dutt CS, Chathoth P, Daghfous A, Khan MS (2020) The adoption of artificial intelligence and robotics in the hotel industry: prospects and challenges. Electron Mark. https://doi.org/10.1007/s12525-020-00442-3
63. Mohapatra S, Kumar A (2019) Developing a framework for adopting artificial intelligence, vol 11, pp 19–22. https://doi.org/10.7763/IJCTE.2019.V11.1234
64. Radhakrishnan J, Chattopadhyay M (2020) (PDF) Determinants and barriers of artificial intelligence adoption—a literature review. ResearchGate, Dec 2020. https://www.researchgate.net/publication/347641074_Determinants_and_Barriers_of_Artificial_Intelligence_Adoption_-_A_Literature_Review. Accessed 21 Jun 2021
65. Field A (2005) Discovering statistics using SPSS, 2nd edn. Sage Publications, Inc, Thousand Oaks, CA, pp. xxxiv, 779
66. Tabachnick BG (2013) Using multivariate statistics, 6th edn. https://www.pearson.com/content/one-dot-com/one-dot-com/us/en/higher-education/program.html. Accessed 23 Dec 2021
67. Liao C, Tsou C, Huang M (2007) Factors influencing the usage of 3G mobile services in Taiwan. Online Inf Rev 31(6):759–774. https://doi.org/10.1108/14684520710841757
68. Cavana R, Delahaye B, Sekeran U (2021) Applied business research: qualitative and quantitative methods. John Wiley & Sons, Australia, Accessed 22 Oct 2021 (Online). Available: https://eprints.qut.edu.au/10523/

An Intellectual Detection and Multiclassification of Anomalies Applying BLPMA System

T. Dhikhi and M. S. Saravanan

Abstract Network infrastructure attacks are the biggest threat to network and information security threats. The massive rise in internet connectivity and accessibility has affected the frequency of unauthorized attackers engaged in illicit activities around the world. As a result of the impressive growth of web applications, the need for information security in a network has inflated. Several methods of intrusion detection are applied to detect irregularities that depend on precision, detection frequency, other parameters and are anticipated to familiarize to vigorously varying risk scenes. In this study, a novel method of detection and classification of the anomaly-based IDS is proposed for the effective detection of intrusions. The data is preprocessed, and the appropriate features are carefully selected from the data collection. A methodology based on feature ranking and subset selection are used to extract relevant attributes from high-dimensional and persistent network data, thus enhancing the detection efficiency of the IDS and its ability to adapt to the current dynamic and multifaceted network environment. Further, an improved classification method based on Bayesian theorem and probabilistic neural network is aimed at classifying the attacks in the dataset. A two-level classifier is ensembled for binary and five class classification model. Data is normalized and discretized by applying the Min–Max Specific Equivalent Frequency Discretization Algorithm to discrete features that help to minimize memory consumption and simplify the representation of data that makes data retrieval faster.

Keywords Cognitive principal data gain feature selection · Specific equivalent discretization method · Bayesian likelihood procurement maximization

T. Dhikhi
SIMATS, SRMIST Ramapuram, Chennai, India

M. S. Saravanan (✉)
Saveetha School of Engineering, SIMATS, Chennai, India
e-mail: saranenadu@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_22

1 Introduction

Communication networks and knowledge exchanges in today's world are a vital factor of the recent society. Whenever such communication has been carried out in a network safety and security is a major concern. In a complex network world, several advanced security policies and mechanisms are constantly researched, planned, and enforced to ensure the integrity, availability, and confidentiality of data for authentic handlers. Different methods are used to protect the computers at risk, by using traditional security strategies like protecting the machine from attacks using firewalls, authentication strategies and access rights or by evading attacks by encryption and detecting intrusions at the accurate period by means of an active and effectual intrusion detection system. The hazard of intrusion has also become a grave threat for the corporate world. Network attacks occur regularly, due to the complexity of the network system and the variety of methods of attack. Therefore, to avoid the increasing hacker technology, network attack detection means should be smarter and more successful than ever before. Nowadays malicious users are smart and advanced enough to constantly monitor and analyse networks to obtain illegal access to delicate data and information sources. Therefore, it is always important to implement improved detection mechanisms capable of analysing user behaviours and organizing those certain activities are legitimate or malicious in an environment of real time. IDSs play an important role in network monitoring and have generally been used in recent years as a network security element. In addition, it helps to detect intrusions by observing the mechanism and intrusion activity characteristics, thereby allowing for a real-time response to intrusion events and the invasion process. IDSs play an important role in network monitoring and have generally been used in recent years as a network security element. In addition, it helps to detect intrusions by observing the mechanism and intrusion activity characteristics, thereby allowing for a real-time response to intrusion events and the invasion process. An intrusion is characterized as any kind of action taken by an illegal user to compromise security aspects such as availability, integrity, confidentiality, accountability, and assurances [1]. Network Intrusion Detection System (NIDS) is a passive system that prevails over an organization's device or network and monitors network traffic to detect attacks. NIDS provides network security based on information about network traffic such as source address, destination address, message size, and length [2]. Network-based IDS systems are more persuaded than host-based systems to be self-reliant. Huge networks can be managed by the installation of certain excellent network engineered devices and is not interrupted when setting up ongoing network operations, as they are passive tools. It cannot inspect encrypted packets, generate some of the traffic that is not detectable by the method and decrease NIDS efficiency. An anomaly IDS (AIDS) will do identification by matching data with standard data profiles taken for analysis. If any variation is initiated, the analysed data is alerted as abnormal. This method detects novel or unknown device attacks based on deviations from the usual use of programs regardless of a basis is a fortunate internal user, or an unlawful external user. Anomaly IDS applies statistical analysis by observing the traffic which is identified to be normal, and a possible reference point is emerged. The key anomaly detection skill is the ability to discover unknown attacks [3].

2 Related Study

Hajisalem and Babaie [4] developed a composite classifying technique on the basis of Artificial Fish Swarm and Artificial Bee Colony for anomaly recognition and then it is trained by means of generated rules for misuse detection. The training dataset is segregated into clusters applying Fuzzy C-Means Clustering and insignificant features are eliminated using Correlation-based Feature Selection procedures. In accordance with the chosen features, rules are generated using CART and those features are categorized as normal and unusual. Computational difficulty study and time cost investigation reveal the complexity of the given process is contrast to equivalent methods. Kaur and Singh [5] developed a deep learning method for hybrid IDS in which it generates a signature of unknown web attacks. D-Sign is competent of effectually noticing and producing high precision, sensitivity, and specificity of attack signatures. This was for the identification of attacks and the genesis of web-based assaults with signatures. Thus, the resultant approach does indeed create a signature to closely distinguish attacks that recorded relatively minimal false positives along with false negatives. Alhakami et al. [6] established a model a detection paradigm for known as well as unknown invasions by means of an emerging Bayesian model by learning patterns of the activities. A newly Bayesian-based MCMC inference for infinite bounded generalized Gaussian mixture models to evade over and under fitting to validate uncertainty through probability. In this approach the number of clusters need not be defined, in view of the ambiguity by appending prior awareness of the model parameters.

Teng et al. [7] suggested in building powerful and active framework for detection for anomalies and introduced a E-CARGO as a tool in which E-C stands for environments-classes, A stands for agents, R stands for roles, G and O for groups, and objects respectively. Further apply SVM Classifiers and decision tree algorithm. In these four different kinds of organized and deployed SVM identification functions for which associated agents are configured with unique characteristics for identifying attacks towards TCP, UDP, ICMP as well as application-level protocols correspondingly. Outcomes demonstrate that the optimized model for the detection relying on binary class SVMs and DTs seems to be more precise and reliable than those of the detection design with a collection of single form SVMs. Saleh et al. [8] designed an IDS that blended different means to adjudicate multiclass classification problem that needs a Naïve Bayes feature selection to shorten the spatiality of specimen data and also maximize the detection rate whilst curtailing RMS error. Not only does NBFS check the impact of each individual feature nevertheless, the joint effect of each of the two-feature pair is also taken into account. Further an Optimized Support Vector Machines, OSVM is utilized for refusing outliers that are noisy data that leads to false classification. Tama et al. [9] recommended a better IDS based on hybrid feature selection and two-level classifier composites. This feature selection exploits the reduction of feature size by a combination of three different methods particle swarm optimization, ant colony algorithm, and genetic algorithm that is applied in two different datasets viz NSL-KDD and UNSW-NB15. The features are chosen in accordance with the results of a reduced error pruning tree, REPT classifier for classification. Subsequently a two-level classifier group is planned with reference to two meta learners such as rotational forest and bagging. Finally, a two-fold statistical significance check is completed to validate the findings. Consequently, provides added value for the experimental outcomes obtained as a result of this particular classifier.

Gao et al. [10] introduced a versatile ensemble model that merges the merits of respective algorithm for various types of data detection and achieves optimum results through group learning. The advantage of integrating the predictions of multiple base surveyors to boost quality of being generalizable and robustness by means of a single estimator. Moreover, MultiTree and adaptive voting algorithm are compared with other methods which expand consequence of detecting anomalies and found superior amongst others. The ensemble learning process has enhanced the detection effect and shown effectual increase in the accuracy of detection. Yin et al. [11] studied a better structural model of the upgraded k-dependency Bayesian network (KDBN) that can reliably define the relationships of dependency amongst system parameters and lessen the difficulty of the Bayesian network framework by diminishing the directed edges of weak dependence. Investigation is done on the accuracy of detecting attacks with three IDCMs viz NB, KDBN, improved KDBN and concluded minimum detection time and higher efficiency for the later one. Additionally, stabilize the small classes of U2R and R2L attacks. Davis et al. [12] presented autoencoders to find out potential feature groups and are skilled to learn consistently the resemblance amongst input features. A DNN framework is defined by merging an unsupervised platform for multiple channel feature learning and a supervised level that exploits cross-channel feature dependencies. In the former stage double autoencoders are studied individually for normal and abnormal. Since the decoder of the autoencoder restores the sample as the input, two new feature vectors could be defined allowing each network flow to be represented as a multichannel sample. In regard to supervised level, the impact of one channel on the other is studied by taking a multichannel parametric convolution. The efficiency of the solution is calculated by recording the autoencoder residual error besides the accuracy and F-score of the models for intrusion detection.

An effective IDS should be able to detect different kinds of attacks accurately including intrusions that incorporate evasion techniques. Developing IDSs capable of overcoming the evasion techniques remains a major challenge for this area of research. For a better result of intrusion detection, a collaborative machine learning approach can be implemented for identifying the novel attacks. This type of system combines two or more intrusion detection systems methodologies to analyse, detect, and match any suspicious behaviour and signature malicious code that attempts to attack network. The power of combination means it can detect more types of intrusion, thereby providing relatively better results as compared to other methods. If the selected classifier output is totally incorrect, the real outcome might also be totally incorrect for a single classifier. Furthermore, the trained classifier may well not be able to handle the problem sufficiently. Ensemble classifier has greater accuracy over single technique of classification. The combination of several trained classifiers thus results in a better performance than any single classifier.

3 Proposed Work

The intended study is designed to incorporate trusted unsupervised and supervised machine learning algorithm to evaluate the accuracy and effectiveness of the process. The scheme is developed with the intention of categorizing novel threats with better detection rates, low classification error levels and therefore less computation complexity in a network environment. The inclusive framework of this intrusion detection and multiclassification system is shown in Fig. 1 with distinct stages to find the anomalies in a network. The architecture involves significant integrant namely dataset, user interface segment, data cleaning/pre-processing segment, feature selection segment, detection and classification of intrusions segment including classifier and different evaluation parameters.

1. Choose the dataset NSLKDD for training and testing.
2. Data cleaning is done for the features of NSLKDD that involves numericalization to convert data into numeric values followed by normalization. Then a discretization scheme is applied that do not require class information to improve the classification model.
3. Apply feature selection algorithm using machine learning algorithm, that applies a feature sub selection and feature ranking followed by a feature selection using autoencoder that requires defining the parameters used for a more accurate representation of a machine. This method helps to enhance the classification process's accuracy and efficiency by choosing only the relevant terms and removing the noisy terms.
4. Building an effective classifier using two different models for binary and multiclassification and further these algorithms are combined to get an improved five class classification model. The classification of the training is considered as a learning phase to construct a classifier or model. The classification is applied on test data to determine the performance parameters such as accuracy, precision, and recall. In this binary category classification for normal and anomaly as well as five class categories for attacks and normal data classification are done.

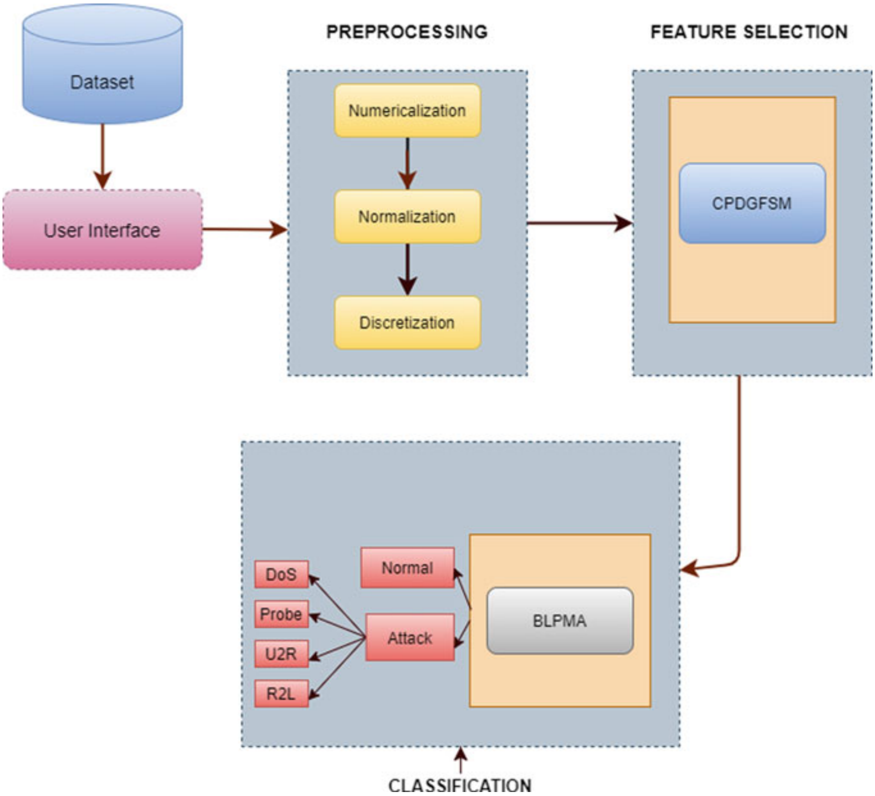


Fig. 1 Intrusion detection and multiclassification system

3.1 *Min–Max Specific Equivalent Discretization Method for Data Cleaning*

Data Cleaning were found to depend primarily on skilled domain experts to define the most essential factor of network activity and to create the actual collection of traffic characteristics. In addition, automated methods have been used to extract features to minimize the data dimensionality and collection of features to identify the most important features from that whole candidate set [12]. Phase of pre-processing minimizes data to the maximum extent without information loss and needs specialized scheduling, planning, and evaluation. NSL-KDD dataset includes numeric attributes as well as few non-numeric ones. Non-numeric attributes such as protocol type, service, and flag attributes would require translating into numerals since numeric matrix for giving the training and testing data input for the next stage. There are three (protocol-type, service, and flag) and 38 non-numeric features and numeric features in the NSL. Therefore, on non-numeric features, a 1-n encoding scheme is applied [13]. Finally, the 41-dimensional features are translated into 122 numeric features. Min–max normalization is employed thus maps the exiting range of values to two new range values. Min–Max Normalization conducts a linear transformation into the specified interval on the original data x using the Eq. 1. This approach scales the data proportionally from $(\min_A, \max_A)$ to $(new\_\min_A, new\_\max_A)$. The benefit of this approach has been that it retains the exact values of all the data relationships. This does not introduce potential prejudices into the data.

$$x_i^{'} = \frac{x_i - \min_A}{\max_A - \min_A}\left(new\_\max_A - new\_\min_A\right) + new\_\min_A \tag{1}$$

where xi denotes the data point, $\min_A$ is the minimal value for all data and $\max_A$ is the upper limit value for all data factors. $new\_\min_A$ and $new\_\max_A$ are the newly mapped minimum and maximum values, respectively [14]. Specific Equivalent Frequency Discretization Algorithm is the method for transforming the continuous domain features into a nominal space. This method leads to the generation of a discrete scheme D for a given continuous attribute F [15]. In this method, the input is given as a continuous attribute value. The input attributes sorted in ascending order to find the intervals initially. Subsequently, calculate the value m as shown in Eq. 2. Then the number of intervals k is calculated from the value m, which is the nearest integer value of the square root m using Eq. 3. Then bins are created by the number of attributes n divided by the number of intervals k. After finding the bins, the boundary of each interval is estimated by adding elements in the current bin and next bin, and this value is divided by the total number of elements in both bins. Finally, the continuous value of the attribute is converted to a discrete one by determining the range to which it belongs. The steps of the discretization method are outlined in algorithm 3.1.

$$m = \frac{n+1}{2} \tag{2}$$

$$k \approx \sqrt{m} \tag{3}$$

In this discretization method, the need for user input for the number of intervals is eradicated. Rather, calculate the interval limits to improve the effectiveness of other discretization methods [16]. This method is an unsupervised discretization method because it does not require information on the label class. Besides, it does not need a learning model that makes this form a static one. The processing time of the mechanism is $O(n)$.

Algorithm 1. Specific Equivalent Frequency Discretization Algorithm (SEFDA)

1. Input the continuous value of attribute A and n is the number of attributes
2. Sort all the values of A in increasing order.

3. Find $m = (n + 1)/2$
4. Calculate number of intervals k $\approx \sqrt{m}$
5. Generate bins at each interval with the number of elements n divided by the number of intervals, n/k
6. Borders of each interval are calculated by adding elements of the current bin and next bin and then divided by total number of elements in both bins.
7. By identifying the frequency to which they pertain, the continuous value attributes are converted into discrete value.

3.2 Cognitive Principal Data Gain Feature Selection Method

A feature is defined as pertinent when its values differ systematically with the values of the class attribute. Composite classification methods often include feature selection, applying a single irrelevant feature to the ML process can make inaccurate comparison and restricts the detection process. Feature subset selection methods provide maximum reduced feature set depending on the algorithms for the best feature subset search without ranking the features [17]. The purpose of selecting a feature is flexible, i.e., improving classifier detection accuracy, timeliness, and cost-effectiveness of classifiers, and awareness of the ML process that obtained the predictions. Each feature does have a distinctive contribution to anticipate the type of attack based on the amount of information a feature holds to identify a specific attack. This is built on converting a considerable number of variables into a smaller uncorrelated variable by seeking a few linear orthogonal combinations. Next, the data is analysed in advance to normalize the mean and variance. Thereby, to formulate the step of reduction, the covariance matrix, Eigen vectors, and Eigen values are determined [18]. PCA is designed to remove overlap in the features and to solve measurable relationships in the new space termed principal space. This is a very renowned statistical concept that applies to large-dimensional data in a variety of applications such as image compression, facial recognition, and pattern detection. For a set of features the task is to map the data and figure out the key components. It could be achieved by minimizing the variance and raising the reconstruction error in the least square. PCA intended to identify the highest variance subspace in the M-dimensional space of the feature. The actual characteristics are converted uniformly into the principal components (PCs) using eigen vectors. This phase is important to rank the feature subset dependent on the information entropy in descending order. For a feature to be relevant, it should not be reliant on the input data and can be reliant on the class labels. The feature that has no impact on class labels can be rejected. The primary peculiarity is that the features are ranked that have an impact on the labels of the class. Depending on this ranking property the top ranked features are mostly important and meaningful for warning categories. Feature ranking is applied using information gain and computed using Eq. (4) by equations entropy of class (5), entropy of feature (6). This is calculated on the basis of the entropy or disorder of the system. The information gain is figured for each subset of the feature acquired from the feature subset selection method by measuring the relevancy of it to the class label. The entropy of the class labels is measured, and the entropies of the feature value are lessened from it. Amongst those instances of feature value, a count is allocated for number of times for each class. This is repeated for all possible values of feature. A matrix in Eq. (7) is reserved to count and tally the class labelled by features. This ranking method is a proportionate evaluating process. The information obtained for y monitoring x by Eq. (8) and vice versa are equal [19].

$$\mathrm{x} = \mathrm{HC} - \mathrm{H(C|A)} \tag{4}$$

$$H_C = \sum_{c\varepsilon \mathrm{C}} p(c[i]) * \log_2(p(c[i])) \tag{5}$$

$$\mathrm{H\_A} = \sum \_a\varepsilon\mathrm{A}\ \mathrm{p(a[i])} * \mathrm{log\_\ 2\ (p(a[i]))} \tag{6}$$

$$M = z + p(c[i]|a[j]) * \log_2(p(c[i]|a[i])) \tag{7}$$

$$z \leftarrow M \tag{8}$$

Feature ranking will assign score for each feature and is given ranks according to the descendant order and can choose the high ranked ones amongst them and discard rest of the features [20]. Gain Factor in Eq. (9) is the enhancement of information gain by extending a standardization to information gain by dividing the dataset D into m partitions equivalent to the test on attribute with the help of partition information in Eq. (10).

$$partitiondata_D = -\sum_{j=1}^{m}\left[\frac{freq\left(c_j, D\right)}{|D|}\right]\log_2\left[\frac{freq\left(c_j, D\right)}{|D|}\right] \tag{9}$$

$$\mathrm{x} = \mathrm{HC} - \mathrm{H(C|A)} \tag{10}$$

Algorithm 2. Cognitive Principal Data Gain Feature Selection Method (CPDGFSM)

1. Procure the features in the preprocessed NSL KDD dataset as the input process for feature upgradation.

2. Calculate Mean and subtract the mean from every corresponding input variable from the training set to find the standard deviation.
3. Compute the covariance matrix after calculating the transpose of standard deviation.
4. Generate Eigen vectors and Eigen values of the covariance matrix.
5. Select ς eigen vectors equivalent to choose the features that are more relevant called principal data.
6. Initialize class labels and attribute values C and A respectively, Dataset D and $z = 0$.
7. For every c in class labels C, find $p(c[i])$
8. Estimate $H_C = \sum_{c\varepsilon C} p(c[i]) * \log_2(p(c[i]))$.
9. For every 'a' in A, determine $p(a[i])$
10. Compute $H_A = \sum_{a\varepsilon A} p(a[i]) * \log_2(p(a[i]))$
11. For every value of a and class c, calculate $p(c[i] \mid a[i])$
12. $M = z + p(c[i]|a[j]) * \log_2(p(c[i]|a[i]))$
13. $z \leftarrow M$
14. $\mathrm{H(C|A) = -HA * M}$
15. $\mathrm{X = HC - H(C|A)}$
16. Find $partitiondata_D = -\sum_{j=1}^{m}\left[\frac{freq(c_j, D)}{|D|}\right]\log_2\left[\frac{freq(c_j, D)}{|D|}\right]$
17. Compute gain factor $GF(A) = \frac{x}{partitiondata_D}$

3.3 Bayesian Likelihood Procurement Maximization Algorithm

The Bayesian Likelihood Procurement Maximization (BLPMA) classification algorithm aims to identify the data as regular or anomaly, and the abnormal data that are attacks are further classified into subcategories such as DoS, Probe, U2R, and R2L. This algorithm uses the Bayesian algorithm principle and that the probability can be determined accordingly. Bayes theorem offers a mode to evaluate posterior probability. Initially this algorithm finds the parameters needed for the function of probability density μ and σ. The posterior probability is the upgraded probability of an occurrence happening after considering novel information and achieved by revising the prior likelihood. This is the likelihood that an occurrence can happen provided that another occurrence has taken place. The posterior probability is depicted in Eq. (11). This can be a good indicator of the important belief of a data generating process than the prior probability, since more information was encompassed in the posterior. When new knowledge emerges and is integrated into the analysis, a posterior probability will subsequently become a prior for a new modified posterior probability. The probability calculation, posterior probability is done using the Eq. (12) in which $p(c/a)$ is the posterior probability of target or class, c considering a. a is data record and c that represents data a. $p(c)$ is prior probability of target. $P(a/c)$ is the likelihood i.e., the probability of predictor given class. $P(a)$ is prior probability of predictor. This calculation can be achieved mainly by making a table of occurrence against the target for each attribute. Then these tables must be converted into yet another likelihood table to eventually apply the formula to find a posterior probability for the entire class.

$$posteriorprobability = \frac{likelihood \times prior}{evidence} \tag{11}$$

$$p(c/a) = \frac{p(a/c) \times p(c)}{p(a)} \tag{12}$$

The likelihood is calculated using the Eq. (13). To estimate every term in likelihood equation conditional probability is specified by Eq. (14) for $a = 1$ to n. Then maximum likelihood is estimated for μ and σ by taking two derivatives of the Eq. (15). One derivative is in relation to μ by considering σ as constant and also the maximum likelihood is computed for μ by determining at which point the derivative is 0. Further derivative to be relative to σ, addressing μ to be constant and then calculating maximum likelihood of σ by finding the point where derivative is 0. To get derivatives, log of the likelihood function is taken for constructing derivative easier as in Eq. (16). Applying log to do lot of transformations to this function. This log changes the distinct likelihood functions and finally will produce the Eq. (14).

$$p(a|c) = p(a_1|c) \times p(a_2|c) \times \cdots \times p(a_n|c) \tag{13}$$

$$p(a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(a-\mu)^2/2\sigma^2} \tag{14}$$

$$ln[p(a_1, \ldots, a_n)] = ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(a_1-\mu^2\right)/2\sigma^2} \times \cdots \times \frac{1}{\sqrt{\sqrt{2\pi\sigma^2}}} e^{-\left(a_2-\mu^2\right)/2\sigma^2}\right) \tag{15}$$

$$\begin{aligned} \ln[p(a_1 \ldots a_n)] = n[p(a_1, \ldots, a_n)] - \frac{1}{2}\ln(2\pi) - \ln(\sigma) \\ - \frac{(a_1 - \mu)^2}{2\sigma^2} - \cdots - \frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{(a_n - \mu)^2}{2\sigma^2} \end{aligned} \tag{16}$$

The above equation of likelihood is further simplified by combining first term with remaining $n - 1$ occurrences for n datapoints. After simplification the log of the likelihood function is depicted in Eq. (17) and the derivative of the same is found with respect to μ which is the slope function for the log curve. The first and second term in the equation doesn't contain μ so derivative of those are zero. As the numerator has μ, chain rule is applied to find the derivative and the denominator remains same because σ is constant. After which the derivative of third term will become $\frac{(a_1-\mu)}{\sigma^2}$, further the same rule is applied for the remaining terms and simplified to get the derivative of log likelihood function with respect to μ is as shown in Eq. (18).

$$\ln[p(a_1 \ldots a_n)] = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{(a_1-\mu)^2}{2\sigma^2} - \cdots - \frac{(a_n-\mu)^2}{2\sigma^2} \tag{17}$$

$$\frac{\partial}{\partial\mu}\ln[p(a_1 \ldots a_n)] = \frac{1}{\sigma^2}[(a_1+a_2) - n\mu] \tag{18}$$

Following the derivative of log likelihood function with respect to σ is formed from Eq. (19). The first term doesn't contain σ so its derivative is zero and second term is n over σ. The derivative of next term is $\frac{(a_1-\mu)^2}{\sigma^3}$ and similarly the remaining terms are done and simplified to Eq. (19). In order to find the maximum likelihood calculation for μ and σ, the derivative must be solved with respect to $\mu = 0$ and $\sigma = 0$ respectively since the slope is zero at the peak curve point for both cases. For this initially set the derivative with respect to μ is zero and solve for μ. Thus, the maximum likelihood measurement for μ is the mean of the measurements as in Eq. (20) and is where the centre of the normal curve is. Next set the derivative with respect to σ to zero and resolve for σ. The maximum likelihood for measurement for σ is in Eq. (21). The prior probability of class or goal is to be determined regardless of the data. In addition, a prior likelihood of the data can also be calculated in the same way. These values are added in the Eq. (12) to determine the posterior likelihood for each attribute. Ultimately, prediction of classes is dependent as to which class really does the nearest predicted value is.

$$\frac{\partial}{\partial\sigma}\ln[p(a_1 \ldots a_n)] = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left[(a_1-\mu)^2 + \cdots + (a_n-\mu)^2\right] \tag{19}$$

$$\mu = \frac{a_1 + a_2 + \cdots a_n}{n} \tag{20}$$

$$\sigma = \sqrt{\frac{(a_1-\mu)^2 + \cdots + (a_n-\mu)^2}{n}} \tag{21}$$

The probability density can be chosen as the class of Gaussian mixtures as shown in the Eq. (22), θ represents the parameter vector $(k_i, \mu_i, \sigma_i)_{i=1}^n$ and $k_i \geq 0$ and $\sum_{i=1}^n k_i = 1$ is applied. The $N(a|\mu_i, \sigma_i)$ in the Eq. (22) are multivariate normal densities and is given in Eq. (23)

$$p(a|\theta) = \sum_{i=1}^{n} k_i N(a|\mu_i, \sigma_i) \tag{22}$$

$$N(a|\mu_i, \sigma_i) = (2\pi)^{-\frac{\mathrm{d}}{2}}|\sigma_\mathrm{i}|^{-\frac{1}{2}}\exp\left[\frac{-1}{2}(\mathrm{a}-\mu_\mathrm{i})^\mathrm{t}\sum_\mathrm{i}^{-1}(\mathrm{a}-\mu_\mathrm{i})\right] \tag{23}$$

Then maximum likelihood is estimated using Expectation Maximization algorithm that contains mainly two steps E-step and M-step. EM algorithm is widely used to simplify the difficult calculation of maximum likelihood problems often found in mixture models and that could not be numerically solved [17]. Firstly, the posterior probability that component *i*, is responsible for the generation of a^k is viewed as Eq. (24) and new parameters are obtained in the next M-step as depicted in Eqs. (25), (26), (27). Thus, the maximum likelihood for the probability function is derived and is applied in the likelihood equation in Eq. (24). Finally, the Eqs. (23) and (24) after the calculations are further given in the Eq. (22) to get the probability of a particular feature for classification.

$$h_i^k = \frac{k_i N\left(a^k \middle| \mu_i, \sigma_i\right)}{\sum_{j=1}^{n} k_j N\left(a^k \middle| \mu_j, \sigma_j\right)} \tag{24}$$

$$k_i' = \frac{1}{m} \sum_{k=1}^{m} h_i^k \tag{25}$$

$$\mu_i' = \frac{\sum_{k=1}^{m} h_i^k a^k}{\sum_{l=1}^{m} h_i^l} \tag{26}$$

$$\sigma_i' = \frac{\sum_{k=1}^{m} h_i^k \left(a^k - \mu_i'\right)\left(a^k - \mu_i'\right)^t}{\sum_{l=1}^{m} h_i^l} \tag{27}$$

A chaining rule applies to the posterior probability in which the likelihood for the second test from the first test becomes the prior one. Chaining rule is useful in updating probabilities as data enters.

Algorithm 3. Bayesian Likelihood Procurement Maximization Algorithm

1. Read the features in the training dataset after feature selection.
2. Calculate the parameters required for probability density function such as mean and standard deviation of the predictor variable in each class using Eqs. (20) and (21) which is the maximum likelihood estimation for normal distribution.
3. The μ and σ from the above step are given as the input for expectation and maximization algorithm to find the updated parameter values of μ_i' and σ_i' using the Eqs. (26) and (27).
4. Compute the probability p(a) using Eq. (14) in each class by using parameters in Eqs. (26) and (27).
5. Determine the likelihood to be given in Eq. (12) using the Eq. (13).
6. Estimate prior probability of class(target) irrespective of data as well as prior probability of data despite class.
7. Employ all calculations in Eq. (12) to obtain the posterior probability.

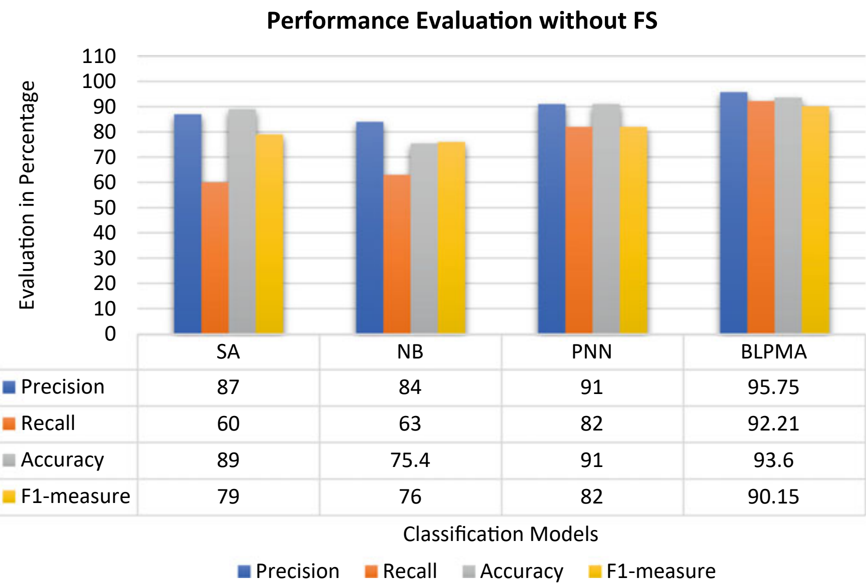
8. Apply chain rule to the posterior likelihood and essentially the class determination is defined on which class does have the closest expected value.

4 Results and Discussions

In the research, different classification algorithms viz. Sparse Autoencoder (SA), Naïve Bayes (NB), Probabilistic Neural Networks (PNN), Bayesian Likelihood Procurement Maximization Algorithm (BLPMA) without employing feature selection technique are compared with different parameters such as precision, recall, accuracy, F1-measure. These classification techniques are further applied with the feature selection method and the performance of the respective algorithms are compared. Then the study is done on the same performance metrics using different classification methods. The performance is assessed using the multiclassification method Bayesian Likelihood Probabilistic Maximization Algorithm without feature selection, followed by feature selection inclusion. The Bayesian Likelihood Procurement Maximization Algorithm is used to classify data into normal, DoS, Probe, user to root, and remote to local attacks. The analysis of different performance metrics for the BLPMA classification without feature selection, and analysis with feature selection is made. The results show that when used in conjunction with feature selection, the multiclassification method produces a better result.

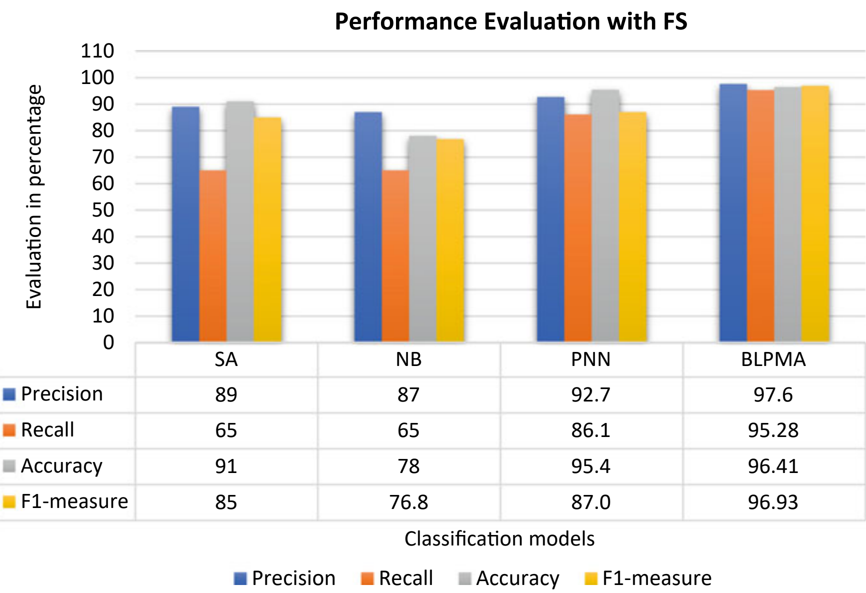
The performance analysis of feature selection is shown in Table 1. It displays the total positive rate, total negative rate, and the accuracy of the feature selection. Figure 2 depicts the performance evaluation of the various proposed classification methods without the use of a combined feature selection technique. The comparison graph of the proposed classification method with the feature selection methodology is shown in Fig. 3. The performance evaluation of the BLPMA for the metrics accuracy, precision, recall, and F1-measure is then plotted in Figs. 4 and 5 respectively, without and with feature selection. The performance evaluation of the classification model shows a better performance in all methods. The feature selection approach used in classification model applies an enhanced performance.

The parameters used to evaluate the effectiveness of IDS are Total Positive Rate (TPR), Total Negative Rate (TNR), Accuracy, Precision, Recall, F1-measure. TPR gives probability that an actual positive will test positive, TNR produces the probability that an actual negative will test negative, accuracy is the proportion of the total number of the correct predictions to the actual data set size, precision is the fraction of data instances predicted as positive that are actually positive, recall is the proportion of correctly predicted attack cases to the actual size of the attack class, F1-measure is the harmonic mean of the precision and recall at given threshold. The classification model BLPMA produces 95.75%, 92.21%, 93.6%, 90.15% of precision, recall, accuracy, and F1-measure respectively without feature selection CPDGSAFST without



	SA	NB	PNN	BLPMA
Precision	87	84	91	95.75
Recall	60	63	82	92.21
Accuracy	89	75.4	91	93.6
F1-measure	79	76	82	90.15

Fig. 2 Comparison of Classification models without feature selection



	SA	NB	PNN	BLPMA
Precision	89	87	92.7	97.6
Recall	65	65	86.1	95.28
Accuracy	91	78	95.4	96.41
F1-measure	85	76.8	87.0	96.93

Fig. 3 Comparison of Classification models with feature selection

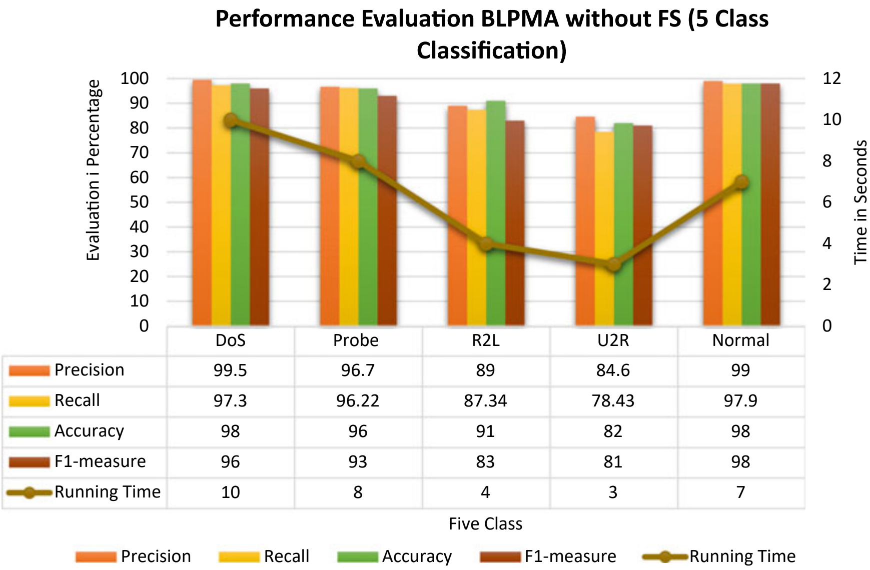


Fig. 4 Evaluation of BLPMA model without Feature selection

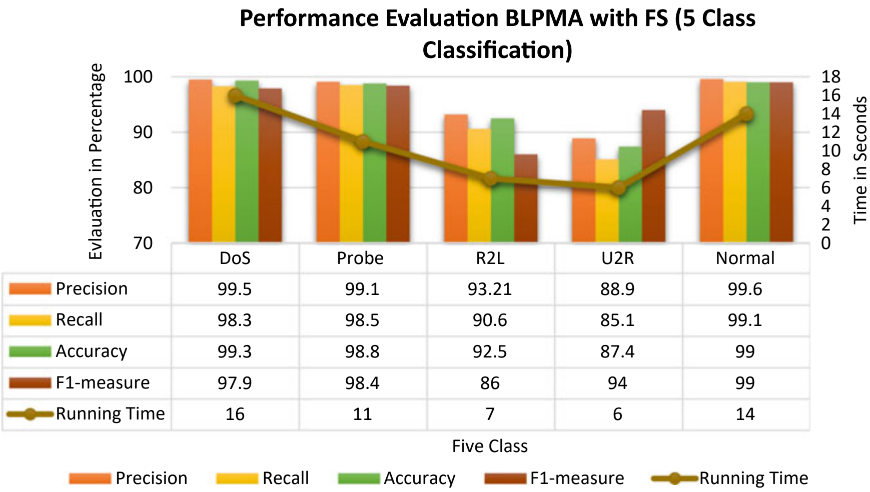


Fig. 5 Evaluation of BLPMA model with Feature selection

feature selection CPDGFS. Applying the feature CPDGFS, the classification model BLPMA achieves 97.6%, 95.28%, 96.41%, 96.43% of precision, recall, accuracy, and F1- measure respectively (Table 1).

Table 1 Performance analysis of feature selection

FS method	TPR	TNR	Accuracy
PCA	0.9259	0.9012	0.9136
CPDGFSM	0.9565	0.9444	0.950
SA	0.9583	0.9565	0.9374

5 Conclusion

The rapid growth in Internet use and the generation of vast quantities of useful digital data draws the attacker to the illicit achievement of economic benefits, and so on. The research analyses the deep learning algorithms used for IDS in different environments and has also done some experiment on the NSL-KDD data set using deep learning algorithms, and the results are also available compared to that. The subset of important features is selected, and few features are rejected that are unacceptable and redundant. This implies feature subset selection followed by a feature ranking and finally a feature selection that surges valuation or accuracy rate of classification. The input to the classification is the selected feature vectors that apply intellectual multiclassification method that provides an enhanced solution for five class classification. The classification method that includes Bayesian Likelihood Procurement Maximization for five class classification. This study shows that the detection rate, the false positive rate, accuracy, and precision are higher when compared to prevailing algorithms. The result is found to be better with accuracy of 95.28%, precision of 97.6%, recall of 96.4%, F1-measure of 96.93%.

References

1. Stoneburner G (2001) Underlying technical models for information technology security: recommendation of the national institute of standards and technology, US Department of Commerce, Computer Security Division, Information Technology, National Institute of Standards And Technology
2. Thottan M, Ji C (2003) Anomaly detection in IP networks. IEEE Trans Signal Process 51(8):2191–2204
3. Li Z (2011) Anomaly intrusion detection method based on K-means clustering algorithm with particle swarm optimization.vol 4, issue 2. Springer, Berlin (2011)
4. Hajisalem V, Babaie S (2018) A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection. Comput Netw 136:37–50
5. Kaur S, Singh M (2019) Hybrid intrusion detection and signature generation using deep recurrent neural networks. Neural Comput Applic
6. Alhakami W, ALharbi A, Bourouis S, Alroobaea R, Bouguila N (2019) Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection. IEEE Access 7:52181–52190
7. Teng S, Wu N, Zhu H, Teng L, Zhang W (2018) SVM-DT-based adaptive and collaborative intrusion detection. IEEE/CAA J Automatica Sinica 5(1):108–118
8. Saleh AI, Talaat FM, Labib LM (2019) A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers. Artif Intell Rev 51:403–443

9. Tama BA, Comuzzi M, Rhee K (2019) TSE-IDS: a two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. IEEE Access 7:94497–94507
10. Gao X, Shan C, Hu C, Niu Z, Liu Z (2019) An adaptive ensemble machine learning model for intrusion detection. IEEE Access 7:82512–82521
11. Yin H, Xue M, Xiao Y, Xia K, Yu G (2019) Intrusion detection classification model on an improved k-dependence Bayesian network. IEEE Access 7:157555–157563
12. Davis J, Clark A (2011) Data preprocessing for anomaly based network intrusion detection: a review. Comput Secur 30:353–375
13. Alqatf M, Yu L, Alhabib M, Al-Sabahi K (2018) Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. IEEE Access
14. Dhikhi T, Saravanan MS (2020) An intellectual detection system for intrusions based on collaborative machine learning. Int J Adv Comput Sci Appl 11(2) (Scopus/WoS)
15. Mizianty M, Kurgan L, Ogiela M (2010) Discretization as the enabling technique for the Naïve Bayes and semi-Naïve Bayes-based classification. Knowl Eng Rev 25(4):421–449
16. Hacibeyoglu M, Ibrahim MH (2018) Ef_Unique: an improved version of unsupervised equal frequency discretization method. Arab J Sci Eng
17. Ghazy RA, El-Rabaie EM, Dessouky MI et al (2020) Feature selection ranking and subset-based techniques with different classifiers for intrusion detection. Wireless Pers Commun 111:375–393
18. Mechtri L, Djemili Tolba F, Ghoualmi N (2010) Intrusion detection using principal component analysis. In: Second international conference on engineering system management and applications, Sharjah, pp 1–6
19. Essam Y, Elsalamouny E, Eltaweel G (2015) Improving the performance of multi-class intrusion detection systems using feature reduction. Int J Comput Sci Issues
20. Cepheli Ö, Büyükçorak S, Kurt GK (2016) Hybrid intrusion detection system for DDOS attacks. J Electrical Comput Eng, Article Id 1075648, 8 Pages

Biologically Motivated Hexagonal Framework for Machine Vision

Prathibha Varghese and G. Arockia Selva Saroja

Abstract In view to the current and traditional state of art, all image analyzing device for capturing, recording and display are exclusively meant to utilize square grids. But biological structure models, however, pinpoint an alternative, evolutionary motivated base structure. Motivated by the optical perception system of human, hexagonal image processing (HIP) offers many robust advantages and benefits for the researchers. Hexagonal structure offers decisive advantages like radical symmetry, equidistant and unique neighbours, homogeneity in the lattice and 13.4% increased transformation efficiency. Even though hexagonal image processing is known for good image quality but the implementation approach made it difficult to work as there is no working hardware that applies hexagonal imaging concept. Although with these much of processing power of computers and graphic devices, hexagonal representation has not yet developed to a standardized framework or library for HIP, so the time is ripe to review the various hexagonal grid structure rendering algorithms and hexagonal sampling.

Keywords Hexagonal pixel · Hexagonal sampling · Spiral architecture · Hexagonal grid

1 Introduction

I'd like to open another avenue of image processing through this paper, the less explorable idea of hexagonal image processing. Even though this area of research has been in existence from last 40 years, due to the ease in implementation majority of existing imaging systems use square lattice instead of hexagonal. Following the 2-D sampling theory [1], as shown in Fig. 1a for circularly band-limited analog

P. Varghese (✉)
Electronics and Communication Department, Noorul Islam Centre for Higher Education, Tamil Nadu, India
e-mail: prathibhavarghese2003@gmail.com

G. Arockia Selva Saroja
Department Electronics and Communication Engineering, Noorul Islam Centre for Higher Education, Tamil Nadu, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_23

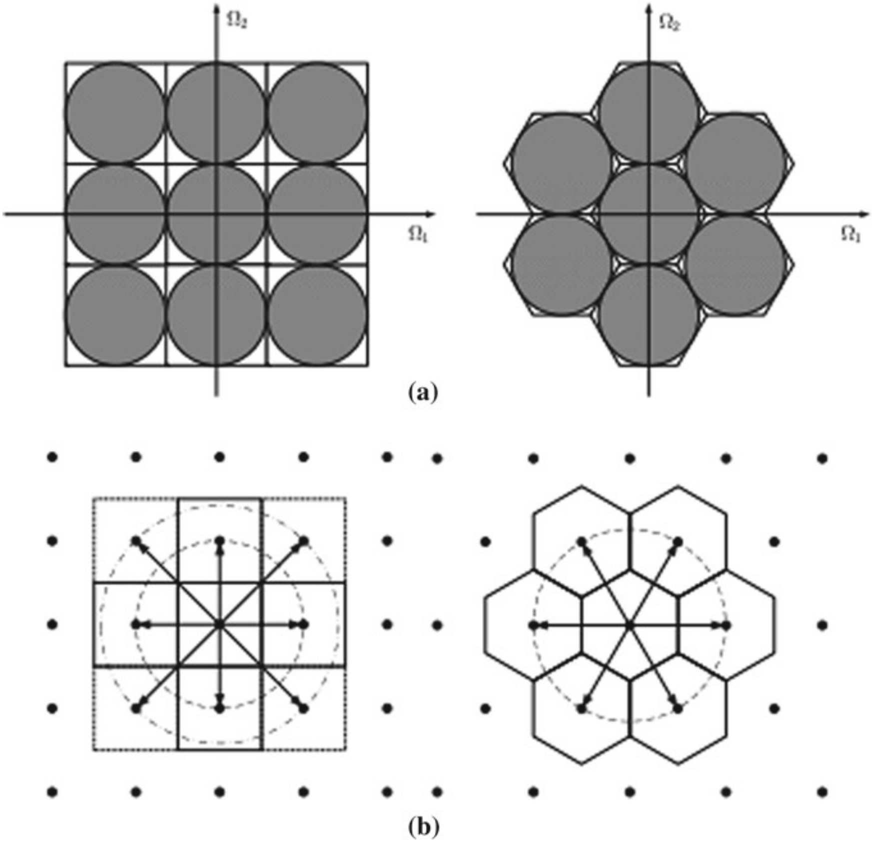


Fig. 1 Hexagonal structure showing **a** High sampling efficiency, **b** better geometric properties

images, the hexagonal sampling will provide 13.4% fewer samples than the square lattice [2]. Hexagonal lattice is more powerful than the traditional square lattice in geometric properties, such as smoother angular symmetry, equii-distance and uniformly connected pixels with consistent six neighbours [2], as shown in Fig. 1b. In order to overcome these fundamental limitations in square lattice systems, there is a need for new ways to capture hexagonal images.

This research work is designed in the following way. Section 2 deals with the regular sampling and hexagonal sampling. Section 3 deals with the Psuedo hexagonal pixel framework arrangement. Section 4 deals with the spiral architecture addressing scheme. Section 5 deals with conclusion and future work.

2 2-D Sampling

Any 2-D band-limited signals can be sampled and reconstructed in their framework. According to Mersereau et al. [2], if $x_a(T_1, T_2)$ denotes analog image, then rectangular sampling can be represented as:

$$x(n_1, n_2) = x_a(n_1 T_1, n_2 T_2) \tag{1}$$

where $T_1$and T_2 represents sampling intervals in horizontal and vertical direction. If $x_a(T_1, T_2)$ is a regular square band limited signals as depicted in Fig. 2a, then to avoid aliasing, sampling intervals must satisfy the condition given in Eq. 2:

$$T_{r1} \leq \frac{\pi}{W_{r1}} and T_{r2} \leq \frac{\pi}{W_{r2}} \tag{2}$$

where $W_{\mathrm{r}1}$ and $W_{\mathrm{r}2}$ represents bandwidth in horizontal and vertical direction in radians.

Suppose $x_a(\Omega_1, \Omega_2)$ be the Fourier transform of $x_a(T_1, T_2)$, then $x_a(T_1, T_2)$ is said to be band-limited in band region R, if X_a (Ω_1, Ω_2) satisfies the condition.

$$x_a(\Omega_1,\Omega_2) = 0 \ , (\Omega_1,\Omega_2) € \ R \tag{3}$$

Two sampling orientations of hexagonal sampling are changed but the sampling intervals remains unchanged in both horizontal and vertical directions as depicted in Fig. 4b.

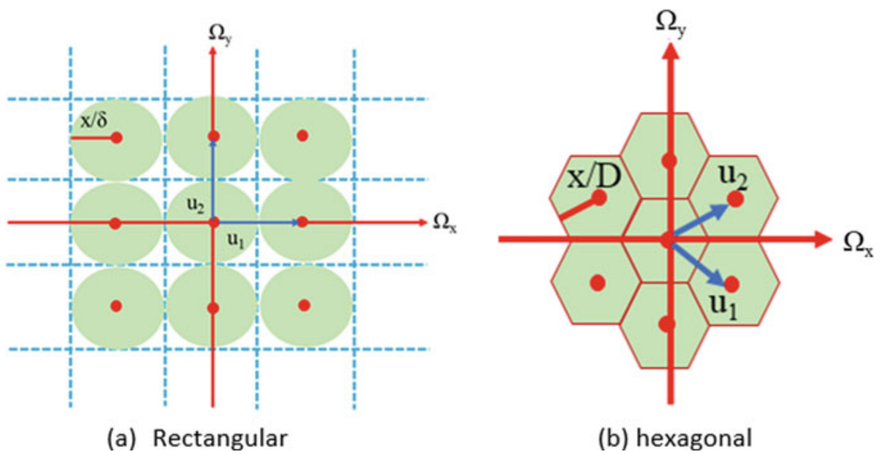


Fig. 2 a Rectangular sampling, b hexagonal sampling

$$x(n_1, n_2) = x_a\left(\left(n_1 - \frac{1}{2}n_2\right)T_{h1}, n_2T_{h2}\right) \tag{4}$$

For $X_a(T_1, T_2)$ to recover completely from the hexagonal sequence it should satisfy this condition

$$T_1 \leq \frac{4\pi}{2w_1 + w_2} \text{ and } T_2 \leq \frac{\pi}{w_2} \tag{5}$$

where $w_1 = w_2 = \frac{2W_2}{\sqrt{3}}$ as shown in Fig. 3.

Figure 4 shows the circular shape inscribed in square structure and hexagon structure. It is evident that vertical sampling intervals are same but horizontal intervals differ.

A circular band region is defined as $x_a(\Omega_1, \Omega_2) = 0$,if $\Omega_1^2 + \Omega_1^2 \geq W^2$ where Ω_1, Ω_2 represents digital frequencies in horizontal and vertical orientations and W

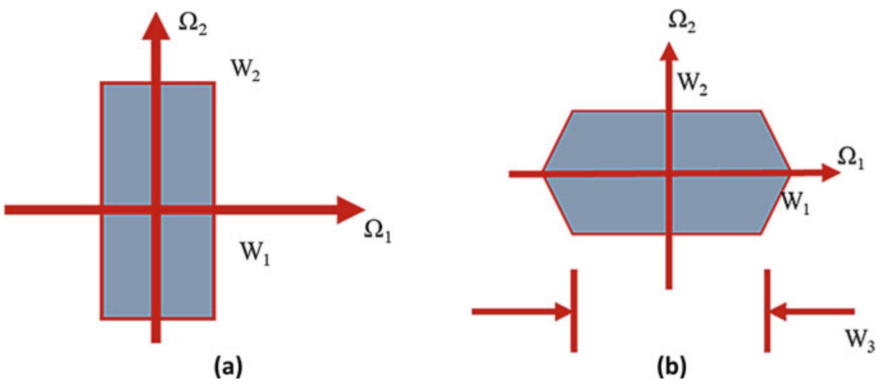


Fig. 3 **a** Rectangular band region, **b** hexagon band region

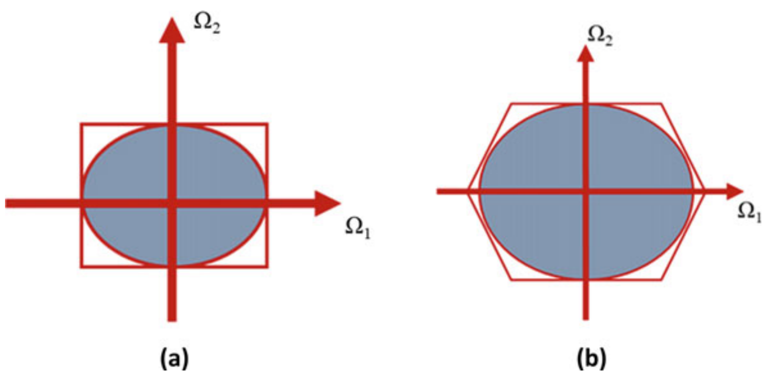


Fig. 4 **a** Circularly band limited structure inserted in (**a**) Square region, **b** hexagon region

shows the inner radius. In order to neglect, aliasing, the circular ring should hold circular regions effectively. Hence, sampling requirements of (2) and (5) will become

$$T_{r1} = \frac{\pi}{W} \text{ and } T_{r2} = \frac{\pi}{W} \text{ for square lattice} \tag{6}$$

$$T_{h1} = \frac{2\pi}{\sqrt{3}W} \text{ and } T_{h2} = \frac{\pi}{W} \text{ for hexagonal lattice} \tag{7}$$

3 Psuedo Hexagonal Pixel

Wuthrich and Stucki [3] implemented an innovative pseudo hexagonal grid structure, in which each hexagonal pixel can be created by grouping finite group of square pixels as shown in Fig. 5.

In constructing pseudo hexagonal pixels from square pixels, Jeevan [4] proposed an algorithm called JK algorithm to create pseudo hexagonal pixels. In JK algorithm, 72 pixels are used for the creation of pseudo hexagonal pixels. But, actually 56 complete pixels take part in the formation of pseudo hexagonal structure as shown in Fig. 6.

The square grid pixels that take part in the formation of hexagonal pixels are shown 1's and others 0's. Hexagonal pixels are completely converted square pixels in the above method. Two basic properties of hexagonal structure [5, 6] are satisfied in the above-mentioned algorithm.

They are: (1) 6-neighbour consistency and (2) angular resolution. The equation for the distance calculation from the centre pixel to the adjacent neighbouring pixel is as follows as shown in Fig. 7.

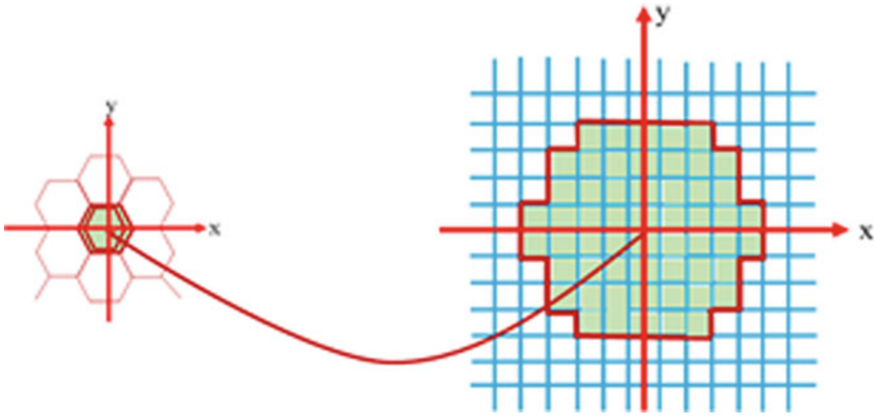


Fig. 5 Arrangement of pseudo hexagonal structure

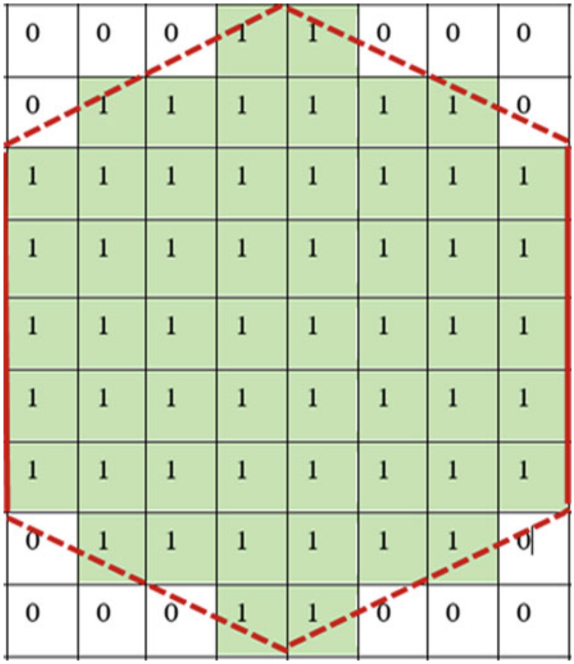


Fig. 6 Single hexagonal pixel

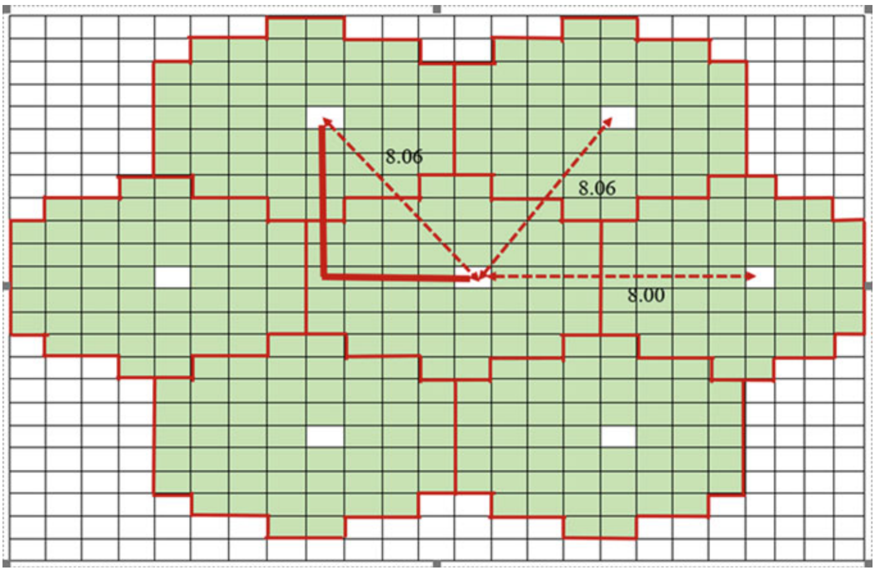


Fig. 7 Six neighbours and consistent connectivity

$$\sqrt{7^2 + 4^2} = 8.06 \simeq 8\,\text{units}$$

For the hexagonal pixel generation method, from Eqs. (6) and (7) it can be inferred that vertical sampling intervals in both traditional square grid and hexagonal grid is same $\frac{\pi}{W}$, but sampling intervals in horizontal directions differ. So we can infer that square lattice to hexagonal lattice conversion is possible only along in horizontal orientation, thus minimizing square lattice to hexagonal lattice. This spark has lead to the generation of psuedo hexagonal lattice grid. Hence, we will prove it mathematically. Let the 2-D sampling intervals for square lattice be T_{s1} and T_{s2} and for hexagonal lattice be T_{h1} and T_{h2}. Then the square sampled image can be shown as $x_a(n_1 T_{s1}, n_2 T_{s2})$ and analog image reconstructed may be represented as $x_r(t_1, t_2)$. Hexagonal resampled image can be represented as $x_h(k_1, k_2) = x_r\big(\big(k_1 - \frac{1}{2}k_2\big)T_{h1}, k_2 T_{h2}\big)$. Let the interpolation kernal $h_r(t_1, t_2)$ used be defined as

$$h_r(t_1, t_2) = h_{r1}(t_1) h_{r2}(t_2) \tag{8}$$

Let us regenerate the original analog image $x_r(t_1, t_2)$ using 2-D interpolation technique

$$\begin{aligned} x_r(t_1, t_2) &= \sum_{n_1} \sum_{n2} x_s(n_1, n_2) h_r(t_1 - n_1 T_{s1}, t_2 - n_2 T_{s2}) \\ &= \sum_{n_1} \sum_{n2} x_s(n_1, n_2) h_{r1}(t_1 - n_1 T_{s1}) h_{r2}(t_2 - n_2 T_{s2}) \\ &= \sum_{n_1} h_{r1}(t_1 - n_1 T_{s1}) \sum_{n_2} x_s(n_1, n_2) h_{r2}(t_2 - n_2 T_{s2}) \end{aligned} \tag{9}$$

In order to get the hexagonal sampled lattice image, we use the reconstructed analog image $x_r(t_1, t_2)$.

$$\begin{aligned} x_h(k_1, k_2) &= x_r\left(\left(k_1 - \frac{1}{2}k_2\right) T_{h1}, k_2 T_{h2}\right). \\ &= \sum_{n_1} h_{r1}\left(\left(k_1 - \frac{1}{2}k_2\right) T_{h1} - n_1 T_{s1}\right) \\ &= \sum_{n_2} x_s(n_1, n_2) h_{r2}(k_2 T_{h2} - n_2 T_{s2}) \end{aligned} \tag{10}$$

We know that,$T_{s2} = T_{h2}$. Also, consider the $sinch_{r2}t_2$ has an interpolation function, it exhibits a powerful and useful property that $h_{r2}(0) = 1$ and $h_{r2}(kT_{s2}) = 0$, resulting in $\sum_{n_2} x_s(n_1, n_2) h_{r2}(k_2 T_{h2} - n_2 T_{s2}) = x_s(n_1, k_2)$. That is, interpolation along the vertical direction can be discarded. Thus, the resultant hexagonal sampled image will be

$$x_h(k_1, k_2) = \sum_{n_1} h_{r1}((k_1 - k_1 - \frac{1}{2}k_2)T_{h1} - n_1 T_{s1} x_s(n_1, k_2) \tag{11}$$

Here 2-D interpolation has happened only in horizontal direction. Hence, we can reduce 2-D interpolation to 1-D interpolation. Here we can see that interpolation is performed only in horizontal direction. Also in hexagonal sampling, pixels are $d/2$ distance staggered where d represents the diameter of the circle inscribed in hexagon as shown in Eq. 11 and this staggering is fixed between odd and even rows in the image.

4 Spiral Architecture (SA) Addressing Scheme

Aroused from anatomical consideration of primates vision, Sheridian et al. [7] proposed spiral architecture addressing scheme. This addressing scheme begins from the middle of the image in terms of powers of seven in a curvilinear fashion as shown in Fig. 8. Spiral addition and multiplication works is based on spiral architecture [7, 8]. In the conventional rectangular grid, group of 3 $\times$ 3 rectangles corresponds to unit vision which is having two different neighbourhood (4 and 8) [9, 10]. So there is time saving in local and global processing. Chain coding scheme is used to extract object contour. The cluster size designed for hexagon is 7^n.

A cluster size 49 hexagons with their corresponding address is given in Fig. 9. This code calculates a set of $\{c[j]\}$, of the spiral architecture that comprise $N_\lambda(w_0)$. The total number of spiral additions calculated in this approach to obtain a set of spiral

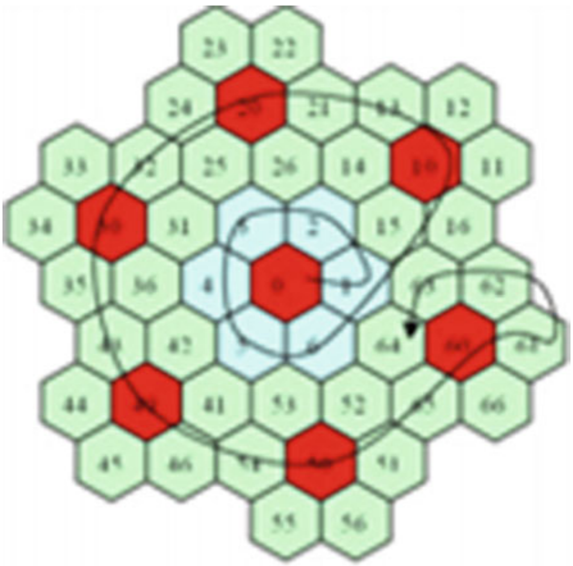


Fig. 8 Spiral addressing method

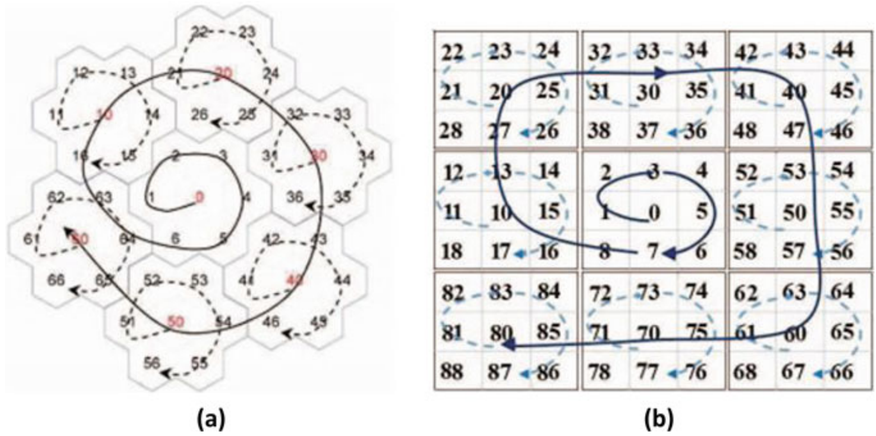


Fig. 9 **a** Spiral addressing scheme for hexagonal image, **b** square image processing addressing scheme

addresses in a layer-ƛ cluster neighbourhood $N_\lambda(w_0)$ is significantly reduced when compared with the required number of spiral additions actually used to calculate a group of spiral addresses for a ƛ cluster neighbourhood $N_\lambda(w)$ with $w \neq 0$, mod 7.

4.1 Locating Hexagonal Pixel: Spiral Address Calculation Method

Let hexagonal pixel with spiral address [11, 12] architecture is denoted by $D(x)$. Then, the centre point will be

$D(0) = [0\,0]$. The points in layer-1 is given as below:

Correspondingly, $D(1) = [-1\ 0]$, $D(2) = [-1\ 1]$, $D(3) = [0\ 1]$, $D(4) = [1\ 1]$, $D(5) = [1\ 0]$ $D(6) = [1]$, $D(7) = [0\ -1]$, $D(8) = (\ -1\ -1)$. Here pixels in layer-1 is numbered from 0 to 8. The layer-2 addressing starts from 10 to 88. Further layers are numbered in similar pattern in clockwise direction.

The shift addresses provided for 0 to 6 are the fundamental units in these recursive algorithm. The algorithm for the next set is multiples of 10 is given by

$$D(X * 10^i) = D(x * 10^{1-1}) + 2D((x+1) * 10^{i-1})$$
$$D(6 * 10^i) = D(6 * 10^{i-1}) + 2D(10^{i-1})$$
$$\text{For } i = 1, 2 \ldots, x = 1, 2 \ldots 5 \ldots.$$

Spiral address for finding the position of hexagonal pixels are

$$X_n X_{n-1} \ldots a1,\ (X_i = 0, 1, 2, \ldots 6. \text{ For } i = 1, 2, \ldots n.)$$

Fig. 10 Single structure of hexagonal pixel

		x	x	x	x	x		
	x	x	x	x	x	x	x	
	x	x	x	x	x	x	x	
x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	
	x	x	x	x	x	x	x	
		x	x	x	x	x		

Can be find out by

$$D\left(X_n X_{n-1}..a_1\right) = \sum_{i=1}^{n} D(x_i * 10^{i-1})$$

where $\sum$ denotes spiral additions and *spiral multiplications. Using the spiral architecture, spiral addition operation and multiplication operation are defined which results in translation and rotation of the image. In order to handle spiral addition, carry rule is definitely required. To design hexagonal shaped pixels [5], every square shaped pixel is sub divided into 7 × 7 smaller set of pixels. The structure of these hexagonal pixels is shown in Fig. 10.

The dimension of the newly implemented pixel will be 12.5% larger than the square pixel. Consequently, for the creation of the hexagonal grid pixels requires only 12.5% less than the traditional square pixel grid. The dimension of each implemented hexagonal pixels is [{(hexagonal sub-pixels total) − (square pixels total)}/Required hexagonal sub-pixel] 100 = 12.5%.

The arrangement seven such hexagon pixels are shown in Fig. 11.

In spiral architecture, hexagonal pixels are arranged in layers. In order to calculate the number of layers required to represent an image using SA, we can use the following Eq. (9)

$$\alpha = \frac{\log M + \log N}{\log 7} \tag{12}$$

where ‘α’ denotes the required number of levels and M × N represents the image.

On further depth analysis, SA [12] has got some distinguishable features when we compare with the square grid (traditional) image processing. First and foremost, single dimensional addressing lead the way to efficient storage and location of the origin at the centre of the image simplifies geometric transformation of image [6]. Secondly, it provides consistent neighbourhood connectivity which is very beneficial

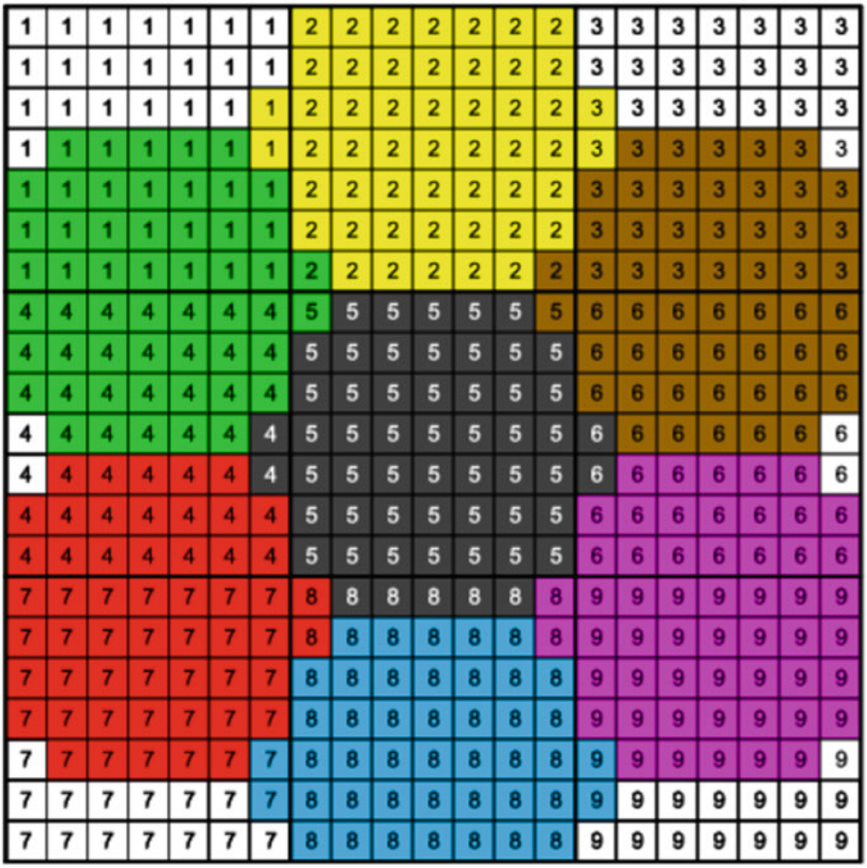


Fig. 11 Cluster of seven hexagon pixels

for computer vision processing application areas. The hyper pixels described above has more pleasing effect when considering the oblique effect of human vision [11, 12] and the main features are highlighted in Table 1.

Table 1 Main features of square and hexagonal images

Main features of square hexagonal images		
Feature	Hexagonal	Square
Indexing	x	(x, y)
Origin	Image centre	Top left
Neighbours	hexagonal	Square
Image boundary	hexagonal	Rectangle

5 Conclusion

Computer vision era in this scenario is experiencing a vast change. For more than 40 years, interest for using hexagonal sampling grid for images is showed in both research and commercial levels. There is not much progress in this area of research compared to research on analyzing the images defined on a traditional rectangular lattice. As there is no direct hexagonal capturing device, the future work is to develop sensors that can capture hexagonal images. HIP framework showcased the utility by different software acquisition method and the addressing schemes which can be demonstrated in different image processing techniques. From above discussions, it is clear that there is improvement and better visualization with hexagonal sampling.

References

1. Petersen DP, Middleton D (1962) Sampling and reconstruction of wave-number-limited functions in N dimensional euclidean spaces. Information Control 5(4):279–323(1962)
2. Mersereau RM (199) The processing of hexagonally sampled two-dimensional signals. In: Proceedings of the IEEE, vol 67, no 6, pp 930–949
3. Wuthrich CA, Stucki P (1991) An algorithmic comparison between square- and hexagonal-based grids. In: CVGIP: graphical models and image processing, vol 53, issue 4, pp 324–339
4. Jeevan KM, Krishnakumar S (2016) An algorithm for the simulation of pseudo hexagonal image structure using MATLAB. Int J Image Graphics Signal Process 8(6):57
5. Schlosser T, Friedrich M, Kowerko D (2019) Hexagonal image processing in the context of machine learning: conception of a biologically inspired hexagonal deep learning framework. In: 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE
6. Luo J et al (2019) Hexagonal convolutional neural networks for hexagonal grids. IEEE Access 7:142738–142749
7. Sheridan P (1996) Spiral architecture for machine vision. University of Technology, Sydney
8. He X (1999) 2-D object recognition with spiral architecture. University of Technology, Sydney
9. Watson AB, Null CH (1997) Digital images and human vision
10. Gupta P, Pahwa K (2014) Square pixels to hexagonal pixel structure representation technique. Int J Signal Process Image Process Pattern Recogn 7(4):137–144
11. Šabanovič E et al (2020) Identification of road-surface type using deep neural networks for friction coefficient estimation. Sensors 20(3):612
12. Wen W, Khatibi S (2018) Virtual deformable image sensors: towards to a general framework for image sensors with flexible grids and forms. Sensors 18(6):1856

A Comparative Study of Deep Learning Techniques for Farmer Query Text Classification

K. Kanchanadevi, J. Arun Pandian, and S. Saranya

Abstract This research proposes Long Short-Term Memory Network models with different stacked LSTM layers for farmer queries classification. The 23 different classes of 105,724 farmer queries were used to train the proposed LSTM models. Word to vector (Word2Vec), Global Vectors for Word Representation (GloVe), and FastText embedding techniques were compared on the query classification. The Word2vec embedding technique produced a better result than the GloVe and FastText embedding techniques in the farmer query classification. After an extensive simulation, the DLSTM network with three stacked LSTM layers achieved a testing accuracy of 90.35%. Classification performance of the proposed DLSTM network with three stacked LSTM layers in farmer queries was superior to Convolutional Neural Network (CNN), LSTM and other DLSTM models.

Keywords Convolutional Neural Network · Long Short-Term Memory Network · Word Embedding

1 Introduction

Methods of cultivation of crops in many parts of India depend primarily on factors such as the situation of the land, type of soils, irrigation facilities, pest and disease management, availability of laborers, market status, intensity and distribution of rainfalls [1]. Low and imbalanced use of fertilizers is a common challenge in plant cultivation [2]. Handling infestation of weeds and pests is also the central problem of plant cultivation [3]. Farmers require assistance for answering queries related to cultivation raised by farmers. An automatic question answering systems are advantageous; they allow the user to ask the question based on specific facts and the

K. Kanchanadevi (✉) · J. Arun Pandian
Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India
e-mail: kanchanadevicse@veltech.edu.in

S. Saranya
Computer Science and Engineering, K.Ramakrishnan College of Engineering, Tiruchirappalli, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_24

system tries to use the context in the supporting facts to answer the queries instead of giving the related keywords. Question classification has a significant role in automated question answering systems [4]. It is used to label a question into a class that represents the question type. The selected question type can be used for filtering the related answers. Question classification is the task of organizing questions into pre-determined classes, generally using machine learning algorithms.

Text classification is crucial in numerous Natural Language Processing (NLP) applications, such as question classification, sentiment analysis, and spam detections. Recurrent Neural Networks can handle a sequence of any length data and capture long-term dependencies. It has achieved remarkable results in document or sentence modeling. It uses recurrent connections to remember the past processed information. A Long Short-Term Memory (LSTM) network is a category of RNN that can learn long-term dependencies between time steps of sequence data. The critical components of an LSTM network are a sequence input layer and an LSTM layer. A sequence input layer in the network inputs sequence or time-series data into the network. Moreover, the LSTM layer learns long-term dependencies between time steps of sequence data using hidden units [5].

A word embedding is a learned representation for text where words with the same meaning have a similar representation. Word embedding is an approach to representing words and documents that may be considered one of the critical breakthroughs of Deep Learning (DL) on challenging NLP problems. The word embedding methods learn a real-valued vector representation for a predefined fixed-sized vocabulary from a text corpus. The learning process is either joint with the neural network model on some task, such as document classification, or is an unsupervised process using document statistics. Word to Vector (Word2Vec) is a statistical method for efficiently learning a standalone word embedding from a text corpus. Global Vectors (GloVe) word embedding extends to the word2vec method for efficiently learning word vectors [6]. This research compares the GloVe and word2vec approaches on word embedding for the classification model.

This article presents the deep learning architecture for classifying the farmer queries using the Deep Long Short-Term Memory (DLSTM) network. The RNN has additional recurrent connections compared to regular neural networks to remember past processed information. These connections, however, make it more computationally intensive to train the RNN. The testing performance of the proposed model was compared with CNN, bidirectional LSTM and other stacked LSTM approaches. Classification accuracy, precision, recall, specificity and F1 score metrics were used to evaluate the performance of the classification model.

2 Literature Survey

Numerous techniques have been concentrated on measuring the text classification from the given set of queries. Due to the vast availability of the supervised and unsupervised learning techniques from studies, the text classification technique has been explained with many algorithms and this algorithm helps the users classify the text document from different resources. In [7], the authors presented a study of text classification, comparisons and various weighing methods to enhance the text classification performance. The authors in [8] focused on the construction and improvement of distributed naive Bayes automatic classification system and the main application of Hadoop cloud computing in web page classification. Because of the Bayesian text classification method, a correlation between the feature sets generated after feature selection has been executed and then features with higher correlation degrees appropriately combined.

In addition, the authors [9] proposed a Hierarchical Attention Network (HAN) model using the Theano framework, indicating more accurate validation for large datasets for text classification. The authors [10] analyzed the performance of SVM and Roccio classifiers on two widely known benchmark datasets with different characteristics. Experimental results showed that modification of term frequency factor in supervised term weighting schemes showed an increase in performance of almost all weighting schemes. In [11], the authors proposed a new simple feature selection method to filter the redundant features and improve efficiency. They introduced word frequency-based relevance and correlative redundancy to calculate the relationship between two words. They have experimented with three datasets wherein support vector machines and naive Bayes are used. The classification accuracy is shown that high compared with other methods and also it runs faster than the traditional methods. The experiments have been done from the KEEL repository to get more balanced and accurate results to overcome the problems with low classification accuracy.

Moreover, the authors in [12] focused on classification problems that include pattern recognition. Their primary focus is on studying the context of granular computing and proposed a fuzzy rule-based system for the recognition-intensive classification of real-life data instances. The experimental setup was implemented via the seven available data sets on life sciences to compare the fuzzy approach with four popular probabilistic approaches in pattern recognition tasks. Likewise, in [13], the authors proposed a multi-label text classification method to combine the Dynamic Semantic Representation Model and Deep Neural Network (DSRM-DNN). Text classifiers can be used to combine the deep belief network and back-propagation neural network.

3 Data Collection and Preprocessing

Question answer text data is naturally sequential data. Also, a question or answer is a sequence of words, which might have dependencies between them. The LSTM neural networks were used to learn and use long-term dependencies to classify sequence data in this research. The research proposed various LSTM models using various LSTM and bidirectional LSTM layers on farmer query classification tasks. The models were named as Single LSTM layered network (LSTM), Two LSTM layered (DLSTM-2) network, Three LSTM layered network (DLSTM-3), Four LSTM layered network

(DLSTM-4) and Bidirectional LSTM network (BLSTM). Based on the performance of the proposed LSTM models, the research suggests the most suitable model for farmer query classification. The subsequent subsections describe the implementation steps of the farmer queries classification model.

3.1 Dataset Preparation

The queries and answers of the farmer in the Kisan Call Centre (KCC) from various centers of Tamil Nadu state data were collected from the Open Government Data (OGD) Platform of India [14]. The farmer queries dataset has 23 different types of 105,724 questions. This dataset contains labeled textual representations of formers queries. The two-dimensional table data set was downloaded as a Comma Separated Values (CSV) file. Out of 105,724 examples, 95,163 examples were used for training and 10,561 examples were used for testing the performance of the proposed and existing text classification models.

The Data Preprocessing step was used to remove the unnecessary data from the dataset. Also, it helps to improve the performance and reduce the training time of the classification. At first, the Tokenization step is used to split the longer queries into smaller pieces or tokens. In the second step, all the text in the dataset was converted to lowercase characters. Also, multiple blank spaces, symbols and digits in the farmer queries were removed. Finally, the stop words in the dataset were removed using the Natural Language Toolkit (NLTK) library. Around 90% of examples were randomly selected for training and the remaining instances were used for the testing process. The proposed dataset was named a farmer query dataset. The class names and distribution of the classes in the farmer query dataset are shown in Table 1.

3.2 Word Embedding

After the data preparation, the word representation process was introduced in this research. The word embedding technique was used to do the word representation process. The word embedding technique maps each word or phrase to an N dimension vector of real numbers from the vocabulary. The three most popular and commonly used word embedding techniques were exclusively applied for embedding the farmer queries. The techniques are word to vector (Word2Vec), Global Vectors for Word Representation (GloVe) and FastText. Each word embedding technique was implemented with a simple LSTM model on the farmer data and the results were compared. Each word embedding technique was implemented with a single LSTM layer LSTM classification model on the farmer data and the results were compared.

Table 1 Class names of farmer query dataset

Class ID	Class name	Examples
1	Agriculture mechanization	777
2	Bio-pesticides and bio-fertilizers	381
3	Credit	134
4	Crop insurance	443
5	Cultural practices	6755
6	Fertilizer use and availability	8801
7	Field preparation	786
8	Government schemes	2388
9	Market information	9668
10	Nursery management	171
11	Nutrient management	17,822
12	Organic farming	537
13	Plant protection	25,382
14	Seeds	2739
15	Seeds and planting material	2701
16	Soil testing	457
17	Sowing time and weather	1428
18	Training and exposure visits	1346
19	Varieties	2846
20	Water management	193
21	Water management, micro irrigation	107
22	Weather	18,076
23	Weed management	1786

4 Query Classification Model

The five different LSTM networks were proposed to classify the farmer queries as 23 classes in this section. One to four LSTM layers were present in the various proposed LSTM networks. The subsequent subsection explains the architecture and training process of a single LSTM layered LSTM Network.

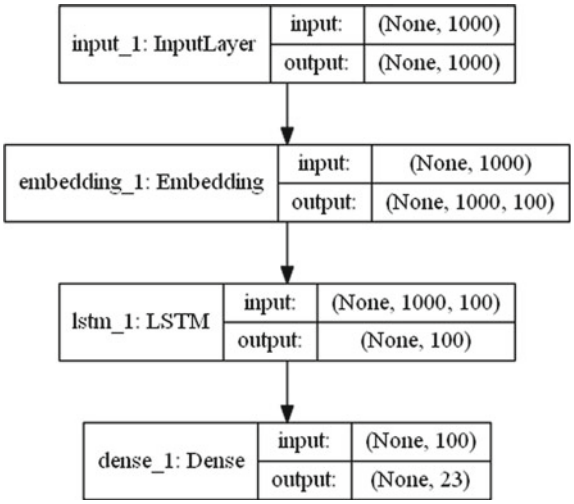
4.1 Single LSTM Layered Network

In this section, a Single LSTM layered network was proposed for the farmer query classification task. The input data was forwarded to the LSTM layer after the data preprocessing and embedding. The LSTM layer produced the output using the Tanh activation function and recurrent activation function of the sigmoid. The output data of the LSTM layer is given as an input of dense layer with the input size of 100 and output size of 23. The output size of the dense layer represents the number of classes. The softmax activation function was used in the dense layer to classify the input data. The layered architecture of the proposed Single LSTM layered network is shown in Fig. 1.

The Word2Vec, GloVe and FastText embedding techniques were compared in the proposed network. Each model was trained using 50 epochs on training data. The validation accuracy of the proposed Single LSTM layered network on the above embedding techniques was compared in Fig. 2.

The proposed Single LSTM layered network was trained using 100 epochs on the training dataset. Figure 3 illustrates the training and validation progress of the

Fig. 1 The layered architecture of the LSTM network



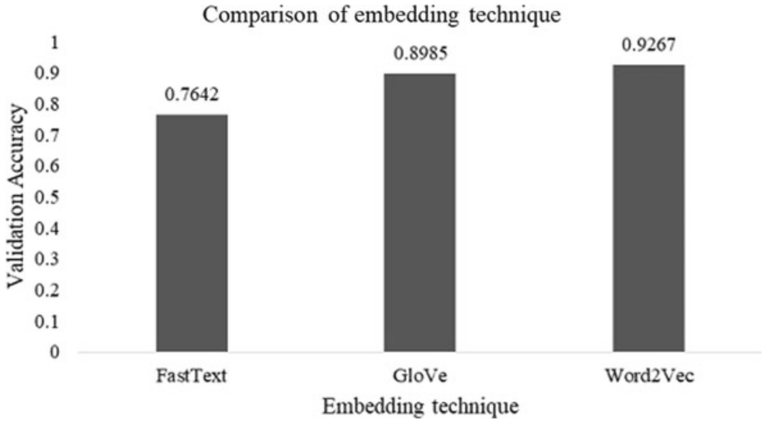


Fig. 2 Comparison of embedding techniques

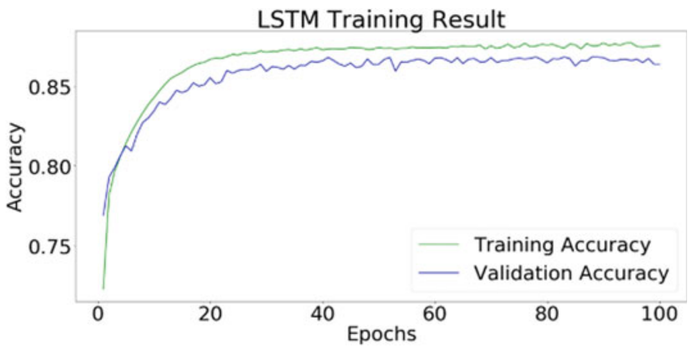


Fig. 3 Training and validation accuracy of LSTM network

model. The comparison result shows that the word2vec embedding technique-based LSTM network provides better performance than GloVe and FastText based models.

4.2 Two LSTM Layered Network

The Two LSTM layered network (DLSTM-2) for queries classification models were proposed in this section. At first, the embedded data was given as an input of the first LSTM layer. The output of the first LSTM layer is forwarded as an input of the second LSTM layer in the model. After the continuous two LSTM layers, the model introduced the dense layer for classifying the output data of the LSTM layer as 23 classes using the softmax activation function. The dimension of each LSTM layer in the DLSTM-2 network is 100 units. Figure 4 shows the layered architecture of the proposed DLSTM-2 network.

The proposed DLSTM-2 network was trained using 100 epochs on the training dataset. The training and validation performance of the proposed DLSTM-2 network is shown in Fig. 5. The training and validation performance of the proposed DLSTM-2 network illustrates that the stacked LSTM layered networks give better performance than the single LSTM layered network. The number of LSTM layers in the model was extended in the upcoming subsections based on this thought.

4.3 Three LSTM Layered Network

Three stacked LSTM layers were used in the Three LSTM layered network (DLSTM-3) for classifying the farmer queries. The embedding layer output is transferred to the first LSTM layer as an input. The second and third LSTM layers were processed and forwarded the data to the dense layer of the model. The dimension of all three LSTM layers in the DLSTM-3 network is 100 units. The dense layer classifies the

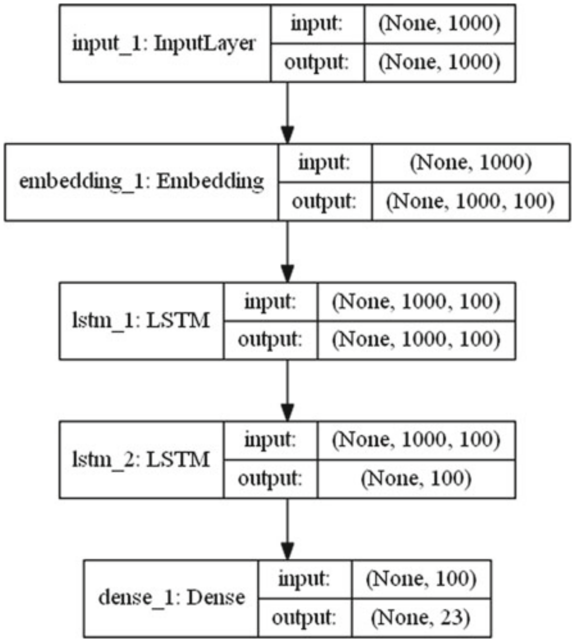


Fig. 4 The layered architecture of the DLSTM-2 network

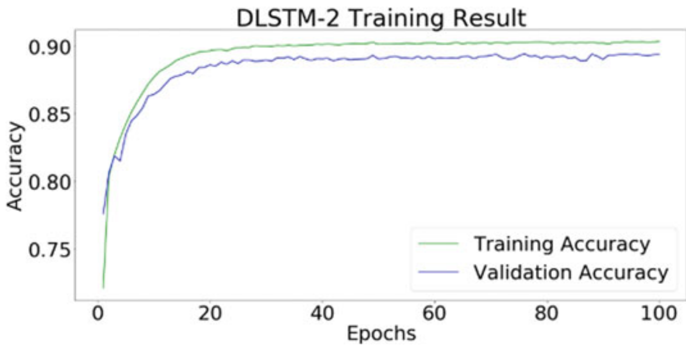


Fig. 5 Training and validation accuracy of DLSTM-2 network

stacked LSTM processed data using the softmax activation function. Figure 6 shows the architecture of the proposed three LSTM layered network.

The model was trained on the query dataset up to 100 epochs. The training and validation performance of the DLSTM-3 is shown in Fig. 7. The training and validation performance of the DLSTM-3 is much better than the previous networks. Based on the results, the research extends to the four LSTM layered networks (DLSTM-4). The training performance of the DLSTM-4 was less than the DLSTM-3 and DLSTM-2. For that reason, extending the LSTM layers in the network process ended up with DLSTM-3 network for farmer query classification tasks. The performance of the

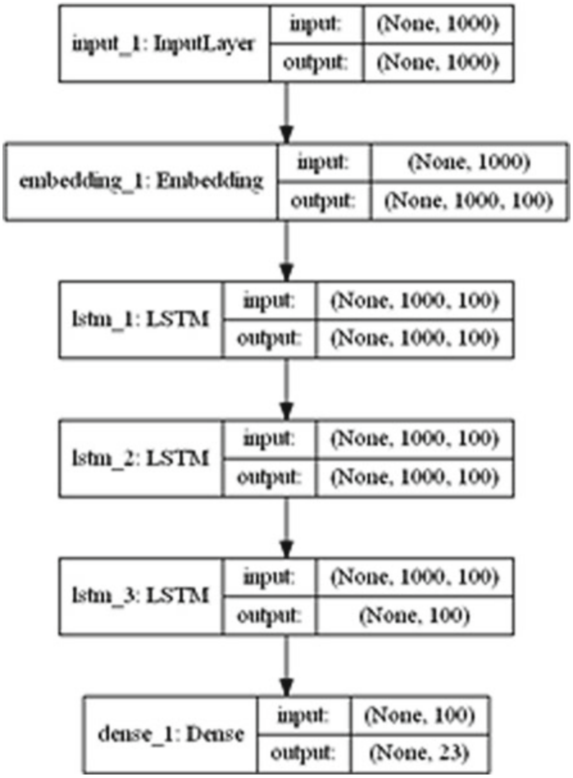


Fig. 6 The layered architecture of the DLSTM-3 network

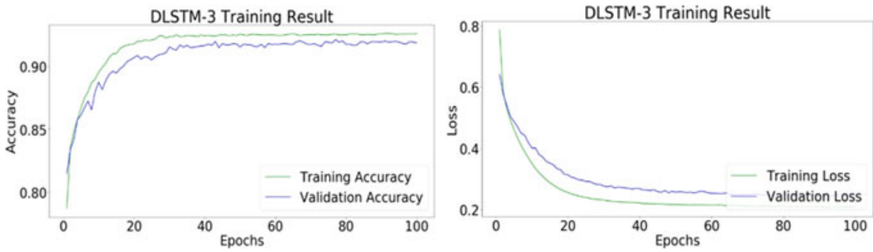


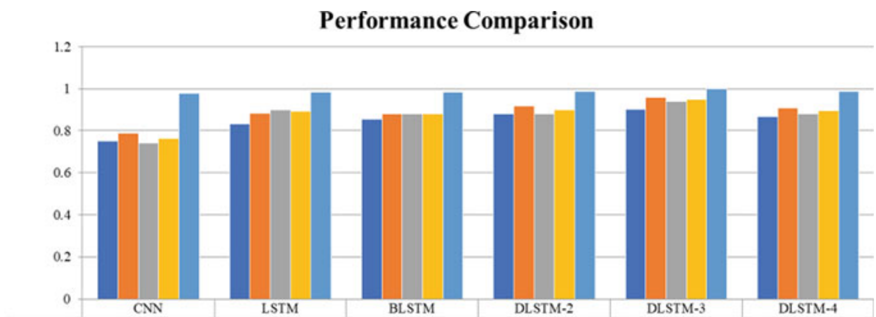
Fig. 7 Training and validation accuracy of DLSTM-3 network

proposed LSTM networks was compared using various performance matrices and testing data in the next section.

5 Result and Discussion

The proposed four different LSTM networks were trained and tested using 10,561 examples of the test dataset. The proposed LSTM, DLSTM-2, DLSTM-3, and DLSTM-4 models were compared using classification accuracy, precision, recall, specificity and F1 score [2]. Moreover, the performance of the proposed models is compared with the Convolutional Neural Network (CNN). Figure 8 shows that the classification accuracy, precision, recall, specificity and F1 score values of the DLSTM-3 are higher than other proposed and CNN models.

Area Under the Curve (AUC) is an evaluation metric that measures the capability of a classification model to differentiate classes and is used as a summary of the Receiver Operator Characteristic curve (ROC). Figure 9 shows the AUC-ROC curves of two randomly selected classes from the farmer queries dataset. Finally, the comparison of classification accuracy, precision, recall, specificity, and F1 score



	CNN	LSTM	BLSTM	DLSTM-2	DLSTM-3	DLSTM-4
Accuracy	0.7495	0.8321	0.8547	0.8791	0.9034	0.8681
Precision	0.7872	0.8823	0.88	0.9166	0.9591	0.9087
Recall	0.74	0.9	0.88	0.88	0.94	0.88
F1 Score	0.762870613	0.891062111	0.88	0.897927196	0.949453952	0.894119752
Specificity	0.978	0.9835	0.9837	0.9859	0.998	0.987

Fig. 8 Performance comparison of query classification models

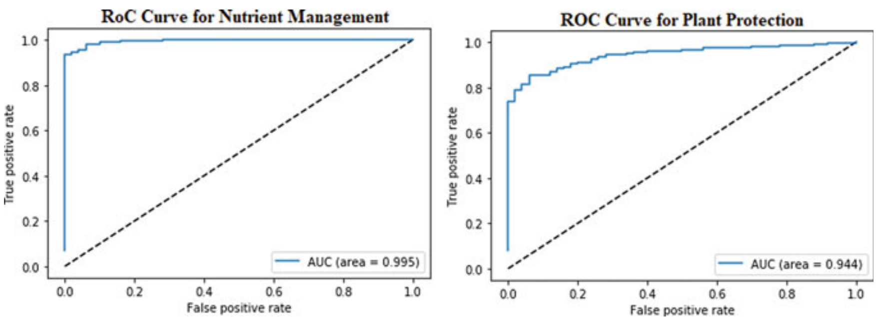


Fig. 9 AUC-ROC curves of nutrient management and plant protection classes of the DLSTM-3 model

results show that the proposed DLSTM-3 model was superior to all the above models in the farmer queries classification task.

6 Conclusions

This article developed a query classification model based on Deep Long Short-Term Memory Networks to classify farmer queries through simple query sentences of 23 different classes. The model training was performed using an openly available database of 105,724 instances, collected and preprocessed from the Open Government Data Platform India portal. The database comprises 120 different plants in 23 distinct classes of agriculture and horticulture queries. Comparison of results shows that the performance of the three LSTM layer models has a better testing accuracy was higher than the CNN, LSTM, BLSTM and Deep LSTM. The most successful Deep LSTM network with three stacked LSTM layers achieved a classification accuracy of 90.35% in the test dataset of farmer query texts. In the future, we plan to enrich the database and improve the model accuracy using different preprocessing and embedding techniques. We can extend this model for the automatic answering system for farmer queries. Future work is to develop a comprehensive cloud application for intelligent devices displaying generated answers based on farmer query inputs.

Acknowledgements The authors are grateful for the infrastructure support of Vel Tech Technology Business Incubator, Chennai.

References

1. Shah D, Shah HL, Dave HM, Mishra V (2021) Contrasting influence of human activities on agricultural and hydrological droughts in India. Sci Total Environ 774:144959 (2021). https://doi.org/10.1016/j.scitotenv.2021.144959
2. Geetharamani G, Pandian JA (2019) Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Comput Electr Eng 76:323–338. https://doi.org/10.1016/j.compeleceng.2019.04.011
3. Pandian JA, Geetharamani G, Annette B (2019) Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques. In: 2019 IEEE 9th international conference on advanced computing (IACC), pp 199–204. https://doi.org/10.1109/IACC48062.2019.8971580
4. Mohasseb A, Bader-El-Den M, Cocea M (2019) A customised grammar framework for query classification. https://doi.org/10.1016/j.eswa.2019.06.010
5. Li X, Cui M, Li J, Bai R, Lu Z, Aickelin U (2021) A hybrid medical text classification framework: Integrating attentive rule construction and neural network. Neurocomputing 443:345–355. https://doi.org/10.1016/j.neucom.2021.02.069
6. Sakketou F, Ampazis N (2020) A constrained optimization algorithm for learning GloVe embeddings with semantic lexicons. Knowl-Based Syst 195:105628. https://doi.org/10.1016/j.knosys.2020.105628

7. Kadhim AI (2019) Survey on supervised machine learning techniques for automatic text classification. Artif Intell Rev 52:273–292. https://doi.org/10.1007/s10462-018-09677-1
8. Gao H, Zeng X, Yao C (2019) Application of improved distributed naive Bayesian algorithms in text classification. J Supercomput 75:5831–5847. https://doi.org/10.1007/s11227-019-02862-1
9. Borna K, Ghanbari R (2019) Hierarchical LSTM network for text classification. SN Appl Sci 1:1124. https://doi.org/10.1007/s42452-019-1165-1
10. Dogan T, Uysal AK (2019) On term frequency factor in supervised term weighting schemes for text classification. Arab J Sci Eng 44:9545–9560. https://doi.org/10.1007/s13369-019-03920-9
11. Wang Y, Feng L (2018) A new feature selection method for handling redundant information in text classification. Front Inf Technol Electron Eng 19:221–234. https://doi.org/10.1631/FITEE.1601761
12. Liu H, Zhang L (2018) Fuzzy rule-based systems for recognition-intensive classification in granular computing context. Granul Comput 3:355–365. https://doi.org/10.1007/s41066-018-0076-7
13. Wang T, Liu L, Liu N, Zhang H, Zhang L, Feng S (2020) A multi-label text classification method via dynamic semantic representation model and deep neural network. Appl Intell 50:2339–2351. https://doi.org/10.1007/s10489-020-01680-w
14. Open Government Data (OGD) Platform India. https://data.gov.in

A Comparative Study of Deep Transfer Learning Techniques for Tomato Leaf Disease Diagnosis

J. Arun Pandian, S. Saranya, and K. Kanchanadevi

Abstract This study developed and analyzed the performance of ten different deep learning models using current transfer learning techniques to detect tomato leaf diseases. The transfer learning techniques were trained using 3000 tomato leaf images from the standard leaf disease dataset. The dataset consists of 3600 images on nine diseased and one healthy class. 300 diseased and healthy leaf images were used to validate the performance of the models. Validation performance of the transfer learning techniques was compared after 300 training epochs. The validation accuracy and loss of the DenseNet201 were 0.976 (97.6%) and 0.05476, respectively, after the 300 training epoch. Also, the test performance of the models was compared using previously unseen 300 images from the dataset. The average testing accuracy of the DenseNet201 model on tomato leaf disease detection was 96.24%. The performance of the DenseNet201 was superior to other transfer learning techniques on test data. Based on the comparison result, the training process of DenseNet201 based leaf disease detection model was extended to 1000 epochs on the dataset. The average testing accuracy of the trained DenseNet201 based tomato disease detection model was 98.31% on the test dataset.

Keywords Deep learning · DenseNet · Leaf disease diagnosis · Graphical processing unit · Transfer learning

1 Introduction

Around thirty-two percentages of food losses occur during the cultivation of food crops worldwide. Sowing, nourishing, pest and disease management, irrigation and harvesting influence farming production [1]. Plant diseases are the most significant

J. Arun Pandian · K. Kanchanadevi
Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai 600062, India

S. Saranya (✉)
Computer Science and Engineering, K.Ramakrishnan College of Engineering, Tiruchirappalli 621112, India
e-mail: saranyas.cse@krce.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_25

factor in cultivation loss in several food crops. Prognosis, diagnosis and treatment of food crop diseases play the most crucial role in improving crop yield. The leaves, stems, flowers and roots of the food crops are commonly affected by various diseases [2]. Most of the diseases are affecting the crop leaves and the diseases are highly visible to the outside. Considering the above factor, diagnosing crop leaf diseases is the most critical process in improved disease management. Advances in artificial intelligence, big data analytics, the Internet of Things and geographical information systems were used to diagnose and manage plant leaf diseases.

The production of tomatoes is around a hundred million tons per year and it is the second most important vegetable crop worldwide. Tomato can grow in a wide range of regions, climates and soil conditions [3]. The seeding to the first yield period of the tomato plant is around 90 days from 120 days. The diagnosis and management of leaf diseases are essential factors in the tomato yield improvement process. Bacterial spot, early blight, late blight, leaf mold, Septoria leaf spot, spider mites, target spot, mosaic virus and yellow leaf curl virus are the common leaf diseases of the tomato crop. Diagnosis and treatment of the above leaf diseases can improve the growth and yield of the tomato. Traditionally the pathologists perform laboratory assessments for the diagnosis of some plant diseases. The traditional process requires more time and cost for diagnosing diseases [4].

Deep learning is part of machine learning methods that use complex artificial neural network architectures. The deep learning techniques could solve various real-world problems such as classification, regression and clustering problems. The plant leaf disease diagnosis is a classification problem and it can be solved using classification techniques in deep learning [5]. A deep convolutional neural network (DCNN) is a deep learning technique recently applied to image classification with large imagery datasets. The DCNN produced better performance on image classification and object detection tasks in various domains. Transfer learning is a standard process used to reduce the model development time and improve the performance in machine learning [1]. Transfer learning is transferring knowledge gained while solving a problem in a different but similar problem [8]. The most common pre-trained models for transfer learning techniques in image classification are VGGNet, InceptionNet, ResNet, DenseNet, InceptionResNet, XceptionNet, MobileNet, NASNet and EfficientNet.

This research applies the recent version of the pre-trained models on tomato leaf disease detection and compares the performance of the pre-trained models for the disease detection task using the tomato leaf image dataset. The following sections of this article have been organized: Sect. 2 addresses the various transfer learning techniques in detecting tomato leaf disease. Section 3 presents the experimental outcomes of transfer learning techniques and related discussions. Section 4 provides conclusions of the survey on transfer learning techniques for tomato leaf disease detection.

2 Transfer Learning Techniques on Tomato Leaf Disease Diagnosis

Transfer learning is adapting the architecture and weights of pre-trained neural networks to a similar problem. The most advanced pre-trained models have been used to develop models for detecting tomato leaf disease in this research. The pre-trained models are VGG16 Net, VGG19 Net, InceptionV3 Net, ResNet152V2, DenseNet201, InceptionResNetV2, Xception Net, MobileNetV2, NASNetLarge and EfficientNetB7. The tomato leaf disease dataset was used to train, validate and test the pre-trained models in this research [6]. There are 3600 images of tomato leaves that have been categorized as nine diseases and one healthy class in the data set. The class names of the dataset are healthy, bacterial spot, early blight, late blight, leaf mold, Septoria leaf spot, spider mites, target spot, mosaic virus and yellow leaf curl virus [7]. The images are split for training, validation, and testing datasets, like 3000, 300 and 300 images. Randomly selected sample images from the dataset are shown in Fig. 1.

The subsequent sub-sections introduced the transfer learning techniques on tomato leaf disease detection. Also, the training performance of each model was discussed.

2.1 VGG16Net

The VGG16Net is a commonly used pre-trained model for transfer learning based applications in different domains [8]. The architecture of the VGG16 network was used to develop this model. The input size of the model was 224 pixels of width and 224 pixels of height. All the layer parameters are enabled to train from the data. Around 27,694,154 parameters are trained while developing the model. The training time of the model was around 3600 s for 300 epochs. The trained VGG16 model produced the training and validation accuracy was 0.9906 (99.06%) and 0.9386 (93.86%), respectively. Also, the training and validation loss was 0.03206

Fig. 1 Randomly selected images from the tomato leaf disease dataset

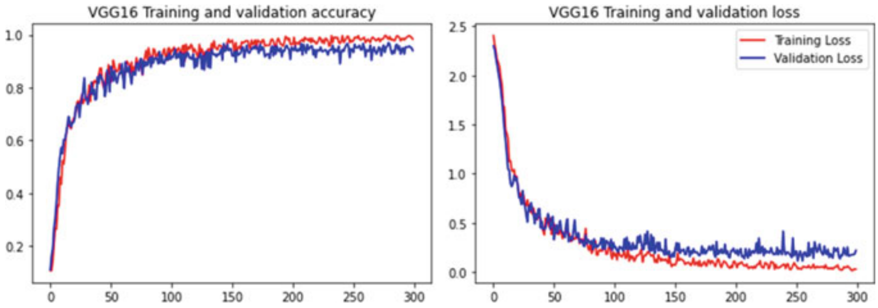


Fig. 2 Training and validation performance of VGG16 network

and 0.21108, respectively. The training performance of the VGG16Net based tomato leaf disease detection model is shown in Fig. 2.

The complete training process of the VGG16Net was done on a GPU workstation with a training batch size of 32, loss function of categorical cross-entropy and optimizer of stochastic gradient descent with a learning rate and momentum.

2.2 VGG19 Net

The VGG19Net is also a widely used pre-trained model for classification tasks and it extended from the VGG16 Net [9]. It produced a better performance for various applications than VGG16Net based models. Moreover, the input size of the VGG19Net was 224 pixels of width and 224 pixels of height. However, around 33,003,850 parameters were identified in the training process of VGG19Net. So, the training process of the VGG19Net on tomato leaf disease classification took 9600 s to reach 300 epochs. The VGG19 also used the same value for batch size, loss function and optimizer on the training process. The training performance of the VGG19Net on tomato leaf disease detection is illustrated in Fig. 3.

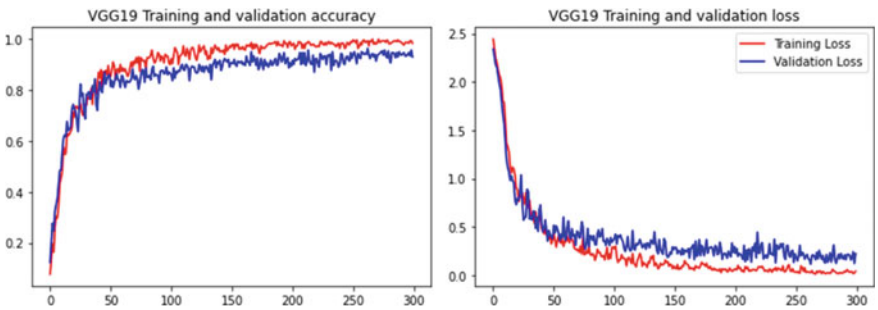


Fig. 3 Training and validation performance of VGG19Net

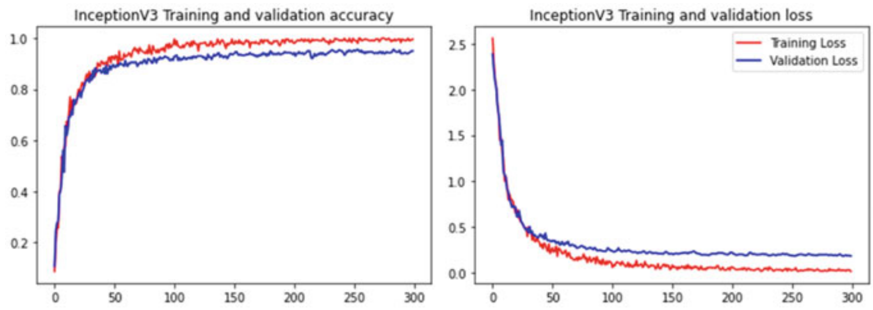


Fig. 4 Training and validation performance of InceptionV3 Net

The training and validation accuracy was calculated after the completion of 300 training epochs. The training and validation accuracy of the VGG19Net on the tomato leaf dataset was 0.98798 and 0.94134, respectively. And, the training loss and validation loss were 0.03234 and 0.18616, respectively. The training and validation performance of the VGG191Net on tomato leaf disease detection was better than the VGG16Net.

2.3 InceptionV3 Net

The InceptionV3 Net is another pre-trained model capable of processing the high dimensional image as inputs than the VGG16 and VGG19 Nets [10]. Therefore, several medical imaging applications were developed using InceptionV3 Net for the classification tasks. The dimension of the input images was 299 pixels of width and 299 pixels of height. The trainable parameter of the InceptionV3 Net was 89,046,058. The batch size only changed in the training process in this model. It was reduced to 16 to avoid the memory overflow issue. Figure 4 shows the training and validation performance of the InceptionV3 Net on tomato leaf disease detection task.

After completing the 300 epochs, the training accuracy of the model in training data was 0.99466 (99.47%). And the validation accuracy of the model on validation data was 0.946 (94.6%). Also, the training and validation loss was 0.02058 and 0.18366, respectively. The training and validation performance of the model was better than the previously developed models.

2.4 ResNet152V2

The ResNet152V2 Net is one of the most successful versions of the ResNet architecture of image classification applications [11]. The input dimension of the ResNet152V2 was 224 pixels of width and 224 pixels of height. The ResNet152V2

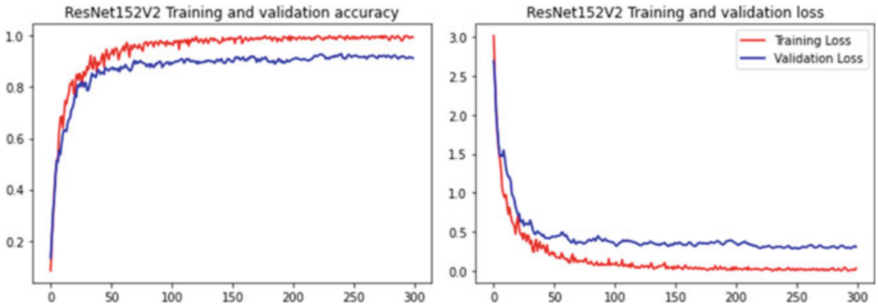


Fig. 5 Training and validation performance of ResNet152V2 Net

Net used more convolutional layers than previous models. So the trainable parameter of the ResNet152V2 was around 109,846,282. So, the batch size of the model was reduced to 16 for the training process on training data. And, the ResNet152V2 Net took 3300 s to complete the 300 epochs of training on the tomato leaf disease dataset. Figure 5 illustrates the training and validation performance of the ResNet152V2 Net.

The training accuracy and loss of the ResNet152V2 Net on training data were 0.994 (99.4%) and 0.02216, respectively. And the validation accuracy and loss of the model were 0.91466 (91.47%) and 0.3088, respectively. The training and validation performance of the ResNet152V2 Net is not better than the InceptionV3 Net based model.

2.5 DenseNet201

The DenseNet201 was the recent version of the DenseNet models for classification; it uses densely connected convolutional layers on the neural network [12]. The DenseNet201 gives better performance than other DenseNet models in different domains. The DenseNet201model was training on the tomato leaf disease dataset on 300 epochs. The input size of the input layer was 224 pixels of width and 224 pixels of height. The model was trained using the batch size of 32 on training data. The trainable parameters of the DenseNet201 model for tomato leaf disease detection were around 66,625,354. The training performance of the DenseNet201 model is shown in Fig. 6.

The model took 2400 s to complete the 300 training epochs on the training dataset. After completing the training process, the training accuracy and loss of the model were 0.99802 and 0.00946, respectively. Validation accuracy and loss of the model were 0.976 and 0.05476. The training and validation results of the DenseNet201 based classification model were superior to the previously developed models for tomato leaf disease detection in this research.

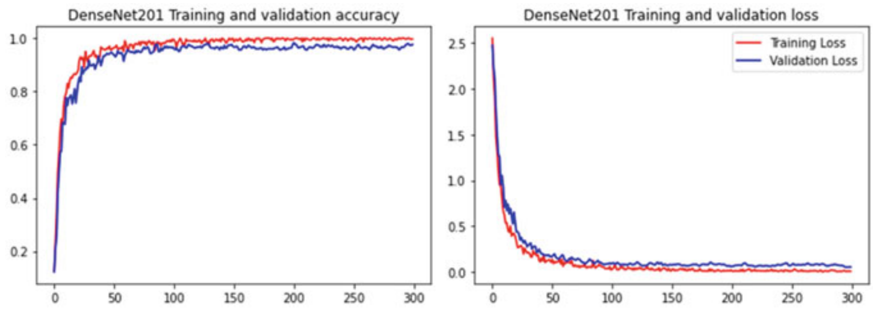


Fig. 6 Training and validation performance of DenseNet201

2.6 InceptionResNetV2

InceptionResNet is the classification model that combines the advantages of InceptionNet and ResNet. The InceptionResNetV2 is the recent network on the series of InceptionResNet for classification applications [13]. The input dimension of the model was 299 pixels of width and 299 pixels of height. The model trains 104,802,794 parameters on the tomato leaf disease dataset. The model trained on the dataset using a batch size of 16. The model trained on the dataset for 300 epochs and it took 4800 s to complete the training process. Figure 7 illustrates the training and validation progress of the model.

The training and validation accuracy of the InceptionResNetV2 was 0.98668 and 0.94132, respectively. And, training and validation loss of the model was 0.03954 and 0.21564 on the training and validation set. The training and validation performance of the InceptionResNetV2 is not better than the DenseNet201 based model on the tomato leaf disease dataset.

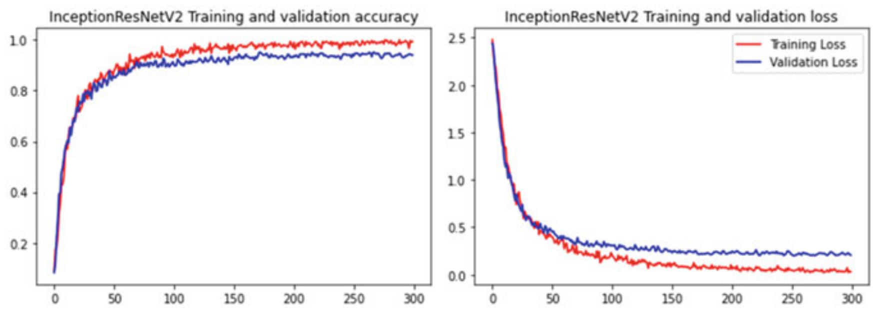


Fig. 7 Training and validation performance of InceptionResNetV2

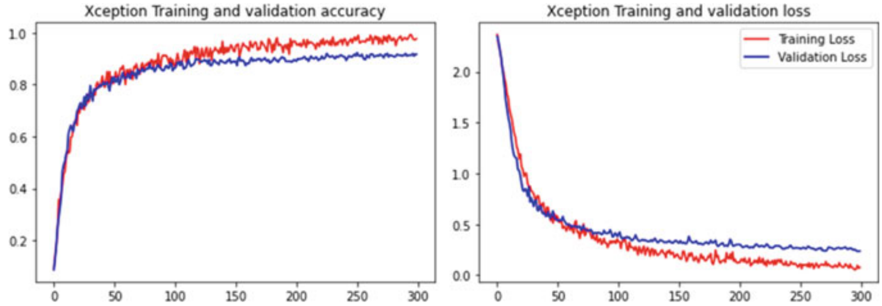


Fig. 8 Training and validation performance of Xception

2.7 XceptionNet

XceptionNet is another classification model used for several applications [14]. The input image size of this XceptionNet model is 299*299*3 pixels. The model trained 72,376,114 parameters on the development of tomato leaf disease detection. The model batch size was reduced to 16 on the training set. The training and validation performance of the XceptionNet for 300 epochs is shown in Fig. 8.

The training process of the model takes 3150 s on GPU for 300 epochs. The training accuracy of the model was 0.992268 and validation accuracy was 0.943864. Also, the training loss validation loss was 0.024816 and 0.189804, respectively. The training and validation performance of the XceptionNet was less than the DenseNet201 based model.

2.8 MobileNetV2

The MobileNetV2 is a lightweight network; the memory size of the saved model was less than previous models [15]. This model was suitable for implementation on portable devices. The input image size of the model was 244*244*3 (Width*Height*layer). The model was trained on the dataset with a batch size of 32. The number of trainable parameters in the MobileNetV2 was 34,505,034 on the tomato leaf disease dataset. The training performance of the model is shown in Fig. 9.

The model was trained in GPU for around 1800s to complete the 300 epochs. Training and validation accuracy of the model was 0.9931256 and 0.9443688 on tomato leaf disease detection. And the training and validation loss was 0.0233112 and 0.1905328. The training performance of the MobileNetV2 was not better than the DenseNet201 model on tomato leaf disease detection.

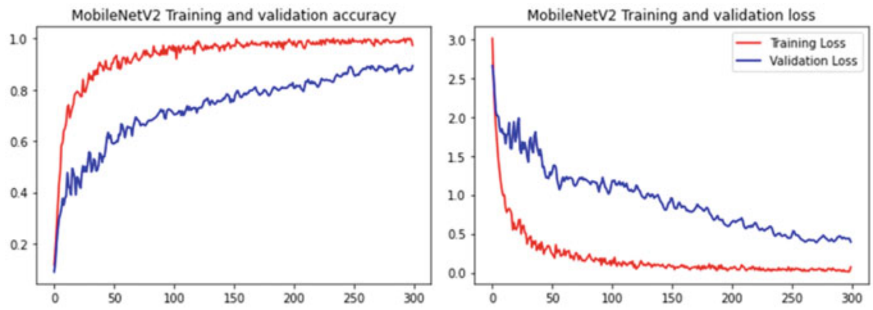


Fig. 9 Training and validation performance of MobileNetV2

2.9 *NASNetLarge*

The NASNetLarge is the recent classification model with large size of the saved model [16]. The model was used to classify the data, which requires more convolutional layers and filters. The input image shape is 331 pixels on width and 331 pixels on height; it was higher than other models discussed in this research. This model trained 334,841,692 parameters on the tomato leaf desire dataset. The batch size of the training data was reduced to 8 for the training process. The training and validation progress of the model is shown in Fig. 10.

The model trained 300 epochs for 12,600 s on the GPU. The training and validation accuracy was 0.978 and 0.834, respectively. And, training loss and validation loss of the model on the tomato leaf disease dataset were 0.05908 and 0.64814, respectively. The training and validation results of the NASNetLarge are not better than the DenseNet201 based tomato leaf disease detection dataset.

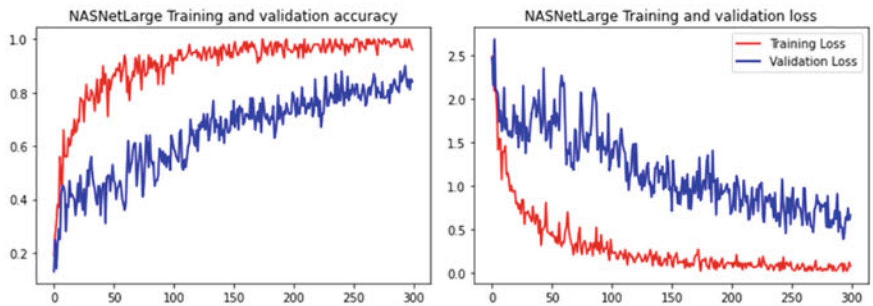


Fig. 10 Training and validation performance of NASNetLarge

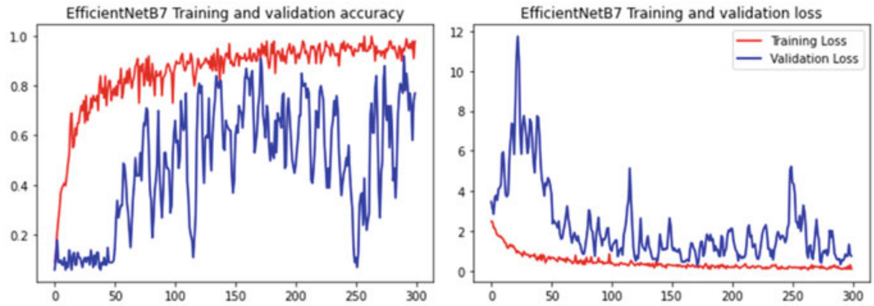


Fig. 11 Training and validation performance of EfficientNetB7

2.10 EfficientNetB7

The EfficientNetB7 is the recent version of the EfficientNetB series of networks [17]. The EfficientNetB7 is also the modern classification network with improved convolutional layers. The input shape of the network is 221*221*3 pixels. The model trained around 128,457,377 parameters on the tomato leaf disease dataset. The batch size of the dataset for the model was 16 and training epochs was 300. The training time of the model was about 2700 s. Figure 11 illustrates the training performance of the EfficientNetB7.

The training accuracy and loss on the tomato leaf disease dataset were 0.958 and 0.15376, respectively. The validation accuracy and loss of the model on validation data were 0.716 and 0.86848, respectively. The training and validation performance of the EfficientNetB7 is not better than the DenseNet201 model. The validation and testing performance of the trained transfer learning models was compared in the subsequent section.

3 Results and Discussions

This section compares the validation and testing performance of the trained transfer learning based models on tomato leaf disease detection. Initially, the transfer learning based models are trained on the training dataset up to 300 epochs. After the validation process, the trained models were tested on 300 unseen test data. The validation and testing performance of each model was compared in Table 1.

The result shows that the DenseNet201 was superior to the other transfer learning based models. And there was a difference of the accuracy around 2% of the following performing InceptionV3 Net based model. Based on the testing performance of the DenseNet201, the training epoch of the model was extended to 1000 epochs. The training process of the DenseNet201 based model took around 8340 s. The training

Table 1 Validation and test performance comparison

Model	Validation loss	Validation accuracy	Test accuracy
VGG16 Net	0.211	0.9386	0.9264
VGG19 Net	0.1861	0.9413	0.9312
InceptionV3 Net	0.1836	0.946	0.937
ResNet152V2	0.3088	0.9146	0.9014
DenseNet201	**0.0547**	**0.976**	**0.9624**
InceptionResNetV2	0.2156	0.9413	0.932
Xception Net	0.1897	0.9438	0.9328
MobileNetV2	0.1904	0.9443	0.9332
NASNetLarge	0.6481	0.834	0.8354
EfficientNetB7	0.8684	0.716	0.7616

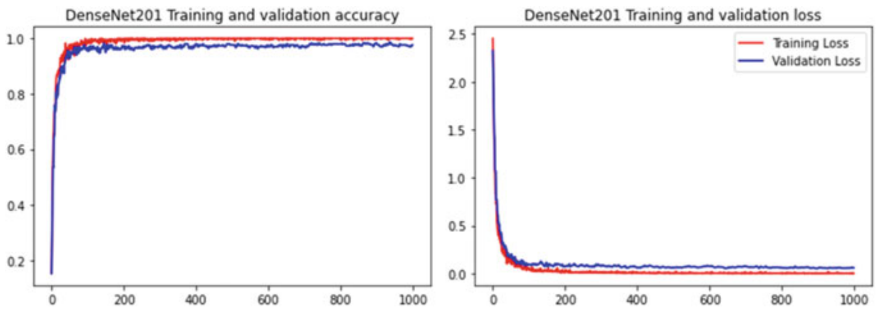


Fig. 12 Training and validation performance of DenseNet201 in 1000 training epochs

and validation performance of the extended DenseNet201 based model is shown in Fig. 12.

The validation accuracy of the extended model was 0.9933 (99.33%) and validation loss was 0.00214 on validation data. The model was tested on the test set after the completion of the training process. The test accuracy of the extended model was 0.9831 (98.31%) on test data. The DenseNet201 based model produced superior results on the test data of the tomato leaf disease detection dataset.

4 Conclusion

This research focused on ten different deep transfer learning techniques for the classification of diseases in tomato leaves. The transfer learning based models were trained using the training set of the tomato leaf disease dataset. After the 300 epochs of training, the validation performance of the transfer learning techniques was compared. The comparison result shows the validation accuracy and loss of the DenseNet201 were 0.976 (97.6%) and 0.05476, respectively. Also, the testing accuracy of the DenseNet201 model on the test dataset was 96.24%. The performance of the DenseNet201 on validation data and test data was superior to other transfer learning techniques. So, the training of the DenseNet201 was extended to 1000 epochs on the training dataset. After the completion of 1000 epochs, the testing accuracy of the DenseNet201 was 98.31% on the test set. The testing accuracy of the DenseNet201 was better than other transfer learning models. The research concludes that the DenseNet201 on tomato leaf disease classification task with minimal augmentation is higher than other transfer learning based models.

Acknowledgements The authors are thankful for the infrastructure support of Vel Tech Technology Incubator from Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai.

References

1. Abbas A, Jain S, Gour M, Vankudothu S (2021) Tomato plant disease detection using transfer learning with C-GAN synthetic images. Comput Electron Agric 187:106279. https://doi.org/10.1016/j.compag.2021.106279
2. Chen X, Zhou G, Chen A, Yi J, Zhang W, Hu Y (2020) Identification of tomato leaf diseases based on combination of ABCK-BWTR and B-ARNet. Comput Electron Agric 178:105730. https://doi.org/10.1016/j.compag.2020.105730
3. Wspanialy P, Moussa M (2020) A detection and severity estimation system for generic diseases of tomato greenhouse plants. Comput Electron Agric 178:105701. https://doi.org/10.1016/j.compag.2020.105701
4. Mirzaee H, Neira Peralta NL, Carvalhais LC, Dennis PG, Schenk PM (2021) Plant-produced bacteriocins inhibit plant pathogens and confer disease resistance in tomato. N Biotechnol 63:54–61. https://doi.org/10.1016/j.nbt.2021.03.003
5. Gonzalez-Huitron V, León-Borges JA, Rodriguez-Mata AE, Amabilis-Sosa LE, Ramírez-Pereda B, Rodriguez H (2021) Disease detection in tomato leaves via CNN with lightweight architectures implemented in Raspberry Pi 4. Comput Electron Agric 181:105951. https://doi.org/10.1016/j.compag.2020.105951
6. Geetharamani G, Pandian JA (2019) Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Comput Electr Eng 76:323–338. https://doi.org/10.1016/j.compeleceng.2019.04.011
7. Pandian JA, Geetharamani G, Annette B (2019) Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques. In: Proceedings of the 2019 IEEE 9th international conference on advanced computing, IACC 2019. https://doi.org/10.1109/IACC48062.2019.8971580
8. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv1409.1556
9. Thankaraj S, Khilar R, Christo MS (2021) A comparative analysis on plant pathology classification using deep learning architecture—Resnet and VGG19. Mater Today Proc. https://doi.org/10.1016/j.matpr.2020.11.993
10. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 2818–2826. https://doi.org/10.1109/CVPR.2016.308

11. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision—ECCV 2016. Springer International Publishing, Cham, pp 630–645
12. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 2261–2269. https://doi.org/10.1109/CVPR.2017.243
13. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: Proceedings of the thirty-first AAAI conference on artificial intelligence. AAAI Press, pp 278–4284
14. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. https://doi.org/10.1109/CVPR.2017.195
15. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) MobileNetV2: inverted residuals and linear bottlenecks. https://doi.org/10.1109/CVPR.2018.00474
16. Zoph B, Vasudevan V, Shlens J, Le Q (2018) Learning transferable architectures for scalable image recognition. https://doi.org/10.1109/CVPR.2018.00907
17. Tan M, Le Q (2019) EfficientNet: rethinking model scaling for convolutional neural networks

An Effective Analysis of Cryptography and Importance of Implementing Cryptography in Fuzzy Graph Theory

Anu Pius and D. R. Kirubaharan

Abstract The study intends to give the essential importance of fugitive graphic hypothesis concepts and applications of dominations in fugitive graphs to distinct real conditions in cryptographic areas. Due to numerous uses in PCs and communication, biomedicine, science and atomic material, interpersonal organizations, natural sciences and different regions, this is an outstanding growth. In this work we provide the ideas of maximum products for fuzzy graph structures, regular furious graphic structures and provide examples and characteristics for these notions. We also give the degree and full degree of a vertex in the maximum product of fuzzy graphical structures and use cryptography to explain several of their features. In addition, we construct a flowchart to demonstrate the general method for applying the fused graph structure, in respect of the cryptography implementation in fuzzy graphs. Graph theory concepts may be applied in various fields of mathematics. A fuzzy set with an interval is an extension of the idea of a fuzzy set. In comparison to fuzzy models, interval-rated fuzzy models provide the system better precision, flexibility and compatibility. In this article, we present in fluffy graph the idea of encryption to secure fluffy computations.

Keywords Fuzzy · Graph · Cryptography · Encryption · Decryption · Set · Lattice · Notions

1 Introduction

Test in many situations, the uncertainty defining judgments and the decision-making process based on statistical rationale may be linked back to the absence of inexact information resulting from the ambiguity in the data being evaluated. Fuzzy set theory comes to the fore in that sense and plausibly plays a significant role as predicted.

A. Pius (✉) · D. R. Kirubaharan
Department of Mathematics, PRIST Deemed to be University, Thanjavur, India
e-mail: anupius06@gmail.com

D. R. Kirubaharan
e-mail: jagkans.dr@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_26

A quick overview of the random variables (which are also known as stochastic variables), their characteristics and the basic statistical ideas of point's assessment, range estimate, hypothesis testing and regression will be provided prior to introducing the fuzziness [1]. They will then introduce their fuzzy analogues. This study provides several key characteristics. Simultaneously an independent definition of fluctuating graphs was offered. The "edge connectivity and vertex connectivity" of fuzzy networks were investigated and exploited in furious graphics clusters. This continued and offered various connectivity-related techniques [2].

They also examined the automaticity of foggy diagrams, fuggy vertices and geodesics in fugitive diagrams. Several characteristics have been developed of fuzzy trees and blocks. They also explored the supplement of fluorescent graphs and various matrices in fluorescent graphs. Identified and presented a method for different kinds of edges in fuzzy graphs. They have effectively described numerous fluffy graphic structures such as fluffy trees, blocks and full fluffy graphs utilizing this identification. In 2010, connectivity of fuzzy edge and vertex was launched. The connection parameters were generalized [3]. The main contribution of this paperwork is current fluid graphic theorems that include saturated and β-saturated cycle as well as fumigation graph complements, which have arbitrary connectivity values and calculate fluid edge connectivity parameters for certain subcategories of fluidized graphs [4].

New values were also calculated for fuzzy cycles, fuzzy trees and entire fuzzy graphs. Some theorems and strategies of existence are also discussed in the wider scenario. At the very beginning is the presence of a super-fuzzy chart with a more fluid edge connection than the fluid chart and the formation of a fluid chart on a t-edge. An example of the clustering of the new parameters in fuzzy diagrams shows that the existing ones are superior to the current ones. The first document in fuzzy graph theory dealt with the topic of clustering utilizing the vertex connection idea. Fuzzy versions of most ideas may be found in graph theory [5]. Cut vertices, bridges, blocks, etc. are widespread. As an intensive area of research nowadays, fuzzy graph theory has grown. Connectivity is the most essential and relevant term in this field. There are numerous traditional problems in network theory that use diverse connection characteristics, such as maximum bandwidth issues, bottleneck difficulties, quality concerns, and so on. No such publications contain adequate generalizations of graph connection parameters [6]. It might be noted that. This is the reason for this article. The writers broaden and compare the fundamental graphic notions to the classical ones. Figure 1 provided concepts based on strong edges of fluidized graphs did not directly generalize cut vertices and bridges in graph theory. In this article, we fix this problem by introducing new definitions for vertex connection and furrowing chart connectivity.

Classical random variables and their basic characteristics have been examined. In order to manage fluid data or observations, it is possible to add fluidic random variables to understand vagueness, in other words randomness and vacuity may now occur at the same time [7]. There are many approaches to the idea of a fuzzy random variable in the literature, especially the definitions (mathematically similar) provided. In what follows, we will use the Kruse-Meyer method by examining a flush random variable as a fluffy perception/observation of a traditional random variable that is

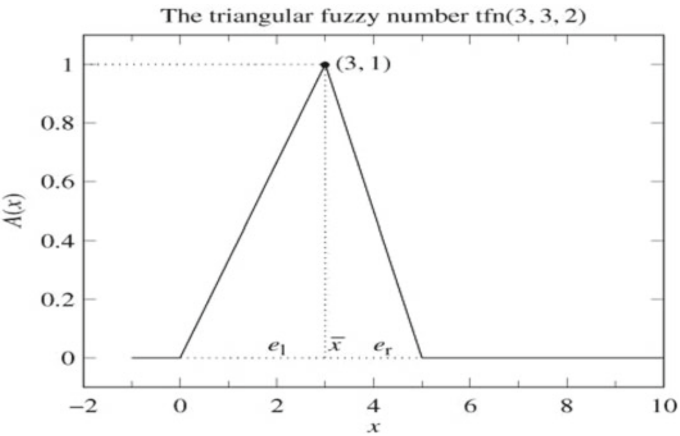


Fig. 1 Triangular fuzzy graph

evaluated in real terms. This method is a mixture of the concerns of the other writers. However, we highlight that for fuzzy random variables we will not introduce the ideas of the expected value, variance, and distribution function. Back to Tanaka's method, it should be emphasized that it can only be used for linear functions [8]. However, the computing is extremely easy, but the fuzzy least-square process has an advantage over Tanaka's approach, in that it maintains the degree of fluidity between observations and estimates to a minimum and therein alludes to some criticisms of the Tanaka model). This possibility model of regression is based on the notion of minimizing fluidity by minimizing the overall support of the fluid coefficient of regression subject to the integration of all observations [9]. In short, the given statistical regression is based on crisp random mistakes, whereas the rotten regression is based on fluid errors. In the event of a fluid environment, the distinction between these two types of uncertainty has led to the need to expand statistical regression. Approach is one of two key techniques to fizzy regression, notably the so-called optional regression analysis based on the idea of option and the approach known as the fuzzy least square method, which seeks to minimize errors between the observed and estimated data. An enhanced S-box is suggested. The downside of their technique is that the existing AES framework cannot deploy their S-Box.

The rationale for the AES S-low box's complexity is revealed. In addition, a new S-box version. Another S-box based on refined mapping with a polynomial content of 253 non-zero terms is created and utilizes the AES implementation. The algebraic complexity of the AES S-box is greatly boosted with the creation of a new S-box based on grey codes. The upgraded S-box consists of 255 non-null terms and uses the complete current AES architecture [10]. A generalization of a Gray S-box: GF (28) is utilized to produce 256 various RTS, including non-linearity, bit independence, stringent law-by-section approaches, differential roughness, algebraic complexity, correlation and histogram, with a highly algebraic complexity. The GF (28) is correct and standard translation. Like the Gray S-box, RTSs are also compatible with the existing implementation of the AES S-box.

In addition, the latest computer assaults such as linear, differential and algebraic attacks may successfully be met by cryptosystems based on multiple S-boxes in comparison to a single S-box security system. The requirement for digital picture security has grown due to the extensive Internet transmission of diverse image files. Therefore, some methods to protect secret pictures are required. The researchers offer many sorts of picture security methods. Cryptography and steganography are two frequently used ways to secure hidden picture material. The fundamental idea of cryptography is to turn a secret image (plain picture) into a dispersed image (counter image).

2 Literature Review

Arya Sebastian [2]: Arya Sebastian In network theory, graphic models are essential. But weight standardization is needed to deal with big networks such as the Internet. Most of the study papers in the literature were confined to algorithms alone. Not much was theoretically researched about the connectedness of normalized networks. Fuzzy theory of graphs addresses most issues in this field. Although the idea of connectedness has been thoroughly explored in fuzzy graphs, there are no suitable generalizations of unweighted graphics' connectivity parameters. This article aims to generalize in graphs some of the current vertex and edge connectivity features. New parameters are compared to previous ones and generic values are obtained for some essential types of such cycles and trees in fluorescent graphs. The presence of super-fuzzy graphs with larger connection values is established for both old and new parameters. For some bigger groups of flush charts, the new edge connectivity values are also reached. The generalization offers significant progress in fluid graphical techniques to clustering and permits theoretical harmonization. A novel family of flush graphs is also studied, called broad fluffy t-connected graphs. A technique for clustering the vertices of a fugitive diagram and a trafficking application is also demonstrated.

This work presents the vertex regular fuzzy graph, the whole degree and the completely vertex regular fluctuating graph. Vertex regular fuzzy diagram and vertex regular flushed diagram are contrasted through different instances and different attributes are presented between the grade and vertex. A required and adequate condition is stated under which they are equal. Some characteristics of the regular vertex fuzzy graph are explored and regular vertex fuzzy graphs are analyzed. In this work we addressed the notion of regular fuzzy vertex graphs and regular fuzzy vertex graphs. A comparative research of regular fuzzy vertex graphs and a completely regular fuzzy vertex graph is performed. Some results on regular vertex fuzzy graphs are also reviewed and examined whether they hold regular fuzzy graphs for the entire vertex. In 1736, Euler initially presented the notion of graph theory. L.A. Zadeh addressed the perception of fuzzy sets in 1965. In 1973 Kaufmann presented the first concept of a fluid graph based on the fuzzy relationship between Zadeh. A. Rosenfeld considers fluid relationships on fluid sets and develops fuzzy graph theory.

3 Proposed Methodology

The substitution box (S-box) is exclusively responsible for generating image dispersion in various cryptosystems. A block cipher is utilized as an Advanced Encryption Standard by the National Institute of Standards and Technology (AES). Currently, the AES cryptosystem is frequently employed. Due to the essential function of the S-box in AES, a large number of cryptographers have studied the AES S-box. The AES S-box is given a new basic mathematical explanation. A polynomial representation of the AES S-box through permutation. The polynomial of the AES S box has only 9 non-zero terms, indicating that AES safety is suspicious of computer assaults. Many academics have suggested improved S-boxes to eliminate this AES vulnerability. Biometric inputs have a crucial role in contemporary cryptography applications.

Because each interpretation of a biometric identification entails a specific allowed degree of mistake, applying fluid logic can be an appropriate option for controlling such a difference between reading and limiting or extending error tolerances. This is one of the major reasons for designing a Fuzzy Identity-Based Encryption or Fuzzy IBE. It is crucial to emphasize that for systems to be safe, the ownership of an identity must be guaranteed. This can be accomplished, for example, by a skilled operator. The Decisional Die-Hellman Bilinear (DBDH) problem is a variation of the Diffie Hellman Decisional problem (DDH). DDH is a cornerstone of security evidence for various cryptographic systems and is used as a premise of computer hardness. This is the assumption of DBDH and the assumption of Decisional Bilinear Die-Hellman (DMBDH). DMBDH is a significantly modified DBDH variant that is utilized in our safety evidence Fuzzy Selective ID Security Model that is a significantly modified version of the IBE Selective ID Security Model.

The original selective ID model was proposed initially in the work and was also utilized in his work. The Fuzzy Selective-ID model of the opponent can only query secret keys for identities that overlap less than d with a target identity. The Fuzzy Selective ID Model is the following game for a FIBE system. In the field of digital fingerprinting, collusion attacks are explored. The combined e-collisional attack by a number of assailants is an attempt to erase the digital fingerprint and produce a replica of the original data. This type of attack may also apply to identity-based encryption systems. Random Oracle (RO) is a hypothetical arithmetic function, believed to map the incoming questions equally and randomly distributed. Since there is no hash function which generates a genuinely random map, a random oracle is a black box. In cryptography, the Random Oracle concept, as set forth. One of the complete ways of accomplishing authentication, key distribution and non-energy using public-key cryptography is a public-key infrastructure (PKI). Simply, what a PKI accomplishes is to offer users with the mechanism that ensures that published audiences are authentic. This objective is realized by certificates and certificate authorities (CAs). A certificate

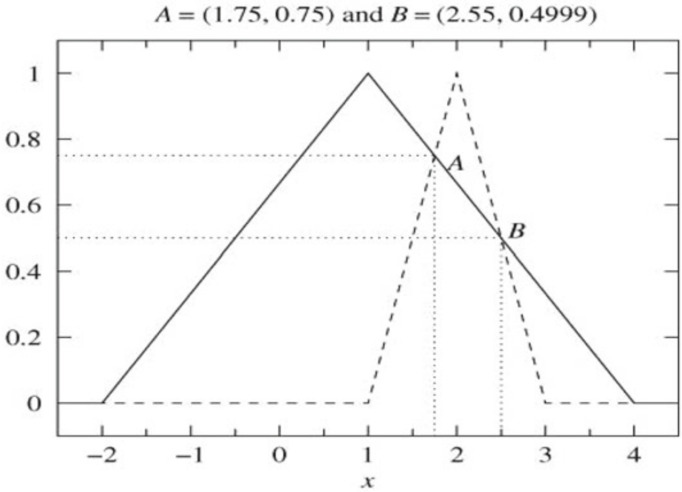


Fig. 2 Cryptography in fuzzy graph

contains digital information for the authenticity verification of a public key. A CA is a trustworthy entity that issues and checks certificates for a particular user by signing them. For example, when Alice wishes to interact with Bob, the identification certificate of Bob is publicly available to her. Alice may decode the certificate with CA's public key and get Bob's identification information together with Bob's public-key.

Now Alice may trust Bob's public key's validity and establish a secure connection. Figure 2 everything is predicated on Alice's first faith in the CA. X.509 is the standard established as the public key infrastructure and privilege management infrastructure by International Telecommunication Union Standardization Sector (ITU-T) (PMI). The idea of a relationship is a fundamental concept in both mathematics and everyday life. For example, if you have a series of keywords and papers, you may link the keywords with the documents and that is a relationship. Fuzzy relationships spread crisp relationships much like fuzzy sets spread crisp sets. If pictures and images are keywords that define document images, one may argue, in a smooth theoretical environment, that these keywords characterize the document to some extent. To have a better grasp of the notions concerning fluid interactions, a thorough understanding of crisp relationships is needed. In the following, we shall present the important topics succinctly. For a more thorough exposition, readers might consult any book on relationships. The appropriate first-order language is provided for all three fuzzy logics (i.e. a corresponding predicate calculus). Below each of these logics lies a language which consists of an unemptied set of predicates and a set of object constants (maybe empty) (i.e. something that we presume or hypothesis to exist in the world). Without exaggerating, pure mathematics can be said to be the only area where humans precisely reason. We are reasoning in approximate terms in all other domains. Since fuzzy set theory involves vagueness, in all these instances one may use it to reason where we approximately reason.

Furious constraints to suggest tangible instruments that would enable one to reason mechanically. We have, however, provided a series of flippant logics and, clearly, these tools should be reformulated in those new languages. Thus, we offer the basic concepts of Zadeh first and then we present their reformulation in various logical languages. The term “logic” is derived from the Greek word “logos,” which meaning “reason.” “Logic” Logic is actually another ancient Greek creation. The creation of logic is ascribed to Aristotle. Logic was formed when the structure of arguments began to be studied. Aristotle named all these forms syllogisms. The following is an example of a syllogism: On Wednesday, John goes to movies; today is Wednesday; thus John goes to movies. Logic concerns the process of obtaining a response by considering known information. And logic may be broken down into formal logic, informal logic, symbolic logic and mathematical logic in general. Thus, a user’s identity may decipher a publicly encrypted cipher text! 0. in this approach! And! 0 do not need to be identical to allow error-tolerant identity measurements.

The approach applies in particular to biometric cryptography. This is because biometric scans entail noise in the data, which means that identities are slightly different. In addition, a public-key infrastructure should not be supported, because the biometric identification itself might be seen as a public key for the owner. We know that a proposition is either true in mathematical reasoning (for example, “3 is an odd number”) or incorrect (e.g., “4 is greater than 5”), as nothing else makes sense. There are, nevertheless, propositions beyond mathematics that are neither true nor untrue.

4 Experiment Result

In the domain of communication, users are more focused on communication than data transmission security. Many cryptosystems were improvised to achieve communication efficiency. RSA cryptosystem is a widely recognized cryptosystem designed to safeguard information and protect communication by making encryption and decryption especially challenging for attackers. Since the safety of the crypto-system is necessary to secure communication, we present a novel RSA encryption system based on the theory of the fluctuations of the plaintext and cipher text in terms of the triangular fuzzy number; (TFN). The outcome of decryption reveals that the message received is the same as the original plaintext. This paper shows that fuzzy set theory may be utilized as an alternate method for safeguarding other cryptosystems. Data communication is data transfer from one location to another. Data communication security is becoming the major problem in data communication.

It can give a great encryption solution. The encryption algorithm is the data encryption mathematical procedure. The method suggested supports the user’s chosen level of security and processing. The technology provides security and its suitable processing levels by generating random keys for the encryption/decryption process.

This facility is obtained by using fluid logic. By comparing the findings with the other current encryption methods, the suggested encryption algorithm will be examined. The project aims to create a new method that uses fuzzy sets to be more sophisticated than existing encryption techniques. The method to prove the security of these three schemes is to show that if an adversary in polynomial time can break the scheme, the scheme itself can be utilized to solve a task which is supposed to be difficult, during the polynomial period.

In brief, we utilize Reduction ad Absurdum to illustrate a false premise (insecurity of a system) leads to an absurd consequence, which proves the safety of this scheme. As we have shown in this paper, the Fuzzy Identity-Based Encryption method has important advantages such as mistake tolerance. We also saw the growth of these systems leading to improved computer science, making them a step closer to practical applications in the real world. Another observation is that the security of the schemes depends directly on the DBDH assumption. There is no evidence available to demonstrate that the problem is difficult to solve, but there is no solution. Diffie-Hellman is considered as difficult as the Discrete Logarithm problem (DLP). The DLP may be solved by means of a polynomial quantum algorithm, which makes the issue part of the so-called bounded error Quantum Polynomial Time (BQP) complexity class. In this way, a quantum computer can solve the problem in the polynomial period. Since these developments are relatively fresh, new methods to this challenge would be intriguing, which is the cornerstone of many cryptographic techniques. If quantum computers are built in a realistically useful manner one day, any strong encryption method that we know in conjunction with the faith in them will last forever. All three systems mentioned here have comparable general characteristics and similar advantages. The most important aspect is the speed of key generation and encryption methods. All three methods may be utilized for both biometric and attribute-based encryption. Figure 3 the key aspect of the safety of any FIBE scheme in biometric applications is that a biometric operator must be able to verify that the user owns the biometric. This prevents so-called assaults of imitation.

The safety of these methods depends on this premise. Because a biometric is a part of the actual user, the public key is always available. Biometric characteristics are unique, prohibiting duplication of their derived identities. These methods enable a flexible fault tolerance through the use of fuzzy logic theory. They are additionally safe from collusion attacks since each private key's components are produced using another polynomial, preventing a new private key from being created from components of multiple private keys. Figure 4 classical approximation theory of true continuous functions using algebraic or trigonometric polynomials has been a topic of rigorous as well as informative presentation study for more than two centuries). Here, we will offer a quick explanation of several fundamental findings in a fluffy environment of approximation theory. The relevance of fluorescent calculus becomes obvious when you consider about the function of classical calculus principles in the mathematical representation of real-world occurrences. The complexity of nature forces us to compromise a lot, namely to formulate assumptions, assertions and premises, not only for building and applying mathematically tractable models, but also due to the imprecision. The fuzzy calculus provides a possible means of dealing

Fig. 3 Fuzzy output set

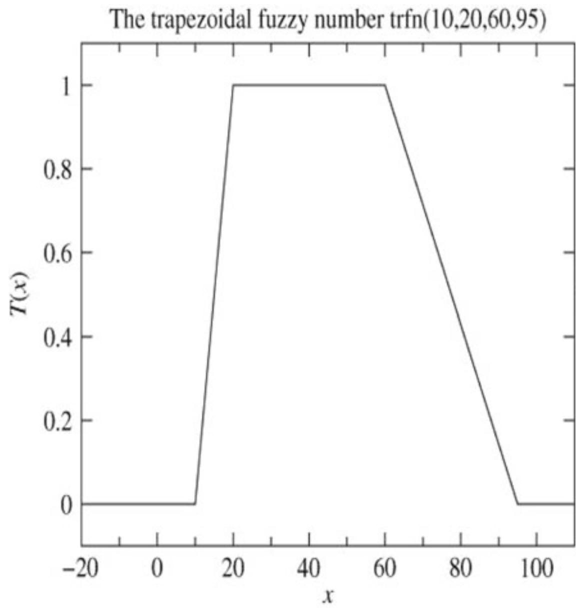
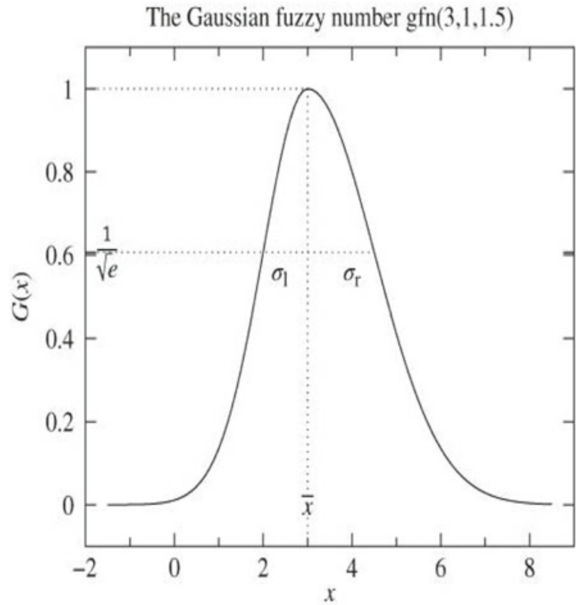


Fig. 4 Cryptography fuzzy set



with the aforementioned ambiguity of mathematical modeling. This chapter first sets up fuzzy functions and their characteristics and then examines their integration and distinguishing. We then explore a theory of flouted boundaries and we shall ultimately study flouted differential equations. Let us first recall briefly what the regular image-sided polygon is in geometry on Euclidean planes, which is a closed, rotating symmetric figure with images sides equiangular, and equilateral, while all of the vertices are noncyclic, i.e., lie on the circumcised circle. Of course, equilateral triangles are the polygons with the lowest number of sides. Each regular polygon really has a registered circle tangent at the middle of each side of the polygon. The circumscribed and the circles inscribed share a same center which is also the center of the polygon.

5 Conclusion

In modeling and managing dynamic interconnection networks, graph parameters are highly essential. An example is also shown which applies these principles in the field of cryptography when the authors work on sustainable development cryptography goals as part of a project. Graphical theory provides the biggest possible generalizations for vertex and edge connectivity and determines their values for some of the relevant subclasses. An example indicates that the new values are preferable to the previous ones and helps to locate more qualitative groupings in the same graph cluster. Super-fuzzy graphs and full flush graphs with pre-fixed connection values are available. The current parametric values for saturated fluid cycles and complements of fluid graphs are also calculated. In this article, several of the existing connection settings can be redefined. Mathematics experts have researched the widest routes of issues and bottleneck problems extensively in the past, but until recently, there were no mathematical theories to support them. Most principles in the current fuzzy graph theory may be applied directly to previously algorithmically establish issues.

References

1. Anjali N, Mathew S (2015) On blocks and stars in fuzzy graphs. J Intell Fuzzy Syst 28:1659–1665
2. Sebastian A (2020) Generalized fuzzy graph connectivity parameters with application to human trafficking. https://doi.org/10.3390/math8030424
3. Mathew S, Anjali N, Mordeson JN (2018) Transitive blocks and their applications in fuzzy interconnection networks. Fuzzy Set Syst, in press
4. Mordeson JN, Malik DS, Richards CS, Trebbian JA, Boyce MA, Byrne MBP, Cousino BJ (2016) Fuzzy graphs and complementary fuzzy graphs. J Fuzzy Math 24:271–288
5. Mathew S, Mordeson JN, Malik DS (2018) Fuzzy graph theory. Springer, Berlin, Germany
6. Mordeson JN, Mathew S, Malik DS (2018) Fuzzy graph theory with applications to human trafficking. Springer, Berlin, Germany

7. Binu M, Mathew S, Mordeson JN (2019) Wiener index of a fuzzy graph and application to illegal immigration networks. Fuzzy Set Syst 384:132–147
8. Binu M, Mathew S, Mordeson JN (2019) Connectivity index of a fuzzy graph and its application to human trafficking. Fuzzy Set Syst 360:117–136
9. Binu M, Mathew S, Mordeson JN (2020) Cyclic connectivity index of fuzzy graphs. IEEE T Fuzzy Syst. Accepted
10. Vengatesan K, Selvarajan S (2015) Maximize pair genes from microarray using the enhanced fuzzy clustering algorithm. J Pure Appl Microbiol 9:611–618
11. Ali S, Mathew S, Mordeson JN, Rashmanlou H (2018) Vertex connectivity of fuzzy graphs with applications to human trafficking. New Math Nat Comput 14:457–485
12. United Nations Office on Drugs and Crime (2014) Global report on trafficking in persons. United Nations Office on Drugs and Crime, Vienna, Austria
13. Mordeson JN, Mathew S (2017) Non-deterministic flow in fuzzy networks and its application in identification of human trafficking chains. New Math Nat Comput 13:231–243

Role of Emotional Intelligence on Final Year Students' Employability

Yasmeen Bano and S. Vasantha

Abstract *Background of the study* Emotional intelligence (EI) is the ability and personal skill of an individual to manage the emotion and enhance the effectiveness at the workplace. The theory of EI came into the existence when Daniel Golman published his book on Emotional intelligence after inspired by the Salovey and Mayer's work. He said EI includes ability to motivate one self, persistent and passion and self-control. This study provides in-depth understanding on the role of Emotional intelligence on employability of final year students. *Objective* The key objective of this research to is to assess and explore the role of emotional intelligence on the final year students' employability. *Methodology* The main focus of this study is to understand the role of emotional intelligence on final year students' employability. Questionnaire was distributed to 150 final year students and data collection was done by using purposive sampling technique. For testing hypothesis, correlation and regression statistics has been applied. *Findings* This study found that all the dimensions of EI were positively and significantly correlated with the final year students' employability. Further the study concluded that dimensions of EI (emotional intelligence) have direct positive effect on final year students' employability. *Implications* The objective of the study is to assess the dimensions of emotional intelligence on final year students' employability. Therefore, findings of the current study help intellectuals and higher education institutions to identify the challenges and issues of final year student' employability. *Originality* This study provides relevant evidence of effect of emotional intelligence on employability by measuring the quantitative data and results were consistent with each variable. It disclosed the dimensions of emotional intelligence that are positively and significantly associated with the final year students' employability. The present study consisted insight understanding of three key dimensions (self-awareness, motivation and communication) of final year

Y. Bano
School of Management Studies, Sathyabama Institute of Science and Technology (SIST), Chennai, India
e-mail: yasmeenbano.soms@sathyabama.ac.in

S. Vasantha (✉)
School of Management Studies, Technology and Advanced Studies (VISTAS), Vels Institute of Science, Chennai, India
e-mail: vasantha.sms@velsuniv.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_27

students' emotional intelligence to positively shape their career and to male them employable.

Keywords Communication · Emotional intelligence · Employability · Motivation · Self-awareness

1 Introduction

Higher institutions and universities have responsibility to provide professional skills, abilities and in-depth knowledge to their students for the employment. In simple word, universities and higher institutions should produce the students who can get employment after completion of their study. Lack of collaboration between higher institution and business world has put pressure on strategist and employers who make policies and process to develop the employability. The areas of employability are wide therefore researchers adopted several techniques to study the employability model. For instance, individual must have various types of skills and abilities to get employment.

As employability is defined as personal, individual and external factors, several types of skills and personal attributes. Many researchers claimed that students' employability is influenced by the various factors e.g., skill, abilities and individual differences.

Emotional intelligence (EI) is a new skill that employers place a high value on because it is seen to significantly impact self-motivation by processing and managing emotions, particularly in professional relationships and interpersonal communications assesses the ability to modify behavior in response to the mood of a co-worker, partner, or family member. Final year students' employability is influenced by the traits, values, emotions, philosophy and their unconscious motives. It also enhances and stables students' employability. As a theory of psychology, EI is the ability to perceive, assess, generate and understand the emotion of the individual to promote the intellectual growth. EI has seen as a key component to enable the students to succeed in their career and to make the strong personal relationships.

Personality factor of emotional intelligence has been observed as predictive power of behavioral process and work outcome. Emotional intelligence is seen as skill and traits in the career of individuals [1]. Employability is considered as the individuals' values, emotions, traits, unconscious motives, philosophy and sense of calling. Dacre and Qualter [2] have proven emotional intelligence as an important predictor of employability.

The study has identified three dimensions of EI for strengthening the final year students' employability because most of previous studies have focused to the concept of EI that has been associated with students' employability but there is still some clarification required on employability. Therefore, this study explores the dimensions of emotional intelligence. This study also examines relationship between dimensions of emotional intelligence (EI) and final year students' employability.

2 Objectives

To explore the dimensions of emotional intelligence among the final year students' employability.

To assess the relationship between dimensions of emotional intelligence and final year students' employability.

3 Literature Review

Several researchers claimed that in twentieth century EI was known as IQ (Intelligence quotient) but in twenty-first century, IQ is replaced as EI. EI is defined as set of skills and personal attributes including personal, social and survival aspects of intelligence that are critical for effective performance. The famous psychologist salovey and Mayer worked on their milestone article "Emotional Intelligence" in 1990, in which they considered emotional intelligence as form of "Intelligence" that defined EI is the ability to observe and understand oneself and other emotions and feelings. He added effective emotional skill helps in gaining personal and professional achievements. Another famous researcher of emotional intelligence, Bar-On [3] who introduced the concept of "Emotional Quotient" defined EI is the ability of understanding one's own self and others, coping with other and immediate surroundings to become successful in their career and profession.

Employers prefer the candidates or freshers who possess higher level of EI. A person with effective EI has more chances of to achieve their goal than a person with less EI, because a person with high EI can perform better than others and also can manage the work better under pressure. Cherniss and Goleman [4] said that employers while recruiting engineering students prefer employable skills as intellectual and personal attributes that include better communication, subject knowledge, self-management, leadership quality and problem solving skill. Researchers considered that final year engineering students should prepare for the social, emotional and intellectual learning to avoid the negative performance. But it has been noticed academics and universities put more focus to make them technically experts, they are not concerned about the emotional concept of the students. Employability is influenced by the values, emotions, traits, philosophy and unconscious motive of the person. It also has the effect on individual's career and better employment [4].

3.1 Proposed Conceptual Model of the Study

Emotional intelligence (EI) framework was developed by Daniel Golman in 1995, it is verified mechanism to measure the emotional intelligence. The study is based on the proposed framework which is based on the Golman's model and presence the direct

Fig. 1 Proposed conceptual model

effect of dimensions of emotional intelligence on final year students' employability. The study examines emotional intelligence with the help of 3 independent variables e.g., Self-awareness, Motivation and Communication (Golman's EI competencies) and dependents variable was final year students' employability [5] (Fig. 1).

3.2 Self-Awareness

Self-Awareness demonstrates the importance **of identifying one's own feelings and performance** and also defined as key of recognizing one's own strength and weaknesses. It helps in the understanding of persons, their own selves, and the recognition of their potential, as well as the work which is required for them. Self-awareness also gives people a clear picture of their job and careers [6].

3.3 Motivation

Motivation is challenge to do beyond the expectations. Locke and Latham suggest people do perform better when people have high and challenging goals. Locke and Latham [7], challenging goals produces high motivation that enhances the performance and he also argues that specific goals enables individual to perform better and monitor their progress as needed. Motivation helps to improve students' employability and proactive career behavior. Locke and Latham [7] explain motivation prepare students engaging in practical behavior contribute and enhance the future employability. Further, he added examine the motivational processes that have more influence on students' employability [7].

3.4 Communication

Communication is the process of exchanging information from the one to another person via non-verbal and verbal manner. Communication is an essential skill for inter-professional practice. Communication skill is identified as important elements in the educational field and personality development. Universities should provide effective communication environment to their student for better academic performance. Hendon et al. [8] showed the positive influence of EI on communication, interpersonal skills and job performance. Effective communication develops the confidence and increases the employment opportunities. Communication in the list of skills that influences employability. Communication is not only a good quality for students to have an employment, it is essential.

3.5 Employability

Employability is the ability or capability to sustain in the labor market. For individual, employability refers the knowledge, skills and abilities to possess the employment. Researchers claimed that students' employability is influenced by the various factors e.g., skill, abilities and attitude. Zhang and Zou (2013) stresses on employability skills such as teamwork skills, innovative skill, problem handling capacity, interpersonal, leadership and presentation skill. He believed these skills as most important skills that a student should have. Previous literature explores the factor affecting employability and suggested academicians should develop the policy that can furnish the students need.

4 Methodology

This study focused on the role of emotional intelligence on final year students' employability. The study has surveyed 150 final year engineering and MBA students by using purposive sampling technique. Descriptive research has been followed for this study.

4.1 Respondents and Data Collection

The population of this research was final year engineering and MBA students from the various colleges and universities. Questionnaire was distributed to 150 students and data collection was done by using purposive sampling technique. Respondents were 83 male students and 67 female students. The questionnaire consisted 23 items.

Each items were in range from strongly disagree-1, disagree-2, neutral-3, agree-4, strongly agree-5. The Cronbach's alpha for overall variable was 0.70.

To examine the relationship between dimensions of emotional intelligence (EI) and final year students' employability, the study has performed Pearson correlation and regression analysis. The study consists three independent variables e.g. Self-awareness, Motivation and Communication and one dependent variable e.g. Final year students' employability.

5 Results

Table 1 determines the correlation between the study variables Self-awareness, Motivation and Communication. Correlation analysis proved that all the variables of the study are significantly correlated at 1% of significant value; therefore, alternative hypothesis (H1) is accepted (Tables 2, 3 and 4).

The study has taken three predictors (Self-awareness, Motivation and Communication) to examine whether they are statistically significant. The average class size (**Self-awareness**, $b = 0.358$, $p = 0.000$) is significant and highly related ($p = 0.000$), so, it indicates the direct relationship between self-awareness and employability of the final year students. Next, the effect of **Motivation** ($b = 0.130$, $p = 0.018$) is significant that indicates Motivation has direct positive effect on final year students' employability. The **Communication** variable ($b = 0.328$, $p = 0.000$) is also highly related and indicates statistically significant positive effect on employability of the final year students.

Table 1 Correlation between the study variables e.g. self-awareness, motivation, communication and employability of the final year students

	Self awareness	Motivation	Communication	Employability
Self awareness		0.425[a]	0.471[a]	0.617[a]
Motivation			0.320[a]	0.432[a]
Communication				0.592[a]
Employability				

[a]Significant at 0.01 level

Table 2 Regression analysis to show the influence of self-awareness, motivation, communication on employability of the final year students

Model	R	R Square	Adjusted R square	Std. error of the estimate	Durbin-Watson
1	0.719[a]	0.516	0.507	0.46030	2.162

[a]Dependent variable: Employability

Table 3 ANOVA influence of self-awareness, motivation, communication on employability of the final year students

Model		Sum of squares	d*f*	Mean square	*F*	Sig
1	Regression	33.040	3	11.013	51.980	0.000[b]
	Residual	30.934	146	0.212		
	Total	63.973	149			

[b]Significant at 0.05 level

Table 4 Coefficient influence of self-awareness, motivation, communication on employability of the final year students

Model		Unstandardized coefficients		Standardized coefficients	*t*	Sig
		B	Std. error	Beta		
1	(Constant)	0.699	0.248		2.816	0.006
	Self-awareness	0.358	0.065	0.380	5.506	0.000
	Motivation	0.130	0.054	0.154	2.400	0.018
	communication	0.328	0.059	0.364	5.510	0.000

6 Discussion

Employability is influenced by the values, traits, emotions, philosophy and unconscious motives of an individual. It also enhances the chance of good and stable employability. The purpose of this study is to examine the role of emotional intelligence on final year students' employability.

EI is one of significant predictor and key element that influences students' employability. Goleman [5] says with strong justification that EI is a prime skill and quality that makes an individual employable.

H1: There is a significant relationship between the study variables Self-awareness, Motivation, Communication and final year students' employability.

This study proved the relationship between dimensions of emotional intelligence (EI) and final year students' employability that means H1 of this study is supported. Emotional intelligence refers ability to perceive and generate the emotions, understand the emotional knowledge, regulate the emotions to promote the intellectual growth. Mayer et al. (2004), EI is the capability to recognize and express thoughts, emotions and reason with the emotions. Emotional intelligence (EI) has become more prominent in the workplace and also demonstrated employers demand and prefer those graduates who acquire high level of emotional intelligence.

H2: There is a significant relationship between Self-awareness, and final year students' employability.

The study found direct and significant relationship between self-awareness and employability of the students; hence H2 of this study is accepted. According to Knight and Yorke [9] employability is influenced by self-awareness of the students regarding their capacity and learning skill that reflect on action. If students are more self-aware, motivated and confidence they can make better decision for their life and career. Graduates should achieve skill, abilities, motivation, self-confidence and self-awareness for the life, not just for the workplace. Lees [10] self-awareness, self-confidence and self-efficacy are more important than individual competence [9].

H3: There is a significant relationship between Motivation and final year students' employability.

H3 of this study is also accepted and indicated that motivation is positively associated with the employability of the students. Employers are more concerned of ensuring individuals right skills, flexibility and motivation in order to achieve and maintain the competitive advantage. Fugate et al. [11] motivational processes are very critical contributor for employability and personal adaptability.

H4: There is a significant relationship between Communication and final year students' employability.

This study examined communication and indicated that it has direct positive effect on employability. Therefore, H4 of this study is accepted. Effective communication develops confidence and enhances employment opportunities of the individual. Oral communication skill is most important skill and very critical attribute for the entry level job candidates. He also stresses on the five classical managerial functions of and said these managerial functions are highly dependent on the effective communication. Peterson [12] leadership, decision making and communication are the highly influential skill in hiring decision.

7 Implications

The objective of the study is to assess the dimensions of EI on final year students' employability. Hence, findings of the present study help intellectuals and higher education institutions to identify the challenges and issues of final year student' employability.

The finding of the research shows that dimensions of the EI are the significant predictor of the final year students' employability. The findings of the study can be supportive for designing the course curriculum in academic field. The research findings of the study add values to the employers and recruiters to assess the students' capability.

8 Suggestion for Future Research

This study has applied quantitative approach to establish the relationship between study variables which is widely acceptable in social sciences studies. But sometimes only quantitative approach will not be appropriate. Therefore, to better understand the concept and dimensions of emotional intelligence, future researchers can adopt the in depth interview method or qualitative studies.

This study has empirically proved that dimensions of emotional intelligence have significant effect on final year students' employability. The limitation of the study is, only three dimensions of emotional intelligence have been taken for the study. So there is a potential scope for future studies to consider more dimensions of emotional intelligence. Future studies may consider additional dimensions of emotional intelligence such as self-management, social awareness, stress handling.

9 Conclusion

This study concluded, emotional intelligence has seen as critical and highly influential component for the students' employability. Individuals with high level of EI enable them to succeed and form strong personal relationship. This study found dimensions of emotional intelligence (self-awareness, motivation, communication) as significant predictor of employability. Further, the study proved that person with high emotional intelligence can have desirable and attractive outcome.

The study finding shows the relationship among the study variable (Self-awareness, Motivation, Communication and employability). The findings indicate the significant positive relationship among the variables. The study also examines the association between the study variables. The strong association between emotional intelligence and final year students' employability has been shown and widely acknowledged in the previous literature. Furthermore, communication which is one of the dimensions of EI has seen as the key element for the employability.

The study found that self-awareness is positively associated with the final year students' employability. Previous studies also show that self-awareness is critical for the students' employability. Self-awareness helps to recognize about yourself because recognizing yourself makes you to understand who you are and what work is required for you. Moreover the study proved that motivation is also a critical dimension of emotional intelligence that is significantly associated to final year students' employability. In the previous studies, communication is seen as most important dimension of emotional intelligence to gain the employment. Concerning all three dimensions of employability, communication is most influencing factor of employability. Overall results of the study suggest that communication skill has an important role on final year students' employability.

References

1. Mayer JD, Salovey P, Caruso DR (2008) Emotional intelligence: new ability or eclectic traits? Am Psychol 63(6):503–517. https://doi.org/10.1037/0003-066X.63.6.503
2. Dacre Pool L, Qualter P (2013) Emotional self-efficacy, graduate employability, and career satisfaction: testing the associations. Aust J Psychol 65(4):214–223
3. Bar-On R (1997) Bar-one emotional quotient inventory: technical manual. Multi-Health Systems, Toronto
4. Cherniss C, Goleman D (2001) The emotionally intelligent workplace: how to select for, measure, and improve emotional intelligence in individuals, groups, and organizations. Jossey-Bass, San Francisco, CA
5. Goleman D (1995) Emotional intelligence. Bantam Books, New York
6. Pathak A, Tewari V, Shankar S (2018) Impact of emotional intelligence on employability of IT professionals. Manag Insight 14(1):14–21
7. Locke EA, Latzham GP (2013) Goal setting theory, 1990. In: Locke EA, Latham GP (eds) New developments in goal setting and task performance. Routledge, Hove, pp 3–15
8. Hendon M, Powell L, Wimmer H (2017) Emotional intelligence and communication levels in information technology professionals. Comput Hum Behav 71:165–171
9. Knight PT, Yorke M (2002) Employability through the curriculum. Tert Educ Manag 8(4):261–276
10. Lees D (2002) Graduate employability-literature review. LTSN Generic Centre, York, pp 1–23
11. Fugate M, Kinicki AJ, Ashforth BE (2004) Employability: a psycho-social construct, its dimensions, and applications. J Vocat behav 65(1):14–38
12. Peterson MS (1997) Personnel interviewers' perceptions of the importance and adequacy of applicants' communication skills. Commun Educ 46:287–291

A Custom Stacking-Based Ensemble Learning Approach to Predict Failure of Stripper Well

Smit Kumbhani and Vishesh Dharaiya

Abstract Prediction of equipment failure has always been a challenging task. Analytical and statistical approaches for prediction of equipment failure have been employed for a long time. Analytical approach is based on criterion, while statistical approach is data driven. Despite its accuracy, statistical approaches fail with large data entries having high dimensionality. Advanced machine learning techniques come to rescue. In this study, an effort has been made to predict failure of stripper well with classical machine learning algorithms followed by a custom stacking-based ensemble learning approach. Classical machine learning algorithms like Support Vector Machine, K-Nearest Neighbour, Logistic Regression, Gradient boosting etc. have been applied to predict failure instance. Micro F1-score has been selected as a measure of prediction accuracy. A novel custom ensemble machine learning approach has been implemented to obtain better prediction accuracy compared to previously applied algorithms. Proposed novel approach has successfully predicted classification case of failure with micro F1-score of ~0.9887.

Keywords Ensemble learning · Failure prediction · Stripper well

1 Introduction

Stripper wells are extensively used in oil and gas industries to extract natural resources from earth's crust. Mechanical failure of stripper wells is not an unfamiliar situation for oil and gas industries. A preventive maintenance approach is used to ensure continuous operation of oil extraction. Replacement of whole assembly or component is preferred over repair to avoid monetary loss in on-site maintenance [1]. Preventive maintenance is not always a successful approach and a significant monetary loss

S. Kumbhani (✉)
Department of Computer Science and Engineering, Marwadi Education Foundation's Group of Institutions, Rajkot, Gujarat, India
e-mail: smitkumbhani080@gmail.com

V. Dharaiya
Department of Materials Science and Metallurgical Engineering, Indian Institute of Technology, Hyderabad, Telangana, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_28

often follows. Solution to this problem is online continuous monitoring through sensors followed by data analysis to predict an instance of mechanical failure. A huge number of sensors create higher dimensionality, where statistical techniques show their limitation. A machine learning or deep learning-based approach can be efficiently used. Table 1 shows previous studies in prediction of mechanical failure of various equipment with machine learning approaches.

It can be concluded from a comprehensive literature review that study on stripper well failure prediction has yet to be conducted. Application of a custom ensemble approach is also a novelty. Dataset used for prediction can be accessed at a given link. Dataset contains 15 input parameters and a target column for instance of failure. Failure instance is shown in binary 0 and 1. 0 shows no failure and 1 shows otherwise.

2 Workflow

Figure 1 shows workflow of conducted study. Data collection is conducted using various data mining tools. Obtained dataset was processed further for data cleaning, which includes removal of NaN values and least influential parameters. However, it has been concluded that data does not contain any unnecessary parameters to remove. Data cleaning was followed by Exploratory Data Analysis (EDA). In EDA we analyzed the distribution of numerical datacolumns and performed phi_k correlation metric to find a correlation score between input and categorical target or ground truth label. So after the feature selection and based on that top highly correlated input columns have been selected for model building.

It has been decided to predict fault with several algorithms. However, this can be divided into two primary approaches. Classical Machine Learning (ML) approach includes working with conventional algorithms like logistic regression, K-Nearest Neighbours, Support Vector Machine etc. On other hand, custom ensemble learning approach is a novel method to be presented in this study. Further discussion will clarify difference between ensemble approach and a custom ensemble approach presented in this study [9].

3 Classical Machine Learning

Logistic regression can also be used for a classification. This machine learning algorithm is suitable for binary classification problems. Like Linear regression, weighted sums of the input is performed. However, it estimates the logistic probability of the result of weighted sum using sigmoid function. Usually it estimates the probability using sigmoid function. If the estimated probability is greater than threshold then it predicts class 1 and 0 if it is lesser than threshold.

Table 1 Literature review of equipment failure prediction using machine learning

S. No	Title (Year)	Author	Technique(s) used	Remarks
1	A neural network auto-associator for induction motor failure prediction (1996) [2]	Thomas P. et al.	Artificial neural networks	A novel candor system is designed for data acquisition and data analysis
2	Machine Learning methods for prediction failures of Hard-Drives: A Multiple-Instance Application (2005) [3]	Joseph F. Murray et al.	Naïve-Bayesian classifiers, Support Vector Machine	Satisfactory performance has been obtained
3	Global model for failure prediction for Rod Pump artificial lift system (2013) [4]	Yintao Liu et al.	Support Vector Machine algorithm	Prediction performance has been improved 65% compared to previously best-available model performance
4	Data driven approach to failure prediction for electrical submersible pump system (2015) [5]	D. Guo et al.	Support Vector Machine	Subject matter experts approved credibility, accuracy and reliability of prediction
5	Road Machinery fault prediction based on Big Data and Machine Learning (2019) [6]	Xue L. et al.	Back-propagation neural network	Prediction accuracy rate of 92% has been achieved
6	Prediction of Extubation Failure for Intensive Care Unit Patients Using Light gradient boosting machine (2019) [7]	T. Chen et al.	Light gradient boosting algorithm	Proposed technique outperforms ANN, SVM and LR predictions
7	Machine Learning Framework for predicting reliability of solder joints (2020) [8]	Sung Yi et al.	Multi-layered feed-forward neural network	Proposed model has accurately predicted thermal fatigue, which is important in industries

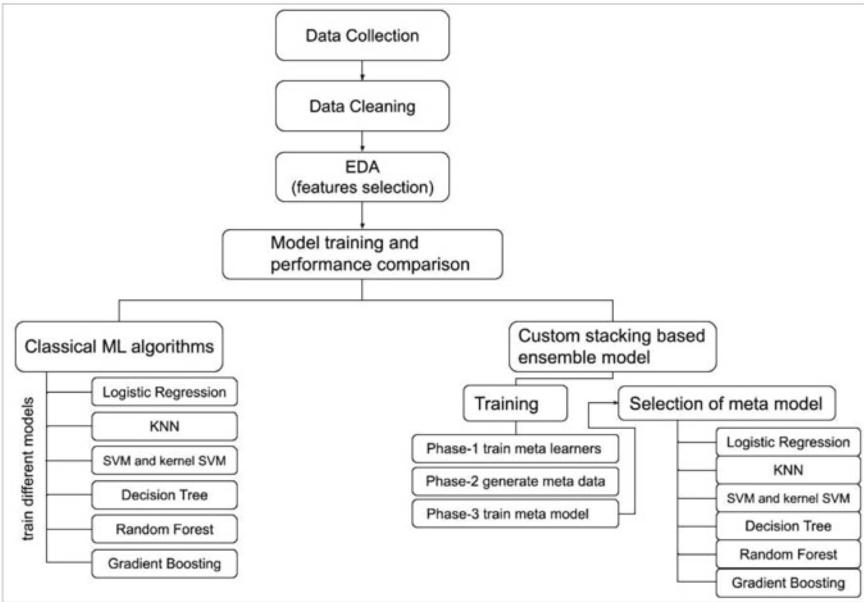


Fig. 1 Workflow

On the other hand, Support Vector Machine algorithm is the most popular and robust machine learning algorithm. It can work for both linear and non-linear classification. It works by drawing central linear line separating objects in classes. It uses other two lines (ve+ and ve−). Ve+ line intersect first positive datapoint parallel to central line and ve− line intersect first negative datapoint parallel to central line. And the distance between ve+ and ve− called as a margin. If margin is maximized, it leads to a conclusion that nearest positive and negative datapoints are quite away from the central lines. Hence wider margin represents the better classification of datapoints. In our study, we have also used the kernel SVM, which is often being used for a non-linear classification.

K-Nearest Neighbour algorithm generates prediction based on the most common ground truth values with its nearest datapoints in the hyperplane. The distance between point in hyperplane is calculated using Euclidean distance. Decision tree algorithm make process look like a tree structure where the leaves of the tree are labelled by categories or classes. Information gain is used to perform the selection of best feature to split the tree further. Random Forest is a Supervised bagging algorithm. Bagging stands for training of learners which have higher variance and predictions of these models are aggregated using voting classification approach to make final prediction. Decision tree with maximum depth is used to make learner with high variance in Random Forest. Number of learners is the profound hyperparameter for random forest algorithm.

Table 2 Performance evaluation of each model

S. No	Meta model	Micro F1-score
1	Logistic regression	0.9151
2	K-Nearest Neighbors	0.9023
3	Decision-Tree	0.9292
4	Linear—Support Vector Machines	0.9311
5	RBF—Support Vector Machines	0.9329
6	Gradient boosting	0.9394
7	Random forest	0.9465

Gradient Boosting anther machine learning algorithm is the combination of two word (Gradient descent + Boosting). Boosting is the conversation of weak models into a strong model. An additive model $\Sigma(t)\_\rho t\_ht(x)$ is fitted by boosting in step or stage wise approach. At each step new weak or highly biased model is trained to compensate gradients of previous weak model. Number of weak learner is the most important hyperparameter for this algorithm.

Dataset contains 15 columns as input and one target column. Seven different machine learning algorithm have been implemented on dataset for prediction of failure instance. In prediction of target column values, which are in form of 0 or 1; 0 shows absence of failure probability, while 1 predicts probable failure. Table below shows micro F1-Score of each algorithm. Micro F1-Score is the performance index. Higher the index, preferable the algorithm is (Table 2).

4 Ensemble Learning

Ensemble learning approach considers multiple models working together/getting merged to perform the task. In ensemble approach, models are trained on the same dataset. Several models are used on the same dataset for training and their combined output is shown in prediction. However, this approach invites overfitting (high variance). To reduce overfitting, models with different algorithms with the same input dataset are used. However, when data is randomized, there is a possibility of each model being trained with the same fraction of input dataset. To overcome this situation, bootstrapping is applied. Bootstrapping uses randomized data with replacement to avoid same data being fed into meta-learners.

In general, ensemble approach considers multiple model for prediction with voting. In multiple classifier models, it evaluates each model and selects a model with the highest votes. In this study, a custom approach is followed. Instead of prediction from a group of models, same group of models is utilized to generate a new 'meta-dataset'. This meta-dataset is then fed in another meta-model through the concept of stacking for predictions.

5 Custom Stacking Based Ensemble Learning

In this proposed custom stacking based ensemble learning approach, we have combined the ensemble learning and stacking approach to create a final model which will be trained using customized splitting of training dataset. For this approach, we have taken inspiration from random forest algorithm and merged it with stacking operations, which helps to optimize the bias-variance trade off.

Now, let's divide the model into fragments and understand the training procedure of this model into different phases. It is comprised of splitting and sampling of a training dataset.

5.1 Phase 1—Training of Meta-learners

1. Split dataset in 80–20 ratio (80% train data—20% test data).
2. Split train dataset into D_1 and D_2 both having equal share.
3. Create d_1, d_2, d_3, …, d_k (*k* samples) from D_1 by sampling with replacement.
4. Train m_1, m_2, m_3, …, m_k (*k* meta learners) using *k* samples as input sets (e.g. (d_1, m_1), (d_2, m_2), …, (d_k, m_k)).

In the process of training, we can see that with the 'D_1' set of data was generated with *k* sample sets through sampling with replacement. It allows training datapoints to be sampled many times for the same predictor. After sampling we passed *k* data samples to *k* meta learners. Here, each meta learning would individually have a higher variance and low bias than the model trained on the original dataset (Fig. 2).

In our study, we have used Decision Trees as a meta learner which is highly preferable for its proposed custom stacking based ensemble approach to leverage the advantages of random forest algorithm. Because decision tree as a meta learner with sampling with replacement (bootstrapping) searches for the best feature among a random selection of features. Random Forest algorithm introduces randomization while generating trees. This results in much more tree diversity, which trades-off a higher bias and a lower variance (Fig. 3).

5.2 Phase 2—Generation of Meta-datasets

1. Pass all datapoints of d_2 in all trained meta-learners for prediction.
2. Collect output of each meta-learner generated by input data from D_2 and make a new meta-dataset. Merge target column in this new meta-dataset from D_2 (Fig. 4).

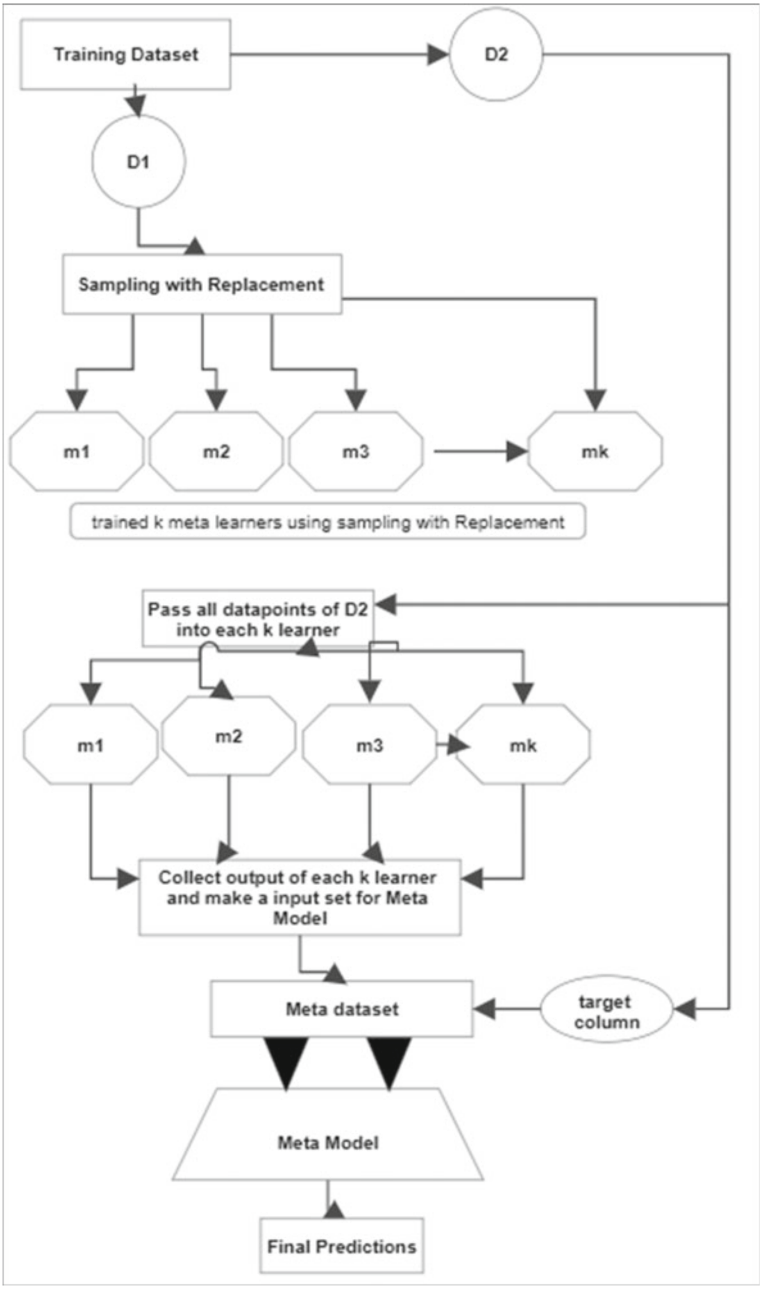


Fig. 2 Complete architecture of custom stacking based ensemble learning model

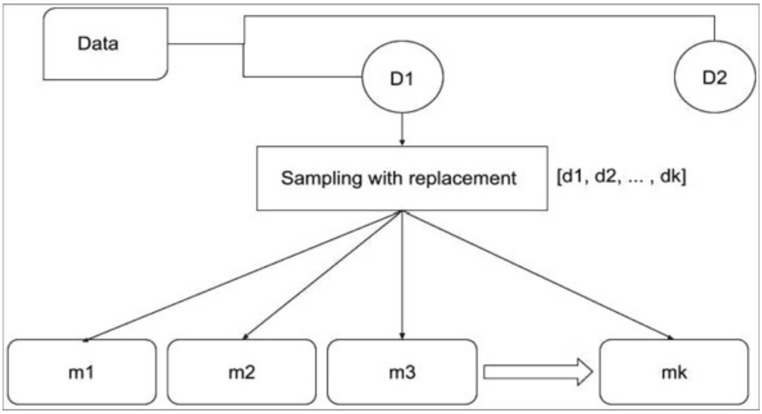


Fig. 3 Train k meta-learners using sample generated with replacement

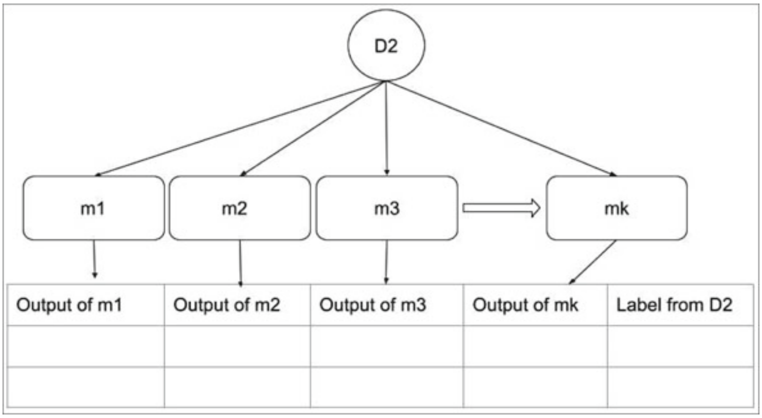


Fig. 4 Generation of meta dataset from meta-learners

5.3 Phase 3—Training of Meta-model

Steps that we have performed in phase-2 and phase-3 will take the "bootstrap aggregation" to next step further by concatenation of outputs of meta learner. As mentioned in phase-1 this approach is inspired by random forest, we merged outputs of meta learner and created a meta dataset. But creation of meta dataset important because it is created by passing data points of D_2 into each meta learner which meta leaner might not have seen during training and when meta dataset would be generated at that time variance would be higher but bootstrap aggregation in form of meta data generation will further reduce the overall variance of this entire complex model while training of meta model. Any classical machine learning model can be used as a meta-model, like Logistic Regression, SVM, kernel SVM, K-Nearest Neighbour,

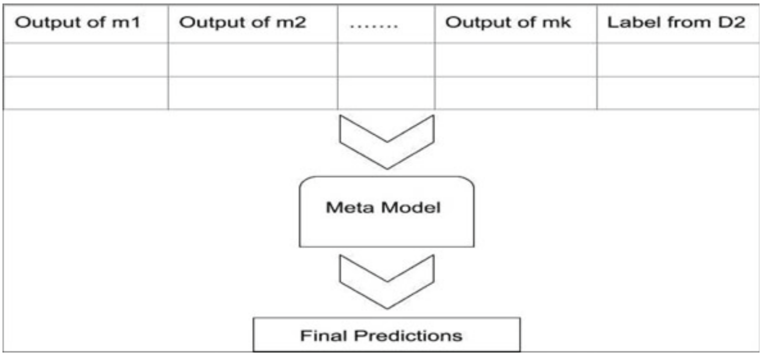


Fig. 5 Training of meta-model

Naive Bayes, Decision Tree and even classical ensemble techniques like Random Forest and Gradient-Boosting. At the phase-1, model had higher variance because of random sampling with replacement and highly depth decision trees, but the bootstrapping aggregation which proposed in phase-2 through merging outputs of meta learners helps in the mitigation of higher variance (Fig. 5).

6 Prediction

1. Pass test data point as an input of each meta learner.
2. Collect the predictions of meta learners and prepare an input data vector for meta model and get the final prediction from meta model for that test data point which was given (Fig. 6).

As explained before, complete stacking based model is carrying multiple number of meta learners with one meta model. It becomes cumbersome when comes to a hyper parameter tuning. But in phase-1 we suggest to use different decision tree as a meta learner with random sampling with replacement, we kept the depth of the meta leaner (decision tree) much deeper. Hence, number of meta learners k becomes a profound hyper parameter which need to be optimized otherwise having a large k value could result into a higher and unnecessary demand of memory and computing power and keeping the value of k lower (~1) could lead the issue of high variance again and which would not be controlled by only one meta learner. Therefore to select best number of k we have taken k from 5 to 1000 in uneven intervals (5, 10, 50, 100, 500, 1000).

Micro F1-Score is used to evaluate the performance of model. It is a harmonic mean of Micro-Precision and Micro-Recall, we performed an addition of individual true positive, false positives and false negative for the different sets to calculate Micro precision and Micro recall. Micro F1-Score does consider label imbalance

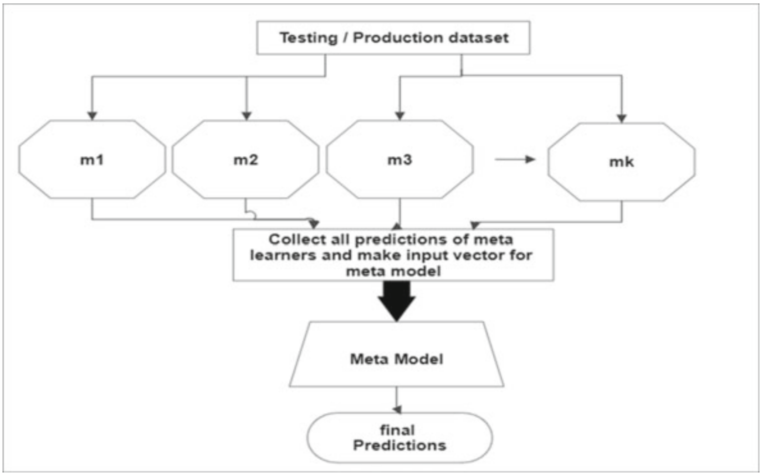


Fig. 6 Prediction through novel trained model

into account. Best value for Micro F1-score is 1 means perfect micro-precision and perfect micro-recall and worst value is 0.

Formula, of Micro F1-Score,

$$\text{Micro f1score} = 2 * \frac{\text{Micro-Precision} * \text{Micro-Recall}}{\text{Micro-Precision} + \text{Micro-Recall}} \tag{1}$$

7 Results

We have tested model on test set of stripper well failure dataset. In the testing we tested model with different values for number of meta learners (k) by using different classical machine learning algorithms as a Meta Model. We have considered Micro F1-Score as an evaluation metric to compare results. After final comparison, with $k = 50$ and Random Forest algorithm as a meta model we have logged best Micro F1-Score on test dataset (Table 3).

8 Conclusion and Future Work

Research has successfully built a model to predict failure of stripper wells from a given dataset. Dataset is first employed to exploratory data analysis followed by model development. Many algorithms have been utilized and compared to find

Table 3 Comparison of results

Meta model	$k = 5$	$k = 10$	$k = 50$	$k = 100$	$k = 500$	$k = 1000$
Logistic regression	0.9651	0.9728	0.9727	0.9726	0.9725	0.9717
KNN	0.9597	0.9593	0.9603	0.9609	0.9608	0.9607
SVM	0.9632	0.9740	0.9722	0.9732	0.9746	0.9742
Kernal SVM (RBF)	0.9645	0.9654	0.9664	0.9662	0.9661	0.9660
Random forest	0.9701	0.9881	0.9887	0.9885	0.9886	0.9885
Gradient boosting	0.9714	0.9721	0.9735	0.9730	0.9732	0.9722

optimum algorithm. Random-Forest algorithm with 500 trees based on highest micro F1-Score.

To enhance performance of model, custom stacking based ensemble approach has been put to test. Prepared as discussed, with 50 meta learners (Decision trees) ($k = 50$) and Random Forest with n_estimators = 100 as a meta model has proved to have higher micro F1-Score of ~0.9887 compared to just Random-Forest algorithm.

Limitation of presented study is no use of deep learning. Having built a highly accurate model with the combination of ensemble and stacking approach, a goal of achieving accuracy higher than proposed can be fulfilled with the use of deep learning neural network in place of Random Forest as a meta learner.

References

1. Covatch GL, Morrison JL (2005) Stripper well consortium offers opportunities and technologies to the stripper well industry. In: SPE eastern regional meeting. OnePetro
2. Petsche T et al (1996) A neural network autoassociator for induction motor failure prediction. In: Advances in neural information processing systems
3. Chen T et al (2019) Prediction of extubation failure for intensive care unit patients using light gradient boosting machine. IEEE Access 7:150960–150968
4. Guo D et al (2015) Data driven approach to failure prediction for electrical submersible pump systems. In: SPE Western Regional Meeting. OnePetro
5. Yi S, Jones R (2020) Machine learning framework for predicting reliability of solder joints. Soldering Surface Mount Technol 32(2):82–92. https://doi.org/10.1108/SSMT-04-2019-0013
6. Murray JF et al (2005) Machine learning methods for predicting failures in hard drives: a multiple-instance application. J Mach Learn Res 6(5)
7. Liu Y et al (2013) Global model for failure prediction for rod pump artificial lift systems. In: SPE western regional and AAPG pacific section meeting 2013 joint technical conference. OnePetro
8. Lige X, Hua SZ, Feng SZ (2019) Road machinery fault prediction based on big data and machine learning. In: 2019 5th international conference on control, automation and robotics (ICCAR). IEEE
9. Sairamya NJ, Susmitha L, George ST, Subathra MSP (2019) Chapter 12—Hybrid approach for classification of electroencephalographic signals using time–frequency images with wavelets and texture features. In: Hemanth DJ, Gupta D, Balas VE (eds) Intelligent data-centric systems, intelligent data analysis for biomedical applications. Academic Press, pp 253–273. ISBN 9780128155530. https://doi.org/10.1016/B978-0-12-815553-0.00013-6

Heart Disease Prediction Using Machine Learning and Data Analytics Approach

Rina S. Patil and Mohit Gangwar

Abstract The heart is the consequent major part connecting to the brain, which has stronger precedence in the Human evidence. It elevates the blood and accumulations to all devices of the entire body. Forecast of circumstances of heart conditions in the medical field is important to work. Data analytics is beneficial for divining more knowledge, and it helps the medical center predict various conditions. A huge number of patient-related data is prepared each month. The collected data can be beneficial for the source of predicting the emergence of future weakness. Unusual data mining and machine learning procedures have been used to indicate heart disease. This research proposed heart disease prediction using various modified Recurrent Neural Network (mRNN) deep learning algorithms. Numerous feature extraction and selection methods have been used to get important features and data collection using custom-generated IoT environments. The system effectively provides heart risk scores with the highest accuracy in a runtime environment.

Keywords Internet of things · Deep learning · Feature selection · Feature extraction · Optimization · Classification · Supervised learning

1 Introduction

Health data mining provides immense promise to discover the secret trends in the scientific domain names data sets. For psychiatric disorders, these structures can be used. The providing raw patient records are therefore widely dispersed, homogenous, and weighty. It is useful to acquire such data in a structured manner. The health monitoring system, data gathered can then be incorporated. A user-oriented methodology to new and secret trends in the information is generated by business intelligence. It

R. S. Patil
Department of Computer Science and Engineering, Bhabha University, Madhya Pradesh, Bhopal, India

M. Gangwar (✉)
Department of IT, IET, Dr. Rammanohar Lohia Avadh University, Ayodhya, Uttar Pradesh, India
e-mail: mohitgangwar@rmlau.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_29

has been estimated from the WHO that 12 million people die internationally each year due to heart failure.

Related to cardio-related disease, half the deaths in the USA and other developing nations occur. It is also the primary cause of casualties in many developed countries. Most of all, it is considered the main reason for deaths worldwide. The word High Cholesterol includes the different diseases that cause the condition. The main cause of fatalities in various countries such as Sri Lanka was breast cancer. In the Western World, heart illness kills one adult every 34 s. Other types of heart disorders include coronary heart disease, cardiomyopathy, and cardiovascular disease. A wide variety of conditions influencing the cardiovascular framework and how air is filtered and distributed through the body are included in the word 'cardiovascular disease.' Several diseases, injuries and deaths arise from cardiovascular disease (CVD). A critical and complicated role in medication is the diagnosis of disorders.

Psychological care is an essential but complex task that remains to be improved precisely and effectively. This program's automation would be highly beneficial. Regrettably, in any sub-specialty, all physicians do not have the experience. Besides, there is a lack of staff at some sites. Therefore, by putting all of them together, and automated medical diagnosis scheme would possibly be extremely helpful. Reasonable judgment on computer-based knowledge help networks will support the achievement of clinical studies at a lower cost. A comparative analysis of the different technologies applied includes effective and precise implementations of the electronic machine. This paper aims to analyze the various statistical and machine learning strategies for evaluating machine learning introduced in past months.

2 Literature Survey

A new deep learning model for sex and age identification using a standard inertial sensor is suggested [1]. This approach is assessed with data obtained from more than 700 individuals using the highest sensor-based training repository value. Ten experiments of Monte-Carlo test sets were given to confirm the robustness and efficacy of the presented method. In protection applications, such as cell phone authorization, biometrics has been commonly embraced. Various biometric features has been tested for each individual object, based on extracted parameters btIoT modules system automatically detect the gender as well as age of user. A supervised neural technique uses a single inertial sensor attached to the lower back of gender and age recognition issues. Machine learning-based techniques have been commonly used throughout vision-based gender detection. Still, they haven't been used for electromagnetic sensor-based gender acknowledgment to our best understanding.

According to [2], a point fulfilment network utilizes a specific dimensional image taken from every suitable angle of view to perform 3d model of objects. The suggested technique has resolved many primary problems in vision-based disease diagnosis relative to previous methods, such as view deformation and scale uncertainty. For testing this method, various experiments were carried out. Still, the feasibility of

SVM and NB, the algorithm was demonstrated in the robust prediction. Tested objects have asked to remember the complete description of food products eaten over the last 24 h and the corresponding portion sizes to measure the calorie consumption. The size of both the food component relies mainly on the entity's personal opinions, which is often too unreliable. Under the circumstances of view occlusion, this approach demonstrates the effectiveness and accuracy of food volume estimation. The finished pixel value of the obstructed food products can be accessed by using the point fulfilment network.

In [3], a comprehensive approach to healthcare monitoring activity recognition using body receptors and a complex Convolutional Neural Network called CNN. They analyze signals from various body sensors in healthcare services, like ECG, wearable sensors, and control system sensors. A CNN based operation is equipped based on the features after removing essential characteristics from the wearable sensors based on PCA. Eventually, the deeply educated CNN used for the identification of data testing events. A publicly accessible standard dataset is applied to the defined solution and then contrasted with other traditional methods. The experimental analysis of this system show it is superior to other traditional classification systems while system works like sensor-based intelligent health systems for behavioral assisting. As a growing field of study, esteemed solutions to implementing a comprehensive smart healthcare delivery system for individuals to extend their independent lives must now be explored. The detectors were positioned on the neck, right wrist, or left ankle of the subject, separately, and connected using elastic bands. Several sensors help us quantify the movement experienced by different body parts, including momentum, turn frequency, and direction of the gravitational flux, thereby better measuring the body's nature.

A new cloud and IoT-based healthcare application was created by Ganesan and Sivakumar [4] to monitor and detect critical disorders. The classifier was trained using data from the benchmark dataset during the training phase. During the testing phase, real patient data was utilized to detect illness and the presence of disease. According to Majumder et al. [5], a numerous sensors system employing a smart IoT that provides an early warning of heart disease risk is suggested. The system gathers data from users through IoT devices and communicates it to a smartphone over Bluetooth using the Body Area Sensor (BAS) system. All the processing and data analysis took place in the application to view real-time user plots of future cardiac arrest. An IoT system with a low power consumption communication model developed regularly collects body temperatures and heart rate using a smartphone. Here ML and signal processing techniques were used to analyze sensor data and predict high accuracy cardiac arrests. A wearable device was implemented based on a smartphone for heart rate detection. It used a combination of ECG and body temperature. A heart rate analysis is done on the Android platform where users can view body temperature and plots of real-time ECG signals.

An adversarial training strategy to multitask learning is developed by Yu et al. [6] to calculate multi-type cardiovascular indices in MRI and CT. These task requirements are shared and taught via multitask forms of collaboration. Finally, they used CT to transfer characteristics learnt from MRI. A set of tests were carried out. The authors used ten-fold cross-validation to improve system efficiency across 2900 myocardial MRI images. The network was then tested on a separate data set of 2360 cardiac CT images. All of the studies on the suggested reverse mapping show that it is quite good at calculating numerous cardiac indices.

In Ali et al. [7] implemented an Optimally Configured and Improved Deep Belief Network (OCI-DBN) to resolve the problems of existing IoT models and boost system performance. System used Ruzzo-approach Tompa's to delete features that do not contribute enough to boost processing speed. They stacked evolutional algorithms stacking two genetic algorithms to have an optimally designed DBN to obtain maximal configuration settings.

In Kumar et al. [8], cloud-based and IoT-based m-healthcare applications have been produced and updated to observe and diagnose the genuine level of severity. To gather the body parametric data using remote devices, like BP, pulse rate, heart rate, pulse pressure, hemogram etc. the approximate estimate can be collected as vital information, connected to the human body, gained from IoT devices. They were using the UCI Repository dataset and clinical sensors to foresee the average person with a major influence on non-communicable diseases medicinal knowledge is produced. The resulting information is securely stored by implementing a new federal storage method in five different steps, such as data storage, data recovery, data gathering, database division, and file merging.

In Park et al. [9] a smart wheelchair method was installed that monitors but visualizes the user's location through a mobile app that attempts to address the inconsistent position of the user. They used strain gauges and shift sensing, using IoT and wireless techniques to communicate with low consumption. It's an extension of Arduino that detects various user stances. This integration complements the user by offering real-time interactive and visualized data for mobile apps to sit properly and understand their present incarnation. With pressure displayed in red, yellow, pink and green triangles, this is a great example of Information systems.

Kumar et al. [10] researchers introduced IoT's three-tier architecture with early detection using a supervised learning algorithm to detect cardiac disease. This system also suggested multi-tier architectures that stored and processed massive quantities of data from wearable IoT environments. First-tier demonstrates on data collection from multiple analog sensors; tier two describes Apache Hadoop that store massive data in cloud storage generated from IoT module, while tier three uses Apache-Mahout to build a reinforcement learning detection and prediction model. In conclusion, ROC demonstrates to get entire system analysis of heart disease prediction.

In Nashif et al. [11] researchers designed a disease prediction system based on the cloud. A real-time monitoring system sensing health parameters like blood pressure, temperature, heartbeat, and humidity was developed using an Arduino microcontroller. The proposed system can detect heart disease using ML techniques as recorded data transmitted to a central server is updated every 10 s.

Suresh et al. [12] implemented an optimized prediction model using a genetic algorithm. System describes various prediction models and important feature selection algorithms. The accuracy and performance of system was better than other traditional prediction models. This method was tested against several prediction models using heart disease datasets and was verified using real-time datasets. To create a balanced training and testing data set, the n–cross-validation process is used.

Researchers employed the rapid miner tool and several ML techniques to enhance the prior accuracy score and forecast heart illness in Alotaibi [13]. The heart disease dataset from UC Irvine was put to the test. The suggested approach enhanced the accuracy score previously obtained.

3 Proposed System

According to many factors, the medical remote management forecast period demand has expanded significantly. When it is very common today in developing countries that older adults typically live separately in their own homes, the elderly population is growing. Besides, the Internet of Things (IoT) makes these remote patient monitoring systems theoretically feasible (IoT as the idea of a capable and modifiable environment of surveillance in which controls or actuators to human and non-living particles) and makes it financially viable because of the even lower expense of sensors. The advancement of smart mobile technology, there are some inbuilt applications are available that provides automatic disease prediction after taking few input parameters, such as temperatures, blood glucose, breathing, mass spectrometry sensor etc. For example, personal health observers such as smart beds instantly notify who is utilizing them. Moreover, they are sure to enlighten about different patients' physical levels, making real smart home drug dispensers automatically warn when medication is not taken, for example. Several virtual care monitoring devices use various technologies inside facilities and their residences to monitor and track clinicians or diagnostic applications. Regrettably, most of these systems are not versatile when new sensors are introduced during runtime, as far as we know. Neither has it allowed regular users with the smart technologies added to generate ad-hoc notifications automatically.

Distributed data processing is one of the processes involved that can resolve a number of the protection, distribution, integration, and management challenges of aggressive data innovation. This propels current research toward thinking about pervasive clinical frameworks focused on the internet. The Predictive Analytic System hardware implementation predicts that disease relative to body temperature, pulse rate, and individual tension. The sensors are mounted on the human chest to monitor the physician's metabolic rate, heart rate, cholesterol levels. It sends data from the information it predicts and diagnoses the disease to the Arduino microcontroller. This information is hosted in the cloud via the end receiver. It can be monitored and anywhere whenever through the Internet. The condition of the physician will be registered on the database and retained. A smartphone framework for machine learning is designed for viewing data from sensors and sending emergency alerts. Healthcare practitioners can track, anticipate, diagnose, and inform their clients at any moment by using this method.

The conceptual system can control several predictor variables on wearable technology, integrated with such a handheld device. Sensor networks work together again to obtain the necessary data, except encountered some problems. It operates by reducing less human involvement to preserve its precision. Using cardiovascular sensors, it is a legitimate diagnostic device for remote areas heart-prone clinicians that monitor heart rate, cholesterol levels, metabolic rate, and many other metrics. N number of measurements obtained can be preserved and viewed subsequently by practitioners to correct acute and chronic treatment.

Figure 1 demonstrates propose system architecture for synthetic and real-time IoT data processing environment. It describes data preprocessing at the first phase, that does null values elimination, class balancing, data acquisition etc. Then various features extraction and selection has been done for next process. In training phase similar features has been considered for module training while for testing from test data. The outcome of system, it provides patient health report according to current parameters. Two major physician and patient interfaces would allow data to be transmitted to each other. In actual environments, this system recognizes fundamental heart problems for the sake of the health of the individual. The system able to provide heart risk score for both genders as well as it having an ability to provide systematic prediction of heart disease based on cardiac index.

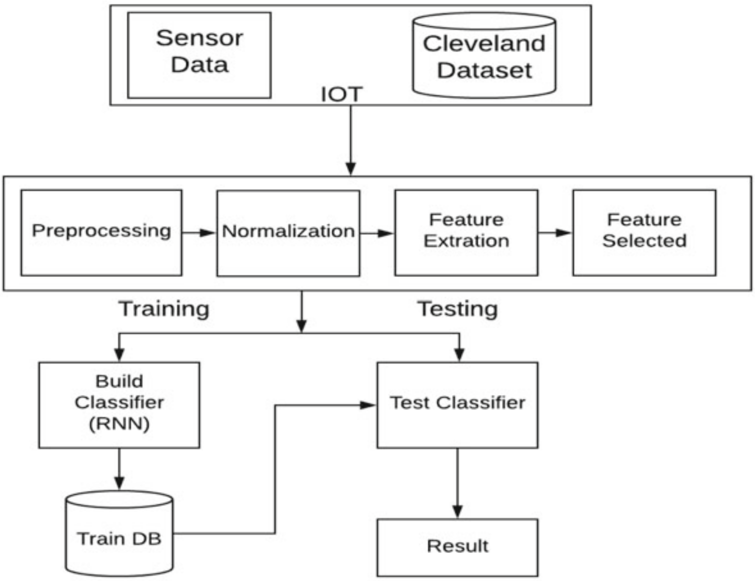


Fig. 1 Proposed system architecture

4 Algorithm Design

Algorithm 1: Proposed modified Recurrent Neural Network Algorithm (mRNN).

Input: Train_Feature set [], // Set of training dataset
Test_Feature set [] //Set of test dataset
Threshold denominator Th
Collection List cL

Output: Generate class label for all test instances based on classification results.
Step 1: Read all attributes from testing dataset using the below function

$$\text{Test\_Feature} = \sum_{j=1}^{n} (T[j])$$

Step 2: Read all attributes from the training dataset using the below function

$$\text{Train\_Feature} = \sum_{k=1}^{m} (T[k])$$

Step 3: Read total attributes from train instances using below function.
Step 4: Calculate the similarity index and generate weight for both feature set

$$\text{Weight} = \text{classifyInstance}(\text{Train\_Feature},\ \text{Test\_Feature})$$

Step 5: Verify with Th.
optimized_Instance_result = Weight >Th ?1: 0;
Add each optimized_instance to cL, when instances = null.
Step 6: Return cL.

5 Results and Discussions

Please It makes sense to interrelate quantitative assessment and clinical scores and could potentially address many challenges in decision-making. The generated training repository was applied with six machine learning models to establish patterns of common, questionable and dangerous activities. To test and rate the machine learning strategies, the threefold, fivefold and tenfold cross-validation model was employed using the behavior classification, training-database. Figure 2 below displays the threefold classification technique used on all parameters and explains all implementations' consistency.

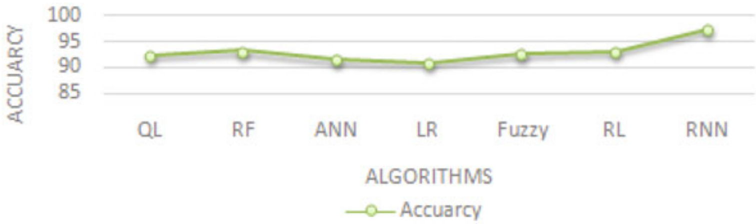


Fig. 2 Accuracy evaluation of various ML and RNN classification

Figure 2 shows the overall accuracy of all approaches, including the suggested RNN. Its accuracy percentage is 97.23%. The minimal accuracy of Linear Regression (LR) is 90.90%, which is greater than other approaches.

The True Positive (TP) of all algorithms, including the proposed RNN, is shown in Fig. 3. It has a TP ratio of roughly 97.40%. The algorithm Linear Regression (LR) has a minimum accuracy of 93.00% when compared to other methods.

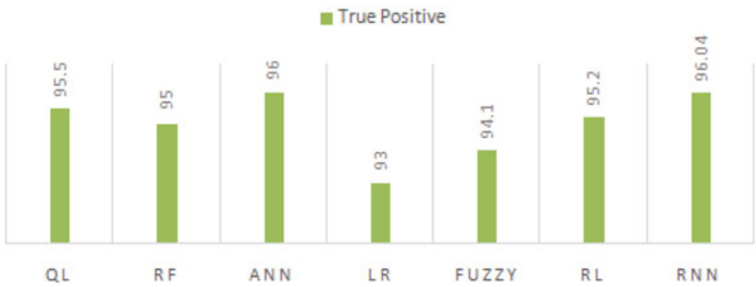


Fig. 3 True Positive evaluation of various ML and RNN classification

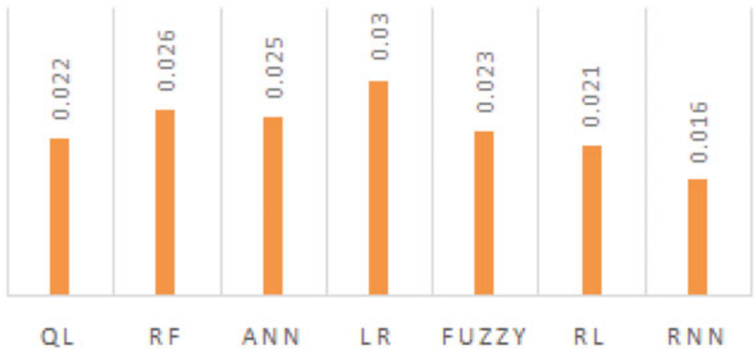


Fig. 4 False Positive evaluation of various ML and RNN classification

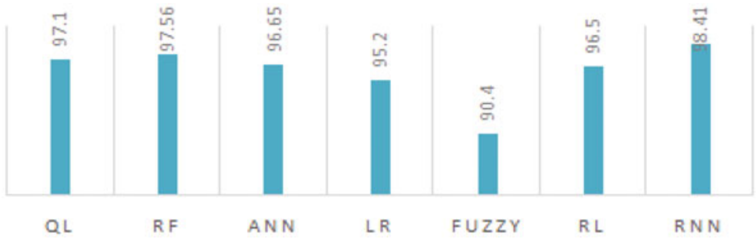


Fig. 5 True Negative evaluation of various ML and RNN classification

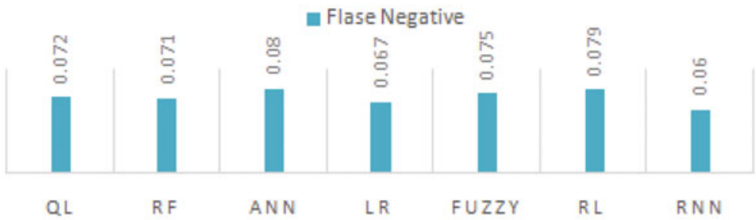


Fig. 6 False Negative evaluation of various ML and RNN classification

The total accuracy of all techniques, including the proposed RNN, is shown in Fig. 4. It has a 97.23% accuracy rate. Linear Regression (LR) has a minimum accuracy of 90.90%, which is higher than other methods.

Figure 5 shows the True Negative of all approaches, including the suggested RNN. It has an accuracy percentage of 98.41%. At 90.40%, the Fuzzy Logic algorithm has the lowest TN of all the approaches.

The false proportion of a system with several algorithms is shown in Fig. 6; LR as well as RNN both provides the lowest negative result ratio consistently.

The above Figs. 2, 3, 4, 5 and 6 improves the importance of different experimental research focusing on various statistical tests with seven distinct algorithms: Q-Learning, RF, Fuzzy logic, Naïve Bayes, Linear Regression, RL and Random Forest with Recommended Perceptron Algorithm. For data management, the mRNN classification algorithm was used during the classifier. The neural network was shown and debated for each model. Both uncertainty metrics show the system accuracy of properly classifying, wrongly classifying, recession, and device recall.

6 Conclusion

This research presented an optimal foundation for real-time prediction models. It may be used by those who have coronary artery disease. It is capable of monitoring and predicting both, unlike many other systems. ML algorithms will be used to forecast cardiovascular disease in the system's diagnostic process. The heart disease dataset served as the basis for the prediction results. The gadget, on the other hand, is incredibly cost-effective; we utilized an enthusiastic pulse sensor and communicated the data to mobile devices using the Arduino suite microcontroller. To monitor the variations and sound an alert if the patient's heart rate rises over the normal range. We conducted experiments using both the tracking and diagnostic methods to demonstrate the system's effectiveness. We tested QL, Linear Regression, Random Forest, Naive Bayes, ANN, and Fuzzy Logic, among other supervised machine classification approaches. The implementation process has validated with and 89% detection accuracy of the defined approach has obtained for mRNN. To extract various feature extraction and selection with embedded deep learning techniques will be an interesting future work for the system. To develop various hybrid feature selection method and evaluate with deep learning classifiers on hybrid dataset will be the interesting task in future direction.

References

1. Sun Y, Lo FP-W, Lo B (2019) A deep learning approach on gender and age recognition using a single inertial sensor. In: 2019 IEEE 16th international conference on wearable and implantable body sensor networks (BSN). IEEE
2. Lo FP-W et al (2019) A novel vision-based approach for the dietary assessment using deep learning view synthesis. In: 2019 IEEE 16th international conference on wearable and implantable body sensor networks (BSN). IEEE
3. Uddin MZ, Hassan MM (2108) Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. IEEE Sens J 19(19):8413–8419
4. Ganesan M, Sivakumar N (2019) IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. In: Proceedings of international conference on systems computation automation and networking 2019. IEEE, 978-17281-1524
5. Majumder AKM, ElSaadany YA, Young R Jr, Ucci DR (2019) An energy efficient wearable smart IoT system to predict cardiac arrest. Hindawi Adv Human-Comput Interact Volume 2019, Article ID 1507465, p 21
6. Yu C et al (2020) Multitask learning for estimating multitype cardiac indices in MRI and CT based on adversarial reverse mapping. IEEE Trans Neural Network Learn Syst. https://doi.org/10.1109/TNNLS.2020.2984955
7. Ali SA et al (2020) An optimally configured and improved deep belief network (OCI-DBN) approach for heart disease prediction based on Ruzzo-Tompa and stacked genetic algorithm. IEEE Access 8:65947–65958. https://doi.org/10.1109/ACCESS.2020.2985646
8. Kumar PM, Lokesh S, Varatharajan R, Gokulnath C, Parthasarathy P (2018) Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Future Gener Comput Syst
9. Park M, Song Y, Lee J, Paek J (2016) Design and implementation of a smart chair system for IoT. IEEE
10. Kumar PM, Gandhi UD (2018) A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases. J Comput Electr Eng 222–235
11. Nashif S, Raihan MR, Islam MR, Imam MH (2018) Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. World J Eng Technol, 22 Nov 2018. https://doi.org/10.4236/wjet.2018.64057

12. Suresh P, Ananda Raj MD (2018) Study and analysis of prediction model for heart disease: an optimization approach using genetic algorithm. Int J Pure Appl Math 119(16):5323–5336, ISSN: 1314-3395 (on-line version)
13. Alotaibi FS (2019) Implementation of machine learning model to predict heart failure disease. (IJACSA) Int J Adv Comput Sci Appl 10(6)

Localized Attention Network for Facial Expression Recognition in an Online Tutoring System

R. Angeline, B. A. Ashwin, S. P. Sri Krishna, and M. E. Gokul

Abstract The world is progressing toward an online mode of education and it requires students to sit in front of a computer for hours, which can be exhausting compared to the traditional method of teaching. Online classes were beneficial at times, but both students and teachers required some time to adapt to this new education model. It is common for students to become inattentive during online classes as they're not at school but at their house and the setting can be a distraction. To overcome this, teachers need to know about the emotions displayed by their students during the class so they can make changes in their teaching style and method to make the online classes more interesting. In this paper, we propose a system that uses facial expression recognition to identify the student's emotions and predicts whether they had a positive or a negative reaction. For this, we use a Localized Attention Network Model for Facial Expression Recognition where the image is divided into target regions, and feature extraction is done; it picks out representative objective information from these divided regions based on the importance of information from the divided image region, it distributes the Attention Network and performs facial expression recognition to assess the students frame of mind in an online learning environment.

Keywords Facial expression recognition · Localized attention network · Feature extraction

R. Angeline · B. A. Ashwin (✉) · S. P. Sri Krishna · M. E. Gokul
SRM Institute of Science and Technology, Ramapuram, India
e-mail: ab2539@srmist.edu.in

R. Angeline
e-mail: angelinr1@srmist.edu.in

S. P. Sri Krishna
e-mail: ss3489@srmist.edu.in

M. E. Gokul
e-mail: ge9259@srmist.edu.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_30

1 Introduction

E-Learning and online classes are slowly replacing the conventional teaching method. However, this platform still faces challenges like evaluating teaching and student's response to it, which plays a crucial role in assessing a teacher [7]. Online classes still face many challenges compared to traditional methods of education. Here we consider the student's positive and negative expressions to evaluate teachers and their role in keeping the class interactive and interesting. So, if a faculty is teaching in an interesting way, then students will be responsive and exhibit positive emotions. In contrast, when a class tends to be boring, the student will display negative emotions [8]. The model can classify emotions expressed by students in online classes into two types: Positive Expressions and Negative Expressions, respectively. The positive expression consists of students being happy, excited, surprised, energetic etc. Negative Emotions consists of students being sad, angry, disappointed, dissatisfied, drowsy [8]. We will implement a model which separates these emotions and classifies student emotions giving an evaluation of faculties Teaching. Deep learning plays an integral role in incorporating technology in daily life & in recent times, Facial expression and Facial recognition using Conventional Neural Network have found its importance in many fields [6]. Due to increased demand in Facial Expression, there's a need to increase the accuracy and multiple human emotion recognition.

2 Existing Concepts/Related Works

2.1 Facial Expression Recognition

Facial expressions are one of the essential aspects of human communication. This field has now become an essential area in data science. It has garnered a lot of attention in recent times due to its potential in various applications. Facial expression recognition can play a significant role in the online tutoring platform to detect how efficiently the teaching is done by detecting the expressions of the student and classifying them accordingly [7]. We face many problems in expression recognition due to the lack of data. These models are sensitive to pose variations. The posture and attributes of an individual's face vary when the position of the head changes. Occlusion is basically blockage; it usually occurs when the face is covered or not fully visible and when all face attributes are not available as an input image. It occurs when a person has a beard, moustache, spectacles or even when wearing a mask [2]. Illuminations are light changes in the environment, and the slightest change in lighting conditions can cause a considerable challenge in recognizing and detecting facial expression recognition processes. An impervious facial expression recognition model should be capable of handling variations in illuminations, expression, pose and occlusion. Here are some of the most sought out models for facial expression recognition (Table 1).

Table 1 Literature review

Author and year	Reference no	Technique	Dataset	Description
XIAO-YUTANG, WANGYUEPENG	[7]	PAD Emotional state model	JAFFE	In this paper, they've proposed a smart teaching evaluation system using classic CNN Alex Net
Xuxu, Z. Ruan	[1]	Graph neural network	CK+	In this paper, they've implemented a graph convolutional neural network for feature extraction and facial recognition
Q. Xuand, N. zhao	[10]	LBP(Local binary pattern)	CK+	In this paper they've proposed an algorithm based on CNN and LBP to perform facial expression recognition
Y. Qaio, J. Yang, X. Peng	[2]	RAN(region attention neural network)	FER-plus	In this paper, they've implemented a novel Region Attention Network(RAN) model for facial detection to overcome various poses and occlusions
J. Lee, S. Kim and K. Sohn	[3]	MRAN(multi-modal recurrent attention networks)	RECOLA	In this paper, they've implemented multimodal recurrent attention networks (MRAN) which takes shades, depth, and the rmal recording clips as input and performs facial recognition
S. D. Deb, et al	[4]	Parallel CNN model	Alexnet	In this paper, they've implemented a parallel Convolutional Neural Network to perform expression recognition
J. Zhou, X. Zhang, Y. Liu	[5]	Spatial–temporal semantic Graph network	CK+ ,	In this paper, they've implemented a Spatial-Temporal Semantic Graph Network(STSGN)which detects spatial and temporal patterns to detect expression

(continued)

Table 1 (continued)

Author and year	Reference no	Technique	Dataset	Description
Saurabh Pal, et al	[8]	CNN	FER-2013	In this paper, they've proposed a method to analyze student emotions by combining multiple expressions

3 Proposed Concept

We propose a convolutional neural network-based localized attention network for facial expression and facial recognition of students during online classes and provide a complete analysis of the learning environment of the class. In previous models, the facial expression recognition was based on a standard convolutional neural network where the weightage was distributed equally to all parts of the image, but this method is time-consuming and fails to recognize important regions of the facial image. So, both the important part and the unnecessary part of the image was given equal weightage, which leads to a decrease in facial expression accuracy. Here in this model, we consider the above problems from past works and propose a concept of localized attention network where the image is divided into target regions and we give more weights to important objective information of the image and distribute attention network based on importance and perform facial expression recognition.

The proposed system shown in Fig. 1 detects the basic emotions of students and sends the real-time data, which is then plotted as a graph and the data is combined by the PAD method to form complex emotions to determine the students' response to the class [8].

3.1 Data Acquisition

In this process, data is gathered for emotion recognition. Images are gathered from the camera and the faces of the students are extracted.

3.2 Feature Extraction

In this process, important features of the face will be extracted and other information will be discarded.

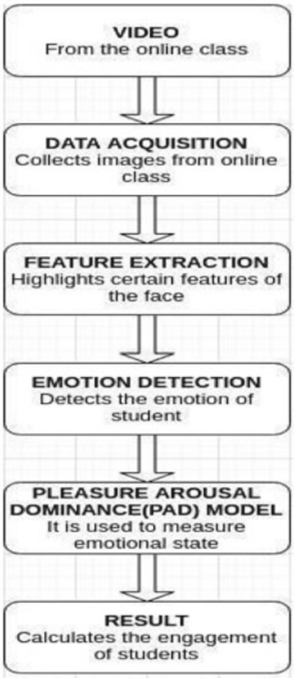


Fig. 1 Online class evaluation system

3.3 Emotion Detection

The localized attention network model will detect the emotions of the students.

3.4 Pleasure-Arousal-Dominance (PAD) Model

After the emotions are detected, they will be clustered based on the PAD model and there silt will be produced.

3.5 Pleasure-Arousal-Dominance (PAD) Emotional State Model

The pleasure-arousal-dominance emotional state model was found by Albert Mehrabian and James A to describe and measure emotional states. It consists of three dimensions, namely dominance, arousal and pleasure to classify and represent all emotions. Pleasure detects whether a person perceives the surroundings as enjoyable or not. Arousal detects the extent to which the environment influences the individual.

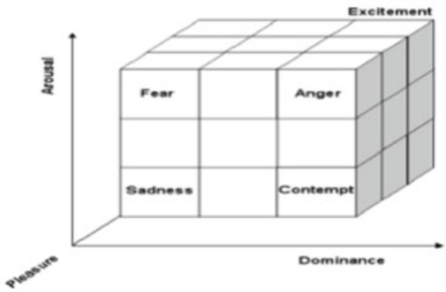


Fig. 2 Emotions in PAD model

Table 2 Weights of dimension of each emotion according to PAD model

Emotion	Pleasure	Arousal	Dominance
Neutral	0	0	0
Happy	2.77	1.21	1.42
Angry	−1.98	1.1	0.60
Afraid	−0.93	1.30	−0.64
Disgust	−1.80	0.40	0.67
Sad	−0.89	0.17	0.70
Surprise	1.72	1.71	0.22

Dominance detects whether the person feels in control or not in the environment. By plotting these three dimensions, we can construct an emotion coordinate x is and recognize emotions based on these three dimensions (Fig. 2) (Table 2).

$$Ip = \sum_{k}^{i} w_p p_i \tag{1}$$

w_p = weight of the corresponding pleasure scale.

$$Ia = \sum_{k}^{i} w_a p_i \tag{2}$$

w_a = weight of the corresponding arousal scale.

$$I_{d=} \sum_{k}^{i} w_d p_i \tag{3}$$

w_d = weight of the corresponding dominance scale.

$$I = \frac{I_{p+} I_{a+} I_d}{3} \tag{4}$$

where,

P_i = Probability of the emotion.

I = Engagement score.

4 Module Explanation

In this paper, we use a localized attention neural network shown in Fig. 3 for enhanced facial expression recognition to assess the student's awareness during an e-learning environment. Most often, in conventional neural networks more importance is given to the number of neural networks and data trained but, in our model, we will be using an optimal conventional neural network with the required dataset to get enhanced accuracy. Not all parts of an image are important for facial expression recognition, so in this model we distribute our attention network to important features in the

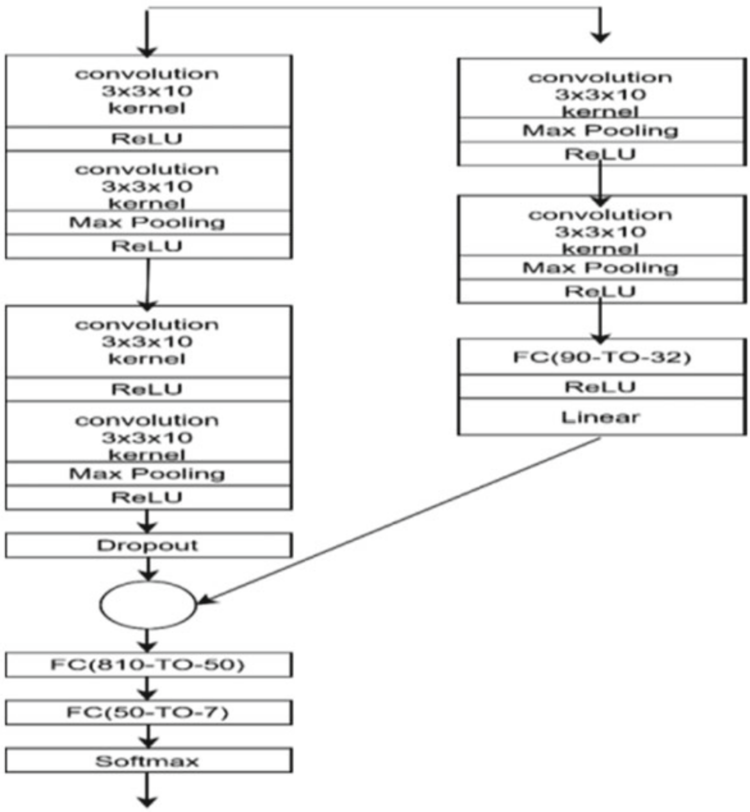


Fig. 3 Model architecture for emotion detection

face required for the recognition process, neglecting the unnecessary parts. While taking alive facial image input in a conventional neural network, an RGB image file is converted into a grayscale image with a grey level lying between 0 and 255 where '0' represents black and '255' represents white and anything lying between 0 and 255 represents a grey level. This image in CNN is received as a matrix with a set of grey level number scales, and then this matrix is multiplied with a matrix of filters such as edge filter and corner filter to get a piece of enhanced feature information from an image which we term as feature extraction [7]. Along with this, in our model concurrently, the same live facial image is divided multiply and converted into a grayscale image and the part of the image which has greater importance with respect to the greyscale grading and facial features goes through a localized attention neural network.

4.1 Convolutional Kernel

In the above flow chart, the convolutional layer represents the typical processs of convolutional neural network which involves conversion of a RGB image into a gray scale image upon which feature extraction takes place.

i. **RELU**
ReLU is an activation function which is required to convert our image input data into a linear form where all the negative values from the grid are removed, taking only the positive points to obtain linearity.

ii. **Max Pooling**
Max pooling is an operation which considers the maximum value in a patch of sub-divided matrix values. This helps in considering the best feature in an image omitting the other features. This operation also helps in reducing over fitting as we are avoiding spoon feeding to our model.

iii. **Dropout layer**
This process involves randomly dropping a neural layer or a grid in a matrix in order to avoid over fitting which indirectly helps in training our model better with an improved testing accuracy.

iv. **Soft Max**
SoftMax is a probability activation function which helps in converting a logical number from a matrix into a number representing probability which involves a formula,

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

v. **FC**
FC represents a fully connected neuron layer that sums up all the features from the before convolutional layers. The model consists of four convolutional

kernels along with two max-pooling layers and an activation function rectified linear unit (ReLU). Later we introduce a dropout layer followed by fully connected layers terminating with SoftMax activation function from which we obtain a probability output. In the localized attention network, we have dual convolutional layers followed by dual max-pooling layers and two ReLU activation functions. The obtained data from these are then passed through a fully connected layer ReLU activation function. After this, we obtain linear grid data which we name as $T(\theta)$, giving us the semi-processed data. From this layer, we obtain data only from the important parts of the image over the unnecessary details. The obtained processed image data from these two networks are merged, forming a fully connected neural layer where we obtained our required information after passing it through a softmax activation function. We will be using the facial expression recognition (FER 2013) dataset to test and train and test our model.

5 Experimental Results

We tested our proposed model in the Fer-2013 dataset and obtained a reasonably good performance compared to the models we referred. In every case, we first trained our model on the Fer-2013 datasets and then validated using the validation set and then tested to obtain accuracy.

Our 4 h of training procedure involved setting up our model's architecture and other parameters than training it for 500 epochs. Using Adam optimizer, we gave attention to network weights with random Gaussian variables and a standard deviation of 0.05. We also used data augmentation on our training images to make our image data invariant. Since Fer-2013 is a huge dataset, it took around 4 h to train. We obtained a result of 74.02% and compared it with other works in Table 3.

Figures 4, 5, 6 and 7 are the results we obtained from our model. Along with facial expressions like happy, neutral, angry, surprise, our model is able to predict sad, disgust, fear, along with considering the fact that FER is a huge dataset. It has lots of variant images and imbalances within classes of emotion. These images are then fed to a PAD model that assesses the performance of the student with respect to the expressions that they display (Figs. 8 and 9).

Figure 4 shows the number of images with a specific expression in the Fer-2013 dataset and Fig. 5 shows the confusion matrix used to evaluate the performance of our model (Fig. 10).

Table 3 Comparisons of results

Work	Approach	Dataset	Avg Accuracy (%)
This paper	Localized CNN	FER2013	74.02
Saurabh Pal et al. [8]	CNN	FER2013	71.83
Linyi Zhou, Yingjie Ma, Tardi Jahjadi [9]	UA-ETN	FER2013	73.5

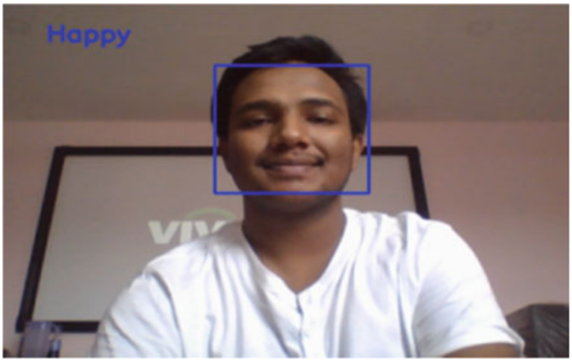


Fig. 4 Happy Emotion

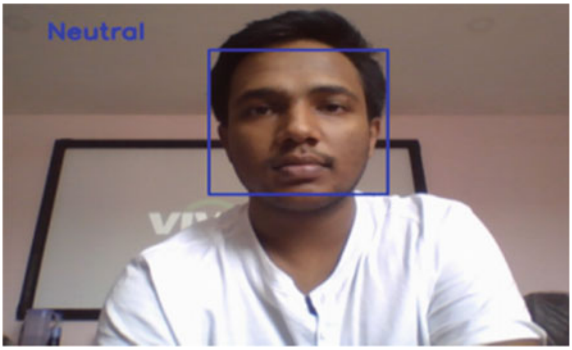


Fig. 5 Neutral Emotion

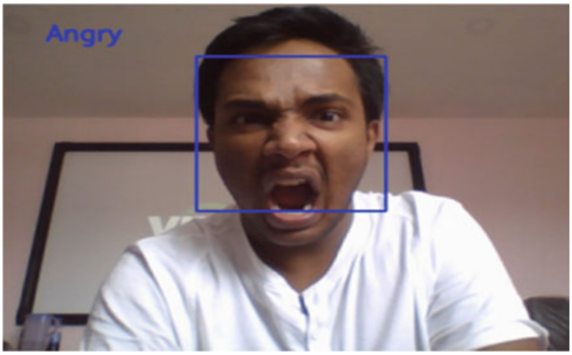


Fig. 6 Angry Emotion

p = *0*	*p* = *0*	*p* = *2.77*	*p* = *0*
a = *0*	**a = 0**	**a = 1.21**	**a = 0**
d = *0* *I* = *0.45*	**d = 0**	**d = 1.42**	**d = 0**

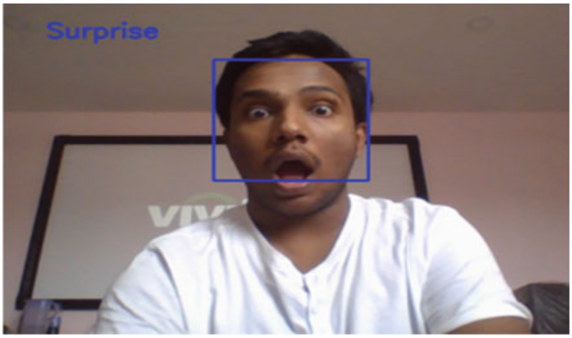


Fig. 7 Surprise Emotion

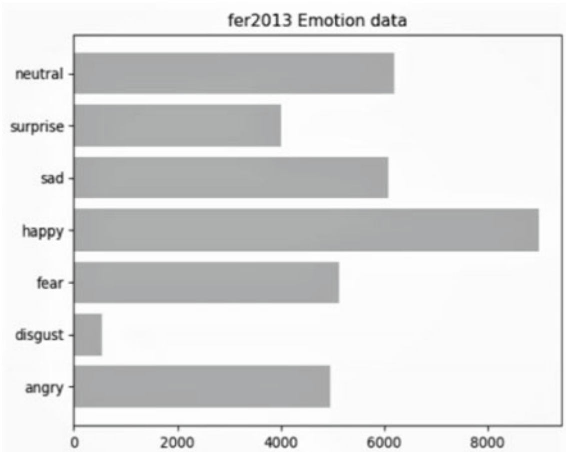


Fig. 8 FER 2013 dataset emotion

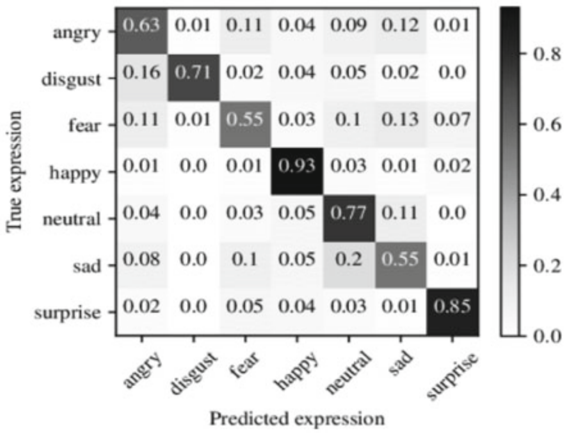


Fig. 9 Confusion matrix for FER2013

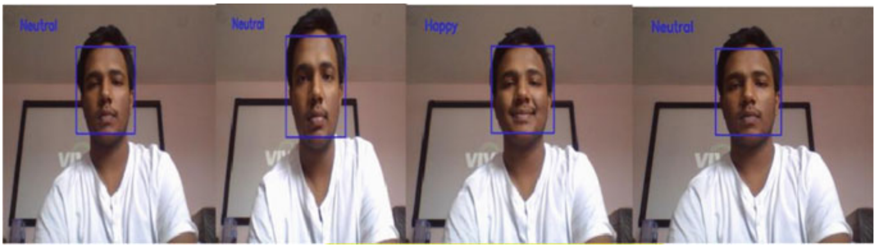


Fig. 10 Expressions detected

Here we take the corresponding weight so *f* *P*, *A* and *D* for the corresponding emotion and calculate the performance score using the formula (4) specified above.

Positive emotions increase the value of *I*, while negative emotions decrease the value of *I*.

The greater the value of *I*, the better the student's response.

Here, the value of *I* is 0.45 from which we learn that students exhibited more negative/dull emotions during the class and that the class was boring and unsatisfactory.

6 Conclusion

In this paper, we propose a method to evaluate the student's response to an online class which helps the teachers to improve their way of teaching suitable for online classes. We have used a localized attention network model to predict the emotions of students which has given 74.02% accuracy, which is a fairly good accuracy compared to the previous models that used this dataset and fed these images to a PAD model to assess the framework of a student's mind. From these results, we fairly judge a student's emotion but human emotions cannot be predicted solely from images because of external factors. Next, we are trying to improve the algorithm and to overcome the issue of occlusions like spectacles and masks and the lighting effect on the face. Overcoming these challenges will makes this model even more refined and suitable for real-world deployment.

References

1. Yang L, Ruan Z, Xu X (2020) Facial expression recognition based on graph neural network. In: 2020 IEEE 5th international conference on image, vision and computing (ICIVC), Beijing, China, pp 211–214. https://doi.org/10.1109/ICIVC50857.2020.9177430
2. Qiao Y, Wang K, Meng D, Peng X, Yan J (2020) Region attention networks for pose and occlusion robust facial expression recognition. IEEE Trans Image Process 29:4057–4069. https://doi.org/10.1109/TIP.2019.2956143

3. Sohn K, Lee J, Kim S (2020) Multi-modal recurrent attention networks for facial expression recognition. IEEE Trans Image Process 29:6977–6991. https://doi.org/10.1109/TIP.2020.2996086
4. Laskar RH, Deb SD, Sharma M, Choudhury C, Laskar RH, Talukdar FA (2020) Frontal Facial expression recognition using parallel CNN model. In: 2020 National conference on communications (NCC), Kharagpur, India, pp 1–5. https://doi.org/10.1109/NCC48643.2020.9056011
5. Liu Y, Zhou J, Lan X, Zhang X (2020) Facial expression recognition using spatial-temporal semantic graph network. In: 2020 IEEE international conference on image processing (ICIP), Abu Dhabi, United Arab Emirates, pp 1961–1965. https://doi.org/10.1109/ICIP40778.2020.9191181
6. Nasoz V, Singh S (2020) Facial expression recognition with convolutional neural networks. In: 2020 10th Annual computing and communication workshop and conference (CCWC), Las Vegas, NV, USA, pp 0324–0328. https://doi.org/10.1109/CCWC47524.2020.9031283
7. Peng WY, Tang XY (2020) Classroom teaching evaluation based on facial expression recognition. In: ICEIT 2020: Proceedings of the 2020 9th international conference on educational and information technology, Feb 2020, pp 62–67. https://doi.org/10.1145/3383923.3383949
8. Nayyar A, Pal S, Mukhopadhyay M (2020) Facial emotion detection to assess learner's state of mind in an online learning system. In: ICIIT 2020: proceedings of the 2020 5th international conference on intelligent information technology, Feb 2020, pp 107–115. https://doi.org/10.1145/3385209.3385231
9. Fan X, Zhou L, Ma Y (2020) Uncertainty-aware cross-dataset facial expression recognition via regularized conditional alignment. In: MM '20: proceedings of the 28th ACM international conference on multimedia Oct 2020, pp 2964–2972. https://doi.org/10.1145/3394171.3413515
10. Zhao N, Zu Q (2020) A facial expression recognition algorithm based on CNN and LBP feature. In: 2020 IEEE 4th information technology, networking, electronic and automation control conference (ITNEC), Chongqing, China, pp 2304–2308. https://doi.org/10.1109/ITNEC48623.2020.9084763

Image and Signal Processing

Fruits Classification and Grading Using VGG-16 Approach

Nishtha Parashar, Aman Mishra, and Yatin Mishra

Abstract Due to availability of human resources in India, nearly all the horticulture work is done manually by workers which involves the possibility of human error, and we get to have some rotten or not so fresh fruits. To avoid this possibility of human error and mistakes it is required to have an automated system in the field of horticulture in India. Numerous research papers have been proposed to address these external issues, however they all have drawbacks and limitations, such as low accuracy due to training on a smaller number of datasets and use of weak algorithms in some cases. In this paper, we present a CNN based approach that separates fresh and rotten fruits of different kinds. To reinforce our experimental results, we trained on a dataset of 14,000 images, resulting in an accuracy of 93.52%.

Keywords Fresh fruit · Rotten fruit · Classification · Texture features · Convolution neural network

1 Introduction

According to a report issued by Export Import Bank of India, "At production level, the major challenge is low productivity, while at post-production stage, the wastage rate is very high". There is a huge difference between the prevailing and probable yields of production. This can be reduced by improved technologies. The study further stated that projected wastage in India varies from 11% in mangoes to 90% in tomatoes. The problem that was estimated is more in post-harvest technologies. The traditional method of grading involves grading at producer's level in regulated markets of different states and it is done manually due to lack of effective extension

N. Parashar (✉) · A. Mishra · Y. Mishra
GLA University, Mathura, UP, India
e-mail: nishtha.parashar@gla.ac.in

A. Mishra
e-mail: aman.mishra_cs17@gla.ac.in

Y. Mishra
e-mail: yatin.mishra_cs17@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_31

services for the dissemination of improved post-harvest technologies of fruits [1, 2]. The fruits are graded by human beings by their weight, color, shape and size where it is not possible at a mass scale and some of the fruits are left rotten and they affect other fresh fruits. To overcome this problem, we need an automated system so that human efforts can be reduced, and the production can be increased [3].

Thus, we intend to develop a novel image processing technique that can more accurately distinguish between fresh and rotting fruit than the currently available methods. The characteristics that distinguish distinct fruits differ from one another, and are mostly determined by size, texture, color, and interior compositions [4–6]. Our goal is to design an algorithm that can differentiate between fresh and rotten fruits based on external variables. For the most part, these requirements and parameters are the same for each variety of fruit. Every variety has its own uniqueness and flavor, resulting in a wide range of pricing and preferences in society [7].

Fruit producers and processing businesses keep a wide range of fruits, resulting in a mishmash of agricultural food in storage and depots [8, 9]. The majority of post-yield processing companies employ physical techniques for arranging and categorization, which is a time-consuming and costly activity, and yet there is inconsistency [10–13]. The primary goal of this study is to reduce manufacturing losses by creating and testing the efficacy of a Convolution Neural Network (VGG-16) algorithm, in order to meet the goal of autonomous fruit recognition and categorization into fresh and rotten categories.

2 Related Work

Digital image processing has been employed by researchers and specialists to solve a range of production difficulties, specifically in the agriculture sector industry [14]. Ronld et al. [11] developed an approach that differentiate distinct varieties of fruits using image pre-processing and the naive bayes technique, with experimentation using 150RGB images gathered by taking images from their handset from the market in firmness of 2048 $\times$ 1024 and cropping it into their required amount, and then converting RGB image to grayscale image and performing filtration and classifying it. They employed random samples to select images for training and validating the dataset, and the remaining images for testing. They applied statistical analysis to analyse the performance and achieved accuracy around 92%.

Expansion of Controlling Systems for Fruits Identification grounded on CNN [15] issued in 2018 used CNN over 971 distinct images in dataset with testing accuracy of 94%, and issued in 2016 developed a machine-based apple variety identification using naive bayes approach with 88% precision.

Hena et al. [9] conducted experimentation on mango variety detection utilising the raw leaf, employing image processing, CNN, and statistical analysis to achieve a 78.01% accuracy. In the first epoch, their learning rate was 0.011%, with 0.1% training accuracy. Following that, they finalise the 78.01% by taking the average of each period accuracy.

Bhargva and Bansal [7] developed a technique in which they distinguish fruit variety and quality, then categorise the fruit into three categories: high, average, and rejected trait. They use k-nearest Neighbor, Support Vector Machine classifiers for classification. They also employed CNN to detect variety. Each method has a varied level of accuracy, but SVM outperforms the others with a score of 95.72%. In [16] antioxidant and antimicrobial properties of dried portuguese apple range were tested.

In Development of Control System for Fruit Identification based on Convolutional Neural Network, Khaing et al. [17] created a method that uses CNN to classify fruits. They used Alexnet for classification, and Matlab's graphic processing unit was used to develop the simulation procedure. Using this classification method, they were able to achieve a 94% accuracy rate.

Li et al. [6] developed a methodology for classifying apple varieties using near-infrared spectroscopy. They employed NIR spectra gathered in diffuse reflectance mode in this model, then used PCA as a data mining approach with spectra statistics, and then used the SPA approach and a back propagation neural network to classify different apple varieties. They tested it over several samples to get accurate results, and they were able to achieve an average calibration accuracy of 99.44% and a forecast accuracy of 96.67%.

3 VGG-16 Based Architecture

In our VGG-16 technique, we use three distinct fruit varieties: apple, banana, and orange. The VGG-16 (Visual Geometric Group) is a more developed CNN that has five convolutional units. Each layer will include comprehensive filters that will be used to gather image features, and the 16 alludes to the 16 unique levels of neural networking. As input, the pre-processed data is feeded into this finely tuned architecture. All the layers are trained here in this CNN architecture irrespective of the basic model. Figure 1 depicts a comprehensive layered description of this model. Each block has a set of three convolutional layers, and a max pooling layer connects adjacent blocks. The number of convolution kernels in every block remains constant, rising from 64 in the first to 512 in the last.

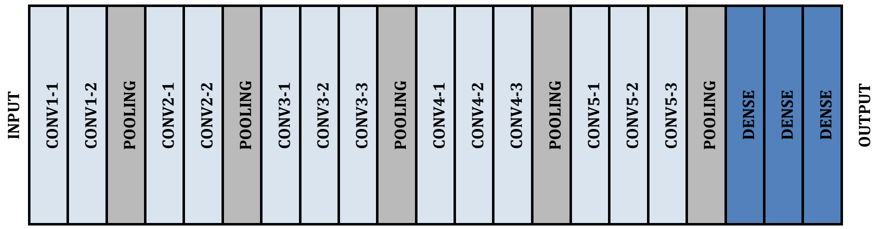


Fig. 1 VGG-16 finely-tuned architecture

A total of 13 convolution layers, 5 max pooling layers, and 3 fully linked layers are employed in this model.

3.1 Convolution Layer

Whenever an image passes through the convolution layer, image filtering as per the input shape is performed and generates as many filters as the filter param specifies, multiplying each of them over the image, and then determining which filter gives the improved result at each epoch.

3.2 Max-Pooling Layer

It is not computed as a layer, but it plays a critical role in this architecture, in that it takes the image matrix and calculates the greatest intensity value for the given pool size, effectively reducing the dimension without losing the image's attributes.

3.3 Fully Connected Layer

Because every output in CNN is related to each input of previous layer, the completely connected layer in CNN is the one that accepts input from the convolution and max-pooling layers' outputs and connects them all. This layer is used last after all other layers because it is the process of real classification.

3.4 Fitting Model

The number of times the data is fed into the neural network is referred to as the epoch. Over-fitting may occur if greater epoch is used, whereas under-fitting may occur if less epoch is used, thus we use epoch of size 5 for this model.

3.5 Layers Implementations

Layer 1	conv-layer; 244*244; 64
Layer 2	conv-layer; 64; 64

(continued)

(continued)

Layer 3	conv-layer;112*112; 128
Layer 4	conv-layer; 112*112; 128 dimension reduced using similar poolsize of 2*2; strides: 2,2
Layer 5	conv-layer; 56*56, 256
Layer 6	conv-layer; 56*56; 256
Layer 7	conv-layer; 56*56; 256 dimension reduced using similar poolsize of 2*2; strides: 2,2
Layer 8	conv-layer; 28*28; 512
Layer 9	conv-layer; 28*28; 512
Layer 10	conv-layer; 28*28; 512 Max-pooling is used with same parameters as previously used
Layer 11	conv-layer; 14*14; 1024
Layer 12	conv-layer; 14*14; 1024
Layer 13	conv-layer; 14*14; 1024 Max-pooling is used with same parameters as previously used

The last three layers are completely connected layers that will be used to do the concluding classification using the supplied weight.

4 Experimental Results and Discussions

4.1 Experiments on Dataset

The dataset, which includes both fresh and decaying images of an apple, banana, and orange, was developed by Sriram Reddy Kalluri in 2018 on Kaggle.com. For training and testing purposes, the dataset was split into an 80:20 ratio, with 10,901 images for training and 2698 images for testing.

Fresh category images of apples, bananas, and oranges from the dataset on which testing was done are shown in Figs. 2, 3 and 4. Figures 5, 6, and 7 show samples of rotten apples, bananas, and oranges from the dataset on which testing was done. Table 1 lists the details of the samples used for training and testing.

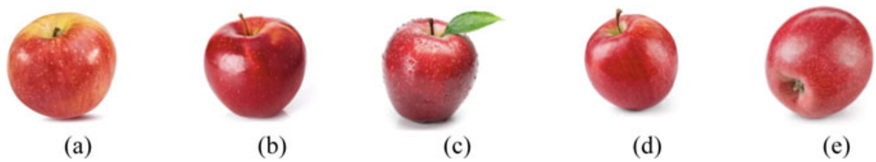


Fig. 2 Fresh apple images from dataset

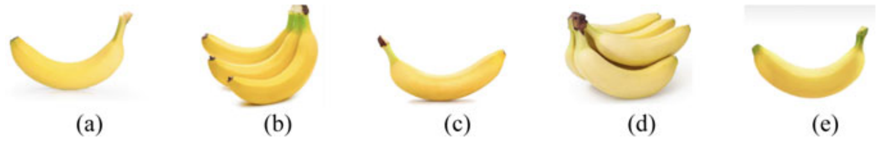


Fig. 3 Fresh banana images from dataset

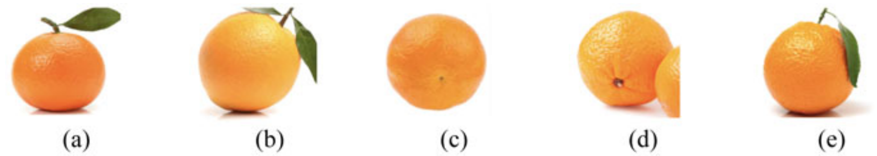


Fig. 4 Fresh orange sample from dataset

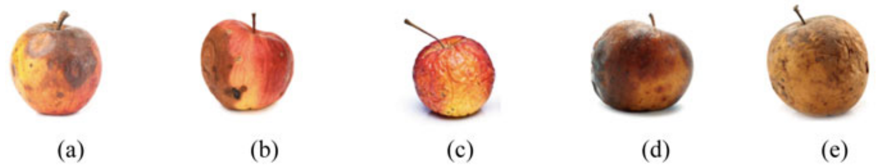


Fig. 5 Rotten apple images from dataset

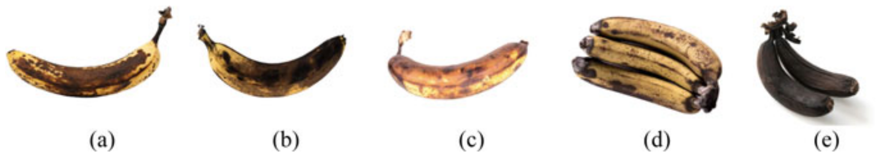


Fig. 6 Rotten banana images from dataset

Figures 8 and 9 illustrate the training vs validation accuracy and loss graphs for the supplied VGG-16 fine-tuned model, respectively. Also, the prediction results for grading of apple, banana, and orange fruits into fresh and rotten are presented in Fig. 10a–f.

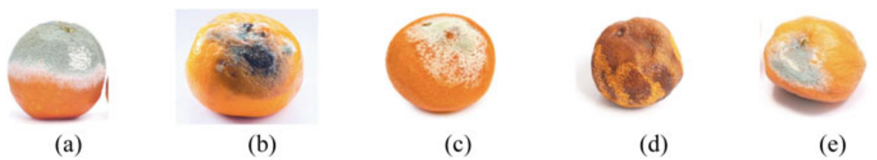


Fig. 7 Rotten orange images from dataset

Table 1 Details of dataset of images

Name of variety	Num of sample (training)	Num. of sample (testing)
Apple fresh	1693	395
Apple rotten	2342	601
Banana fresh	1581	381
Banana rotten	2224	530
Orange fresh	1466	388
Orange rotten	1595	403

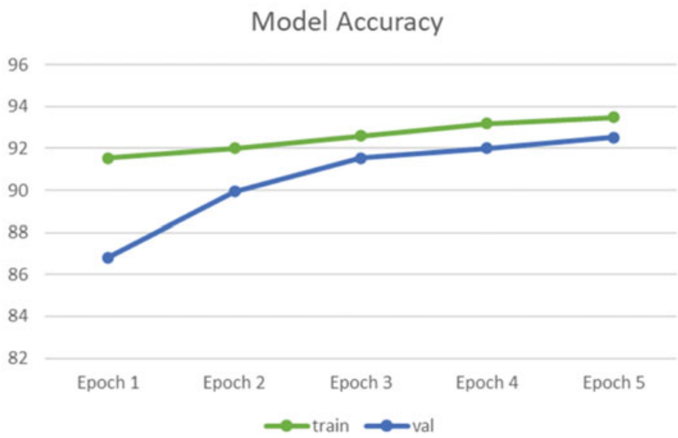


Fig. 8 Training versus validation accuracy for VGG-16 finely tuned architecture

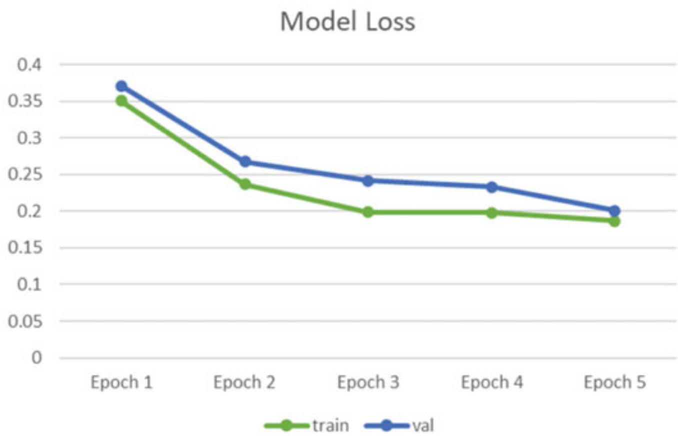


Fig. 9 Training versus validation loss for VGG-16 finely tuned architecture

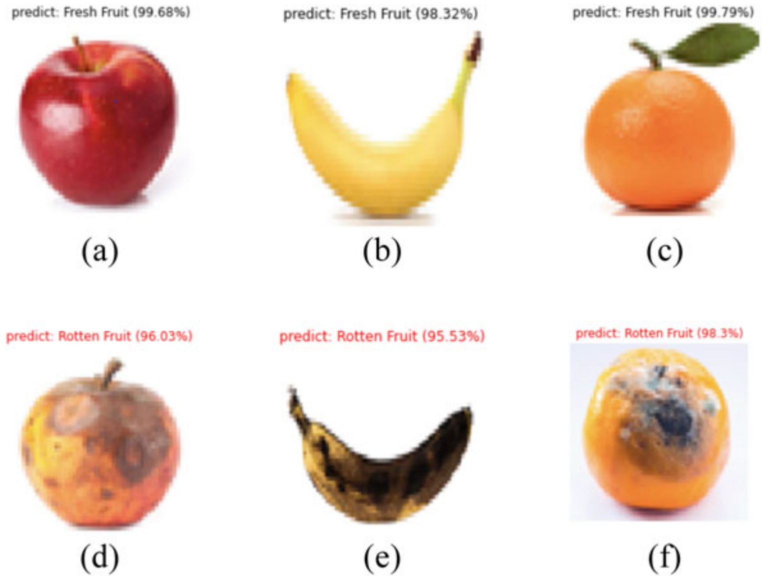


Fig. 10 Results for grading of fruits into fresh and rotten by fine-tuned VGG-16 model

Table 2 Performance assessment with other approaches

Method	Accuracy (%)
Technique in [3]	85
Technique in [4]	86.52
Technique in [13]	92.93
VGG-16 finely tuned method	93.52

4.2 Comparison with Other Methods

Some of the past studies for fruit variety identification that were established in previous years have attracted attention, as detailed in Table 2. When compared to other current approaches, the suggested system performs better and is more accurate with the given dataset. As a result, our method aids in the more accurate detection of fresh and decaying fruits.

5 Conclusion

This research experimented with a fine-tuned VGG-16 architecture based on deep-learning. The accuracy of the fine-tuned model was tested using an 80:20 splitted dataset of fresh and decaying apples, bananas, and oranges, which surpassed other previously available architectures. The impact of numerous additional factors, including the detailed analysis of activation function, can be investigated further in future research.

References

1. Aprea E, Helen G, Carlin S, Vrhovsek U, Mattivi F (2011) Metabolite profiling on apple volatile content based on solid phase microextraction and gas-chromatography time of flight mass spectrometry. J Chromatogr A 1218(28):4517–4524
2. Magwaza L, Opara UL, Nieuwoudt HH, Cronje PJR (2012) NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. Food Bioprocess Technology 5:425–444
3. Prabha DS, Kumar JS (2013) Assessment of banana fruit maturity by image processing technique. J Food Sci Technol 52(3):1316–1327
4. Ashok V, Vinod DS (2014) Automatic quality evaluation of fruits using probabilistic neural network approach. In: International conference on contemporary computing and informatics (IC3I), pp 308–331
5. Loncaric A, Pilizota V (2014) Effect of variety, growing season and storage on polyphenol profile and antioxidant activity of apple peels. Food Health Dis Sci-Prof J Nutr Diet 3(2):96–105
6. Li C, Li L, Wu Y, Lu M, Yang Y, Li L (2018) Apple variety identification using near-infrared spectroscopy. J Spectrosc
7. Bhargava A, Bansal A (2019) Automatic detection and grading of multiple fruits by machine learning. Food Anal Methods
8. Narwojsz A, Tanska M, Mazur B, Borowska EJ (2019) Fruit physical features, phenolic compounds profile and inhibition activities of cranberry cultivars (Vaccinium macrocarpon) compared to wild-grown cranberry (Vaccinium oxycoccus). Plant Foods Hum Nutr 74:300–306
9. Hena H, Sheikh M, Reza S, Marouf AA (2020) Mango species detection from raw leaves using image processing system. In: Smart trends in computing and communications: proceedings of SmartCom 2020, pp 183–192
10. Wang H et al (2015) Fruit quality evaluation using spectroscopy technology: a review. Sensor 15(5):11889–11927
11. Ronald M, Evans M (2016) Classification of selected apple fruit varieties using naive bayes. Indian J Comput Sci Eng 7(1)
12. Cornejo JYR, Pedrini H (2016) Automatic fruit and vegetable recognition based on CENTRIST and color representation. CIARP, pp 76–83
13. Jhawar J (2016) Orange sorting by applying pattern recognition on color image. In: International conference on information security & privacy, pp 691–697
14. Behera SK, Rath AK, Mahapatra A, Sethy PK (2020) Identification, classification & grading of fruits using machine learning & computer intelligence: a review. J Ambient Intell Humanized Comput
15. Rodriguez FJ, Garcia A, Pardo PJ, Chavez F, Luque-Baena RM (2017) Study and classification of plum varieties using image analysis and deep learning techniques. Progress Artif Intell
16. Pires TCS et al (2018) Antioxidant and antimicrobial properties of dried Portuguese apple variety (*Malus domestica* Borkh. cv Bravo de Esmolfe). Food Chem 240:701–706
17. Khaing Z, Naung Y (2018) Development of control system for fruit classification based on convolutional neural network. In: IEEE conference of Russian young researchers in electrical and electronic engineering
18. Ntasiou P et al (2015) Identification, characterization and mycotoxigenic ability of *Alternaria* spp. causing core rot of apple fruit in Greece. Int J Food Microbiol 197:22–29

Multifactor Biometric Authentication for Cloud Computing Security

Dhanshree More, Bhushan Deore, and Surendra Bhosale

Abstract Cloud services offer major advantage of the privacy of outsourced data. To alleviate this problem, sensitive data needs to be transmitted in encrypted form. The proposed scheme represents a way to maintain the confidentiality of the data offloaded from mobile devices with limited resources. The system developed makes use of the AES encryption algorithm to encrypt data, and uses a combination of biometric data of the user's face and iris for multi-factor biometric authentication to retrieve files from the cloud. The proposed system has three stages namely; preparation stage, identification stage, authentication stage. Our goal is to create an effective data encryption and authentication system without compromising data confidentiality. With proposed method processing time is reduced.

Keywords Multifactor biometric authentication · Open CV · LBPH algorithm · etc

1 Introduction

Cloud computing provides varied services through the Internet. These resources include tools and applications, admire information warehouses, servers, databases, networks, and computer code. We will use cloud storage to store files particle cloud server rather than storing files on our own drive or the other disk. As a device can access the network, it may also access data and software to control it. Multifactor authentication has recently become one in all the foremost effective strategies to

The original version of this chapter was revised: The belated corrections has been corrected in References 8 and 14 in this chapter. The correction to this chapter is available at https://doi.org/10.1007/978-981-19-0976-4_58

D. More (✉) · B. Deore · S. Bhosale
Department of EE, Veermata Jijabai Technological Institute, Mumbai, India
e-mail: mymore0@gmail.com

B. Deore
e-mail: bsdeore_p18@ee.vjti.ac.in

S. Bhosale
e-mail: sjbhosale@ee.vjti.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_32

supply additional security. Multiple layers make sure that the user requesting access is what they say. Even though cybersecurity steal credentials, they're forced to use other methods to verify their identity. In addition, by exploitation biometry admire fingerprints, iris, and facial patterns which will be captured by varied sensors, Master of Fine Arts identification is being widely utilized in several fields.

In biggest organizations, companies that store large amounts of sensitive data prefer files in the cloud because these data cannot be stored on hard drives that take up a lot of space and storage space. Many companies choose to store their data in the cloud because the cloud provides many storage advantages, which means that users can access data stored in the cloud at any time from any device with an identity or credentials. Existing systems use encryption algorithms to encrypt data with public and private keys to make it more secure, but this encryption method takes a long time because encryption is cumbersome and only uses biometric identification codes to verify user identity.

2 Literature Survey

Yue et al. [1], paper explains the correctness, safety and processing time required for the execution of the algorithm. The main algorithm of the system is implemented using python. The main algorithms of the system initialization, template creation, authentication request, template creation and matching are realized. The schemes used are OLPPBA and eFinga. Information users can use public parameters on their own devices to generate their requests in untrusted networks. It can reduce the confidentiality of the encrypted template and the impossibility of replacement, and solve the problems of the previous solutions. The OLPPBA program proved to be more effective than the eFinga program.

Zhu et al. [2], paper proposes a cheap and secure authentication theme that can counter user-initiated consent attacks and the cloud. In order to perform biometric authentication, the information owner encrypts the problem information and sends it to the cloud. Encrypt the database and return the result to the owner of the database. In biometric authentication, this role refers to fingerprints and uses finger codes to represent fingerprints. The attack model and system model are given. It provides a completely unique biometric authentication scheme that can protect privacy in cloud computing and resist potential attacks.

3 Proposed Scheme

The proposed scheme is primarily based totally on the present machine. Cloud protection is an extreme difficulty day through day. Many of the companies hesitate the use of cloud carrier due to cloud protection. The proposed machine makes use of

AES encryption set of rules with the most powerful non-public key. For authentication reason proposed machine makes use of multifactor biometric authentication (MFA) for authenticating legitimate consumer. If the right suit is determined with the consumer who needs to get entry to the cloud statistics in database, then the consumer is legitimate and may get entry to the statistics. For multifactor biometric authentication, machine makes use of mixture of face and iris popularity as this gives excessive accuracy.

3.1 Datasets

An employee's database collected from the eforexel.com site, consists of record of thousands of employee names, IDs, user names, passwords, genders, cities, departments, joining dates, etc. Human face database collected from vis-www.cs.umass.edu site, consists of facial images of individual employees working in the company.

3.2 Haar Cascade Classifier Using OpenCV

Haar's algorithm is also used to select or extract features from objects in an image, using edge detection, line detection, and center detection, around the eyes, nose, mouth, etc. in the image. Basic imaging functions and extract these functions for face detection. OpenCV offers a trainer and detector. We can use OpenCV to train classifiers on any object like cars, planes, and buildings. The cascade image classifier has two uses for training the cascade classifier opencv_haar training and opencv_traincascade. These two applications save the classifier in different file formats. Haar properties are properties of digital images that are used in object recognition. They owe their name to their intuitive resemblance to haar waves and were used in the first real-time face detector. Paula Viola and Michael Jones used the idea of the classifier of haar properties based on haar waves in their article entitled 'Fast Object Detection Using an Enhanced Cascade of Simple Features'. This classifier is widely used for tasks such as face recognition in the computer vision industry. Haar's cascade classifier uses a machine learning approach to visual object recognition that can process the images used by the detector to compute them very quickly. The learning algorithm is based on AdaBoost. Highly efficient classifiers and more complex classifiers are combined into a 'cascade' that discards any other region as a face in an image and thus expends more calculations for promising object-like regions.

As shown in the Fig. 1, the first the data owner who owns the biometric data of all users or employees in the organization. The most secure and unbreakable AES encryption is performed by the data owner to protect employee privacy or security and prevent data threats. Based on the owner's encrypted data, it is composed of

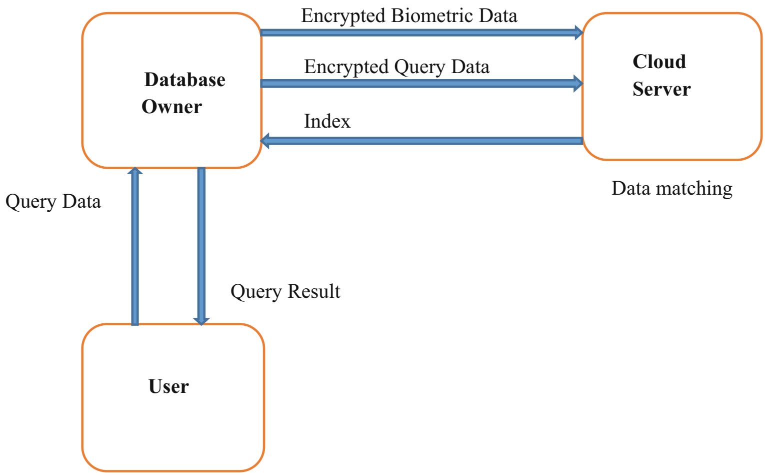


Fig. 1 Block diagram of proposed system

encrypted biometric data of all employees. Finally, the owner of the encrypted data uploads all the encrypted biometric data to the cloud server.

Data Owner:

In this module, the owner of the data uploads his biometric images with the data of its content to the server in the cloud, for security reasons the owner of the data encrypts and then stores it in the cloud and performs the following operations, such as uploading biometric images with encryption. Biometric images, check the biometric image details, and remove the biometric image details.

Cloud Server:

The cloud service supplier manages the cloud to supply storage services and per-form the subsequent operations, as an example save all biometric image files together with your face, view all biometric image files and their knowledge, view all biometric image comments, and look at all data house owners and users, and view all intruders.

User:

The cloud users whose great deal of knowledge is to be kept in cloud servers. The users can search the information and accessing the biometric image data if he/she approved and performs the subsequent operations corresponding to search biometric image, access biometric image and its details.

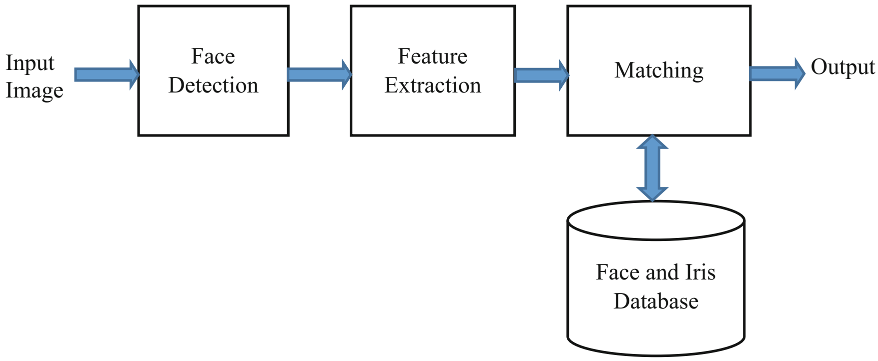


Fig. 2 Block diagram of face and iris recognition system

3.3 Face and Iris Recognition System

As shown in above Fig. 2, face recognition and iris recognition as multifactor biometric identity verification could be a technique of distinguishing or verifying the identity of a personal exploitation of their face and iris. Face recognition is want to establish individuals in photos, video, or in real-time. An automatic face identification scheme is a technology capable of identifying valid face from a digital image or a video frame against information of faces, generally utilized to manifest users through ID verification services, works by pinpointing and measurement countenance from a given image. Face recognition is a class of biometric security. Alternative types of biometric software system include voice recognition, fingerprint recognition, and eye tissue layer or iris recognition. The technology is generally used for security and law enforcement, although there's increasing interest in alternative areas of use. Iris recognition from eye came below face recognition. Firstly face recognition is finished when this eye and iris is detected by exploitation feature extraction.

3.4 LBPH Algorithm

Local Binary Pattern (LBP) is an easy however very effective texture operator. It labels the pixels with inside the photo with the brink of every pixel's community and treats the end result as a binary number. In addition, it changed into located that the aggregate of LBP and the Oriented Gradient Descriptor (HOG) histogram can extensively enhance the popularity overall performance of positive information sets. Combining LBP with histogram, we are able to use an easy information vector to symbolize facial photographs. The first calculation step of LBPH is to create an intermediate photo that higher describes the authentic photo with the aid of using extracting facial features. To the cease, the set of rules makes use of a sliding window idea primarily based totally on radio and neighboring item parameters. LBPH set of rules makes use of four parameters radius, neighbors, Grid X and Grid Y. First we want to educate the set of rules, so we want to apply dataset containing the face photographs we need to apprehend. We additionally want to set an ID for every photo in order that the set of rules can use this facts to apprehend the enter photo and offer it to output. Images from the equal individual should have the equal ID.

The first calculation step LBPH is to create an intermediate photo that higher de-scribes the authentic photo with the aid of using extracting facial features. To this cease, the set of rules makes use of a sliding window idea primarily based totally on radio and neighboring item parameters. According to the discern beneath, we can destroy it down into numerous small steps. This means, think we've got a grayscale photo of a human face. We can use part of this photo as a 3×3 pixel window. We also can think about it as 3×3 matrix containing the intensity (zero-255) fee is used to outline a brand new fee for eight neighbors. We set a brand new binary fee (threshold) for every neighbor of the common. We set 1 for values identical to or above the brink, and zero for values beneath the brink. The array now contains best binary values (no common fee). We should join every binary fee at every function with inside the array row with the aid of using row to a brand new binary fee (e.10001101). Then we convert this binary fee to a decimal fee and set it to the common of the array, that is genuinely one pixel far from the authentic photo. At the cease of this process (LBP process), new photo that higher displays the residences of the authentic photo will appear.

At this point, the set of rules has been trained, and every generated histogram is used to symbolize every photo within side the education dataset. Given the enter photo, we once more carry out the stairs for this new photo and create a histogram representing the photographs. For a photo that fits the enter photo, we best want to examine the two histograms and go back the photo with the nearest histogram. We can use special techniques to examine histograms (calculate the gap among histograms) along with Euclidean distance, chi-square, absolute fee, etc. The set of rules have to additionally go back the calculated distance, which may be used as a degree of 'self assurance'. Use threshold and self assurance to routinely examine whether or not the set of rules efficaciously detects the photo.

3.5 Algorithm is Developed Using the Tools Used

1. Python.
2. OpenCV Library.
3. NumPy Library.
4. OS Module.
5. OpenCV Video Capture.

4 Algorithm and Result

The following flowchart shows the process and proposed algorithm for implementing encrypted biometric multi-factor authentication to successfully verify the user's access to cloud computing.

As shown in the above Fig. 3, the user first logs in with a personal username and password to access data from the cloud. Login failure means that the user is invalid. After logging in, the user sends the requested data to the data owner. The data owner uses the AES algorithm to encrypt the requested data and sends the encrypted biometric data to the cloud server. Initiate identification by capturing real time images of a user through webcam. From captured images face and eye detection is done by LBPH algorithm. By training haar cascade classifier with number of images, face detection of captured images is done. By keeping 80% of training and remaining

Fig. 3 Proposed algorithm flowchart

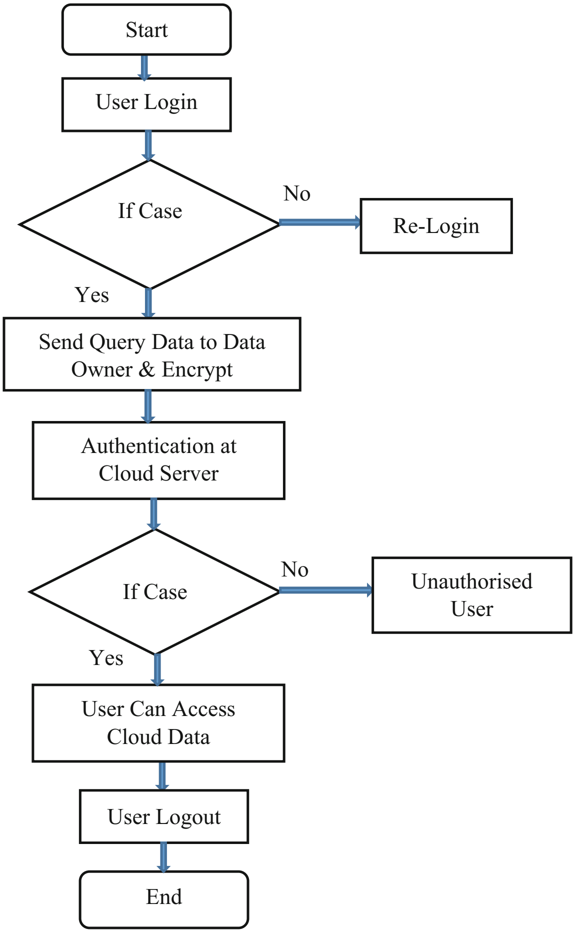


Table 1 Comparison of parameter used and processing time

Sr.No	Paper	Parameter used	Processing time
1	ZelunYue, Yiliang Han and Tanping Zhou	FingerCodes	1 min.67 s
2	Liehuang Zhu, Chuan Zhang, Chang Xu, Ximeng Liu, and Cheng Huang	Fingerprints	1 min. 37 s
3	Shengshan Hu, Minghui Li, Qian Wang, Sherman S.M. Chow and Minxin Du	Fingerprint and Face	53.517 min
4	Proposed algorithm	Face and Iris	1 min. 26 s

20% of testing, accuracy can be evaluated. For comparison use a threshold of 80% as the 'Confidence' and 'Valid user' or 'Invalid user' to estimate correct face and eye recognition. Verify if correct match not found, the algorithm output is 'Invalid user with accuracy'. Verify if the input captured images of user matches with the database images, the algorithm output is 'Valid user with accuracy' and the user will have full access to the cloud. Processing time required for the execution of entire algorithm is calculated using time module is 1 min 26 s as shown in below Table 1.

In the ZelunYue, Yiliang Hanand Tanping Zhou [1] paper, uses OLPPBA with FingerCodes and eFinga solutions. It reduces the confidentiality of the encryption template and the impossibility of exchange, and also solves the problems of the previous scheme. In this task, the system selects the biometric code measurement value from 100 to 1000 (interval of 100). The time required here is 1 min. 67 s., which is more for system-wide work than our solution. Processing time required for execution of complete algorithm also depends on system specifications.

Liehuang Zhu, Chuan Zhang, Chang Xu, Ximeng Liu, and Cheng Huang [2], proposed a fingerprint biometric authentication system, but it takes time to test or perform various identification steps, and this system takes longer than our system to perform. It takes 1 min and 26 s for the entire system to work.

Shengshan Hu, Minghui Li, Qian Wang, Sherman S. M. Chow and Minxin Du [3], for single solution the computation load has successfully moved to the remote server (RS). Processing time required only for user identification and authentication is almost greater than 53.517 min.

5 Conclusion

The proposed system provides the improved security of data exchange and data access through the cloud through strong AES encryption and biometric multi-factor authentication. Due to the use of multi-factor authentication, including user login, face recognition, iris and data comparison. Verify valid users. This reduces the time required to perform the computational complexity of the entire system, and also makes the authentication of large databases more convenient, making the system more efficient than the existing systems. In the key generation and distribution, this scheme is widely used, causing high computational costs to the data owner.

References

1. Yue Z, Han Y, Zhou T (2020) Efficient and privacy-preserving online biometric authentication scheme. In: Conference, pp 358–394
2. Zhu L, Zhang C, Xu C, Liu X, Huang C (2018) An efficiency and privacy-preserving biometric identification scheme in cloud computing. IEEE Access 19025–19033
3. Hu S, Li M, Wang Q, Chow SS, Du M (2018) Outsourced biometric identification with privacy. IEEE
4. Naresh V, GopiVenkata Ajay T, NagaSai Reddy T, Srinivas M (2020) An efficient and privacy preserving biometric authentication scheme in cloud computing. Journal 1966–1969
5. Liu C, Hu X, Zhang Q, Wei J, Liu W (2019) An efficient biometric identification in cloud computing with enhanced privacy security. IEEE Access 105363–105375
6. Chebrolu G, Dinesh S, Joqim D (2019) An efficiency and privacy-preserving biometric identification scheme in cloud computing. Journal 7521–7524
7. Kumar GS, Kandavel N, Madhavan K (2020) To discovery the cloud services authentication an expert based system using multi-factor authentication. IEEE
8. Mansour A, Sadik M, Sabir E, Jebbar M (2017) AMBAS: an autonomous multimodal biometric authentication system. In: 2017 13th international wireless communications and mobile computing conference (IWCMC), Valencia, Spain, 2017, pp 2098-2104. https://doi.org/10.1109/IWCMC.2017.7986607
9. Dharani, Priya L, Nagajothi (2019) A novel privacy preserving biometric identification scheme in cloud computing. Journal 1399–1404
10. Mendonca SN (2018) Data security in cloud using AES. Journal 205–208
11. Padma P, Srinivasan S (2016) A survey on biometric based authentication in cloud computing. In: Conference
12. Jiang Q, Ma J, Wei F (2016) On the security of a privacy-aware authentication scheme for distributed mobile cloud computing services
13. Han Z, Yang L, Liu Q (2017) A novel multifactor two-server authentication scheme under the mobile cloud computing. In: Conference
14. Mansour A, Sadik M, Sabir E, Jebbar M (2015) Multi-factor authentication based on multimodal biometrics (MFA-MB) for Cloud Computing. In: 2015 IEEE/ACS 12th international conference of computer systems and applications (AICCSA), Marrakech, Morocco, 2015, pp 1-4. https://doi.org/10.1109/AICCSA.2015.7507257
15. Alizadeh M, Hassan WH, Khodadadi T (2014) Feasibility of implementing multi-factor authentication schemes in mobile cloud computing. IEEE

Proposing a Novel Scheme for Improving LSB Image Steganography with Color Image

Annu Sharma, Kalpana Sharma, and Amit Kr. Chaturvedi

Abstract During the most recent two years of Covid-19, the vast majority of the work will be performed web based utilizing normal correspondence channel by associations including significant reports trade. Trading significant reports are unsafe to send straightforwardly, even we are seeing different extortion cases, data burglary cases, identity robbery, marital cheats, Internet shopping fakes, and so on. In these cases assailant takes your own data or reports and abuses them for some off-base purposes. To ensure your data or report during trading them online steganography is a valuable methodology. In this paper, we will propose a significant and inventive plan utilizing LSB method with shading pictures. To work on the security of the message, we will apply AES encryption plan and afterward utilize the picture steganography with LSB method.

Keywords Steganography · LSB · Cryptography · AES · Encryption · Security

1 Introduction

Data stowing away has acquired a lot of consideration from the most recent multi decade and particularly over the most recent two years during Covid-19 when the uses of data concealing instruments are seen more. Data can be hided either by utilizing the Steganography or Cryptography. In cryptography the data is scrambled utilizing crypto- graphic calculations like DES, RSA, AES, and so on and after encryption data will be transformed from its unique structure. The scrambled data won't be justifiable and any one can find that something is hided and afterward encoded to conceal the data. Assail-ants will catch such data and attempt to decode the data for uncovering the mysterious message. There are various methods of concealing data like content, picture, sound, or video under a cover. Concealing data under a

A. Sharma (✉) · K. Sharma
Computer Science Department, Bhagwant University, Ajmer, India
e-mail: anu.anusharmaajmer@gmail.com

A. Kr. Chaturvedi
CSE Department, Government Engineering College, Ajmer, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_33

picture cover is known as Image steganography. Image steganography is a developing exploration field that permits clients to insert touchy worries in pictures.

The LSB (Least Significant Bit) coordinating installing is displayed with regards to added substance commotion. It is accepted that the histogram of the secret message is a complication of the disorder likelihood bulk capacity (PMF) and the first capacity. The LSB strategy is extremely straightforward and unsurprising. As these pieces are least critical for the picture, on the off chance that we conceal a message in such pieces, the progressions comes in the pictures pieces of LSB doesn't reflect a lot of progress in the first picture. Consequently, the nature of the picture stays flawless. A shading picture has eight separate piece planes and first piece plane is known as LSB bit plane. In this paper, we have utilized this piece plane to conceal the mysterious message and to work on the security of the message content AES cryptographic calculation is utilized.

2 Related Work

A versatile LSB replacement technique utilizing uncorrelated shading space, expanding the assets of subtlety while limiting the odds of discovery by the human visualization framework. In the projected conspire, the information picture is gone over a picture scrambler, bringing about an encoded picture, safeguarding the security of picture substance, and afterward changed over to HSV shading space for additional preparing. The mysterious substance are scrambled utilizing an iterative enchantment network encryption calculation (IMMEA) for better security, delivering the code substance. A versatile LSB replacement strategy is then used to insert the scrambled information inside the V-plane of HSV shading model dependent on secret key-coordinated square wizardry LSB instrument. Using HSV shading space for information covering up is roused from its assets comprising de-relationship, cost-adequacy in preparing, better stego picture quality, and appropriateness for steganography as confirmed by our trials, contrasted with other shading spaces like RGB, YCbCr, HSI, and Lab. The quantitative and subjective trial aftereffects of the proposed system and its application for tending to the security and protection of visual substance in online interpersonal organizations (OSNs), affirm its viability rather than state-of-the-craftsmanship strategies [1].

The study of getting a data by encryption is Cryptography while the methodology for hiding secret messages in various messages is Steganography, with the goal that the secrecy's very occurrence is hidden. The word 'Steganography' represents the strategy for stowing away psychological substance in another medium to keep away from discovery by the gatecrashers. This paper presents two new techniques wherein cryptography and steganography are joined to encode the data similarly as to cover the mixed data in another medium so the manner that a message being sent is covered up. One of the strategies tells the best way to get the picture by varying above it into figure text by S-DES shrewdness operating a mysterious key and disguise this content in an-other image by steganographic approach. Other technique shows another method of concealing a picture in another picture by scrambling the picture straight by S-DES calculation using a key picture and the data got is covered in another picture. The proposed strategy forestalls the potential outcomes of steganalysis additionally [2]. Steganography is the craft of "covered up composition," and it alludes to procedures for disguising information inside apparently harmless things called "Cover Objects." There are different kinds of covers for scrambling touchy data, yet photos are omnipresent in regular applications and have a significant degree of repetition of portrayal. Subsequently, they are engaging possibility for use as cover things. This paper assesses the LSB calculation. Perhaps the most notable steganographic strategies is Least Significant Bit. Installing with another kind of encryption can make it safer. Large numbers of the researchers recommend the mix of encryption with Steganography adds more deterrents to get broken. The paper is intended to present the idea of LSB with relative code encryption procedure with the constraints of steganography [3].

The mysterious picture is first encoded by utilizing BLOWFISH calculation which has awesome execution and is a most remarkable strategy contrasted with different algorithms. Presently this scrambled picture is installed with video by utilizing LSB method of steganography. Our projected prototypical provides two layers of safety for privileged information, which completely fulfill the essential key elements of data security framework that incorporates: Confidentiality, Authenticity, Integrity and Non-Repudiation [4]. Cryptography and Steganography are two techniques for shielding information from interlopers while moving completed a sweeping channel organization. Cryptography is a technique to scramble information and steganography is the craftsmanship and study of concealing mystery message in a cover picture. A Hash Least Significant Bit (H-LSB) with Affine code calculation has been proposed for giving greater security to information in an organization climate. First we scramble the information with the new cryptography calculation and afterward insert in the picture. Eight pieces of the mysterious message are isolated into 3, 3, 2 and implanting into the RGB pixels up-sides of the cover picture separately. A hash work is utilized to choose the specific situation of inclusion in LSB bits. This framework permits a message sender to choose keys to scramble the mysterious message prior to inserting into the picture and a beneficiary is utilized by the keys to unscramble the message. Recipient can be decoded the scramble message with mistaken the keys however to an alternate structure from the first message. This framework can give better security while moving the mysterious message from one finish to the opposite end in network climate [5].

Steganography is the craft of concealing a message with the goal that a future snoop is unconscious of the message's quality. The goal is to obscure the occurrence of the communication from unapproved work force. It presents an errand of moving the inserted data to the objective without being recognized by the assailant. Numerous applications have effectively been produced for Steganography. The creators proposed to improve the application's transportability and convenience, and in this manner have fostered the application on the Android stage. Nath et al. have effectively proposed different approaches to scramble and shroud secret messages in various cover records. In this investigation, the mysterious message was encoded utilizing a calculation motivated from the Vigenère figure method and afterward the message was installed into the LSBs of the bytes of the cover record. It intends to shroud one byte of a mysterious message; the creators utilized the LSBs of eight bytes of the cover record without annihilating the property of the document essentially. The proposed bit trade is reversible, for example the unscrambling is done in the converse method of the encryption. The creators applied the current steganography calculation on picture documents and the outcome found was agreeable [6]. Steganography is utilized to move the information with picture procedure. Its simpler to move information safely without adjustment. The idea utilizing steganography presenting calculation "Upgraded LSB ALGORITHM FOR IMAGE STEGANOGRAPHY." In steganography it conceals secret message to the picture handling [7].

Computerized pictures are generally imparted over the web. The security of advanced pictures is a fundamental and testing task on shared correspondence channel. Different methods are utilized to get the advanced picture, like encryption, steganography and watermarking. These are the techniques for the security of computerized pictures to accomplish security objectives, for example classification, trustworthiness and accessibility (CIA). Exclusively, these strategies are not exactly adequate for the security of advanced pictures. Mixed security strategy is utilizing encryption, steganography, and watermarking. It involves three key segments: (1) the first picture has been scrambled utilizing huge mystery key by turning pixel pieces to directly through XOR activity, (2) for steganography, encoded picture has been modified by least critical pieces (LSBs) of the cover picture and got stego picture, then, at that point (3) stego picture has been watermarked in the time area and recurrence space to guarantee the proprietorship. The proposed approach is proficient, more straightforward and got; it gives critical protection from dangers and assaults [8].

Here they utilize the unchanged SIYu plan for the encryption and installing cycle and utilize 3-3-2 LSB inclusion strategy to insert information to cover picture. It has been shown here, how is the consequence of encryption measure and inserting measure utilizing SIYu map. Moreover, we will show our calculation execution dependent on arbitrariness of arrangement created by SIYu map with NIST test and entropy test, histogram investigation, decency of fit test, nature of the stego picture, nature of unscrambled information, starting worth affectability, and key space. The outcomes show scrambled information is consistently circulated; arrangement produced by SIYu map breezes through all NIST assessment and entropy test, which implies the created grouping is arbitrary; stego picture have a decent quality, it tends to be seen from the relationship coefficient and PSNR esteem in excess of 40 dB; decoded information has similar quality as the first information; beginning worth affectability comes to $10-16$; key space comes to 2.56×1094; and add commotions can be discretionary to build information privacy. For the end, our calculation can be accustomed to keeping information and data privacy on advanced picture [9].

Presently a day's sharing the data over web is turning into a basic issue because of safety issues. Subsequently more methods are expected to ensure the common information in an unstable channel. In steganography only the disguised correspondence which is like for securing our information from unapproved individual. In this procedure privileged intel can be supplanted or hide behind the cover data unsuspicious. Cryptography is the procedure, where privileged intel is supplanted in indiscernible organization. The current work is center around blend of cryptography and steganography to get the information while sending in the organization. Initially the information which is to be sent from sender to recipient in the organization should be scrambled utilizing the encoded calculation in cryptography. Besides the encoded information should be stowed away in a picture or video or a sound record with assistance of steganographic calculation. Thirdly by utilizing decoding procedure the beneficiary can see the first information from the secret picture or video or sound document. Communicating information or archive should be possible through these ways will be gotten. In this paper we executed three scramble strategies like RSA calculation alongside steganographic calculation like LSB replacement strategy [10]. This investigation suggests a combination of two safety approaches, in particular steganography and cryptography. The projected steganography strategy is LSB on the grounds that it enjoys benefits in information implanting limit and message impalpability. The LSB strategy is exceptionally straightforward and unsurprising, so to further develop communication safety HBV encryption methods are applied. HBV encryption is pragmatic to the communication prior to sticking it. HBV encryption is a blend of the Beaufort calculation and Vigenere figure so the message security will be more grounded. By consolidating HBV and LSB encryption, the message can be covered up appropriately, and it's conscious. In light of the exploratory outcomes, the encryption and inserting interaction can route sound and yield great quality encryption and stego pictures. Likewise, the extraction and unscrambling cycle can run consummately [11]. This investigation is tied in with getting message by carrying out the mix of steganography utilizing Least Significant Bit (LSB) strategy and cryptography utilizing Vigenere figure. The system use is cyclic cascade model. The stages additionally dont cover with one another. The testing includes the way toward inserting plaintext and code text into cover picture. The outcomes from the testing are the Mean Squared Error (MSE) and Peak Signal Noise Ratio (PSNR) worth of the stego-picture. The lower the MSE esteem and the higher the PSNR esteem is better picture quality [12].

3 Proposed Model and Methodology

The basic requirement to perform this experimental work with proposed novel model for image steganography using LSB (Lease Significant bit) bit plane using color image are (1) Color Image, (2) Secret Message, (3) Steganography Tool, (4) Cryptographic algorithm. So, in this experimental work the image used is shown in Fig. 1 and the secret message is displayed in Fig. 2.

Fig. 1 Cover image

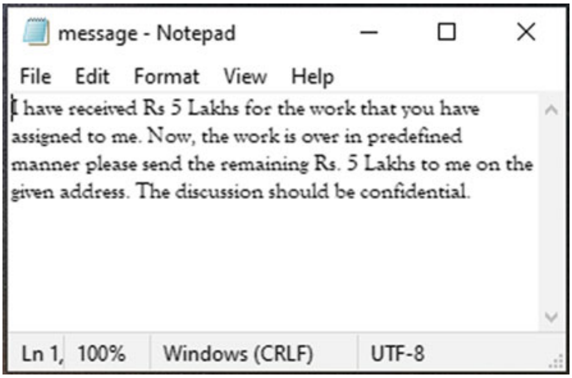


Fig. 2 Secret message

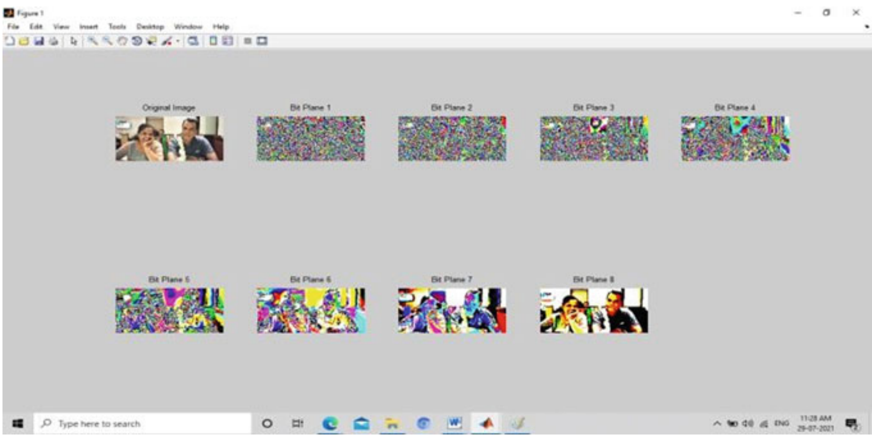


Fig. 3 Eight BitPlanes separation

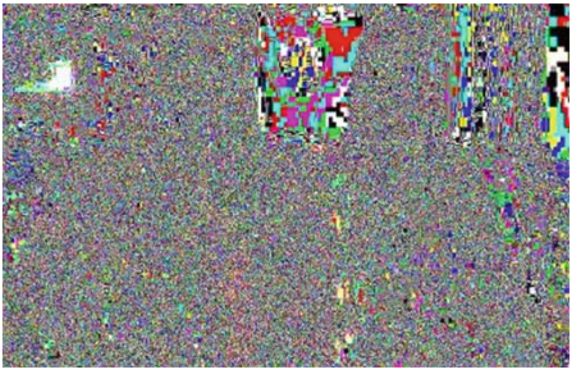


Fig. 4 BitPlane1or LSB plane of the cover image

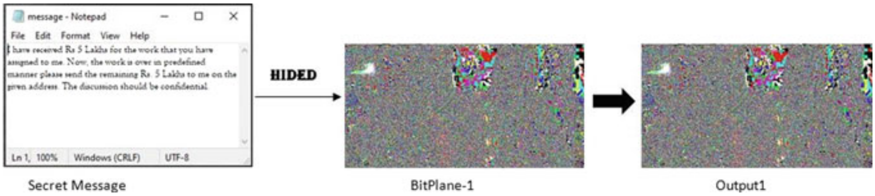


Fig. 5 Secret message hided in the BitPlane-1 and resulted as Output1

The color image is having 8 bit planes and these bit planes are separated by using bit plane separation algorithm and this process is presented in Fig. 3. The first bit plane i.e. LSB bit plane is first separated, which is shown in Fig. 4.

The secret message then encrypted using AES-128 algorithm to increase the safety of the communication and after encryption, the encrypted message then hided under the BitPlane1 i.e. LSB Plane of the colored cover image, which is a stego object and here we marked it as Output1. This process is shown in the Fig. 5.

Now we will recombine the Output1 and remaining seven BitPlanes to generate the final stego object as recombined image and process is illustrated in Fig. 6.

Recombined image is the final stego object to send by the sender to the recipient on the common shared channel like Internet and at the received end, the recombined image will be again separated into eight bit planes and Output1 is obtained. The secret message will be extracted from the cover i.e. Output1 using steganalysis. Here another very important point to be considered regarding security of the secret message is the secret key that has been applied on the secret message during the cryptography of the secret message. The same key will be used again when we perform steganalysis on the Output1 image, which is containing the secret message

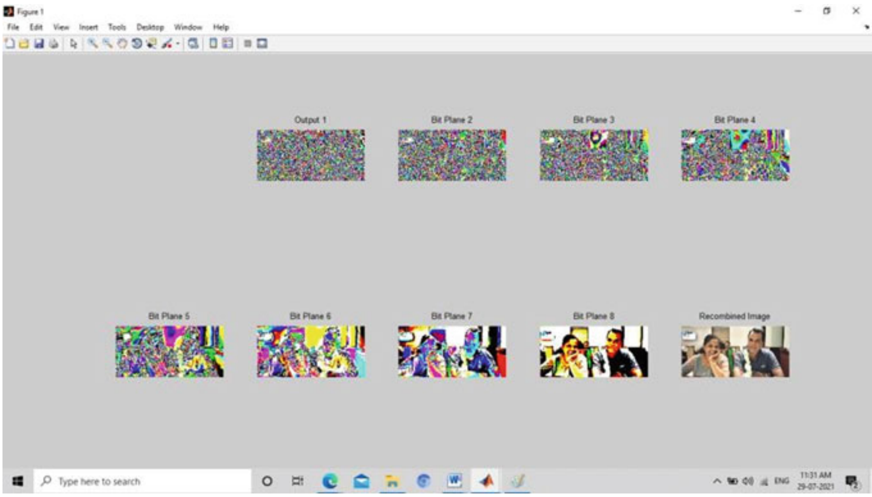


Fig. 6 Recombined image from Output1 and remaining seven bit planes

4 Conclusion

The necessity of message security is consistently a requirement for the general public, business, associations, or government. During the Coronavirus, it has been noticed the necessity of such methodology increments more through which the mysterious correspondence can happen proficiently and viably. There are numerous proposition put together by specialists however an alternate methodology is consistently a prerequisite forgot correspondence. The projected model is effectively supported and the represented in this paper. The creative thought that we skimmed in this paper is the utilization of LSB plane to shroud the message and afterward producing the stego object. The last stego object created through this methodology is safer to break and get the mysterious message than other comparable methodologies. Consequently, this proposition is helpful for steganography and effectively played out the ideal errand.

References

1. Mohammad K, Sajjad M, Mohammad I, Rho S, Baik SW (2018) Image steganography using uncorrelated color space and its application for security of visual contents in online social networks. Future Gener Comput Syst Issue 86:951–960
2. Narayana S, Prasad G (2010) Two new approaches for secured image steganography using cryptographic techniques and type conversions. Signal Image Process: Int J (SIPIJ) 1(2)
3. Verma A, Gupta A, Suman SK, Khan MM, Sarangi PK (2021) Encryption and decryption of images based on steganography and cryptography algorithms: a new model. Turk J Comput Math Educ 12(11):2839–2846

4. Sharma MH, MithileshArya M, Goyal MD (2013) Secure image hiding algorithm using cryptography and steganography. IOSR J Comput Eng 13(5):01–06, e-ISSN: 2278-0661, p-ISSN: 2278-8727
5. Abdullah AM, Aziz RHH (2016) New approaches to encrypt and decrypt data in image using cryptography and steganography algorithm. Int J Comput Appl (0975–8887) 143(4)
6. Nath A, Roy S, Gopalika C, Mitra D (2017) Image steganography using encrypted message. Int J Adv Res Comput Sci Manage Stud 5(4):7–11
7. Rupashini PR, Gowthami D, Hariprada P, Nandhini V (2017) Enhanced LSB algorithm for image steganography. Int J Comput Sci Eng (IJCSE) 6(04):111–114, ISSN: 2319-7323
8. Razzaq MA, Baig MA, Shaikh RA, Memon AA (2017) Digital image security: fusion of encryption, steganography and watermarking. (IJACSA) Int J Adv Comput Sci Appl 8(5):224–228
9. Satria Y, Suryadi MT, Cahayadi DJ (2020) Digital text and digital image encryption and steganography method based on SIYu map and least significant bit. In: ICOMPAC 2020, journal of physics: conference series, pp 1–11
10. Rote GD, Patil AM (2015) Steganography with cryptography technique for data hiding. Int J Sci Res (IJSR), ISSN (Online): 2319-7064, 4(1):2312–2318, Paper ID: SUB15849
11. Susanto A, Ignatius Moses Setiadi DR, Rachmawanto EH (2019) Dual security method for digital image using HBV encryption and least significant bit steganography. In: International conference on electronics representation and algorithm (ICERA 2019), pp 1–9
12. Zamri NM, Asraf SMH, Idrus SZS (2019) Two level security in delivering message using encryption and steganography techniques. In: Journal of physics: conference series, JICETS 2019, pp 1–9

Mining Minimal Non-redundant Image Association Rules Using Statistical Texture Feature Based Frequent Itemsets

Nishtha Parashar, Akhilesh Tiwari, and R. K. Gupta

Abstract A method for mining images using minimal non-redundant image association rules based on frequent itemset is presented in this paper. It is a two-phase approach. Firstly, we pre-processed the images and extract statistical texture features. In the second phase, a feature vector is created using statistical texture features which are then further optimized with the help of feature selection and discretization. Using this optimized feature vector, transaction database is prepared for mining image association rules using Apriori algorithm. We further apply HCIAR algorithm to obtain minimal non-redundant image association rules based on frequent itemsets. Outcomes of the experiment reveal that mining images is viable and gives highly correlated image association rules.

Keywords Image mining · Feature extraction · Feature vector · Image association rules

1 Introduction

Due to the massive advancement in the generation of digital images worldwide, the task of extracting information of use has become a critical task by the humans from the bulk of images being produced on daily basis. At this stage, the task of image analysis plays a key role in mining information of use from the digital images [1, 2]. Thus, image mining systems play a vital role in assessing semantically significant information by extracting knowledge from vast sizes image datasets. Extraction of implicit knowledge and correlation between image and data is considered as the means of image mining [3, 4]. Unlike, other image processing procedures the purpose of image mining is not to identify a particular pattern in an image. The emphasis is instead on discovering image patterns and deriving the knowledge from images within a substantial group of image datasets centered on the low-level pixel details [5, 6].

N. Parashar (✉) · A. Tiwari · R. K. Gupta
Madhav Institute of Technology and Science, Gwalior, M.P, India
e-mail: nishtha2909@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_34

In contrary to mining of images, frequent patterns associated with similar types of images can be used to generate image association rules which can further improve the procedure of image mining [7, 8]. The most selective features during the mining procedure are utilized to obtain the strongest association rules [9]. This paper presents a method for mining minimal non-redundant image association rules using texture-features-based frequent itemsets for the input image dataset.

2 Related Work

Among the most significant work done by the researchers in the field of image mining, Ribeiro et al. in [10] proposed a two-stage approach to improve the mammogram image diagnosis using association rules. It is comprised of two algorithms namely PreSAGe and HiCARe. Results prove that proposed approach is effective in the task of classification of mammogram images. PSO, SVM, and association rules-based method were proposed by Abdi et al. [11] for the diagnosis of erythemato-squamous disease resulting in classification accuracy upto 98.91%. A comparative analysis between field of data mining and image mining is provided by Tin et al. in [12]. A review to check how far the field of image mining can stand in medical science was made by Sollini et al. in 2019 [13]. A detailed review for the diagnosis of pancreatic diseases using image mining was made by Abunahel et al. in 2020 [14]. Neethu et al. [15] used the fusion of multimodal features of images and performed mining of image data using clustering and ARM techniques of data mining. Kumar et al. [16] performed the task of classification on a dataset of brain tumor MRI images using a hybrid approach. A combination of DWT, SVM, and genetic algorithm was used to analyze the results, resulting in improved accuracy.

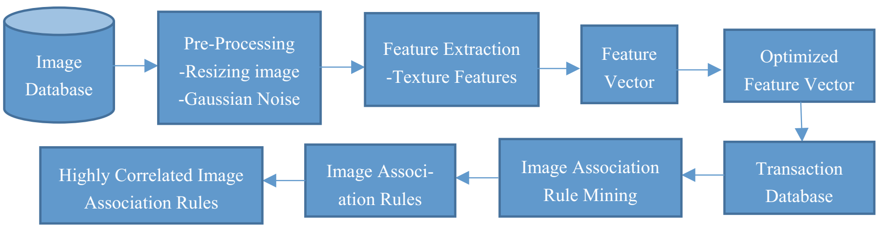


Fig. 1 Block diagram of image association rule mining method

3 Proposed Methodology

Figure 1 illustrates the block diagram of the proposed methodology in which the input is the image database, and the output is the highly correlated image association rules. Figure 2 illustrates the algorithm for the proposed methodology.

3.1 Pre-processing of Input Image

Initially, image resizing and gaussian noise removal is performed for the input images in the image database.

3.2 Extraction of Statistical Texture Feature and Generation of Feature Vector

From the pre-processed images the statistical texture features are obtained using gray-level co-occurrence matrix (GLCM) which is given by the comparative frequency of forthcoming of two gray level pixels x and y. Further these statistical texture features are transformed into feature vectors for further processing.

3.3 Formation of Optimized Feature Vector

This step involves the selection of statistical texture features and their discretization. This pre-processing step for image association rule mining intends to have least inconsistency in feature values. Thus, from the feature vector obtained in the previous

Algorithm 1: Proposed Image Association Rule Mining technique

Input: Images, min sup, and min conf
Output: Highly correlated image association rules.
1. Initialization of input images (Pre-processing).
    1.1. Perform image resizing.
    1.2. Perform removal of gaussian noise.
2. Obtain statistical texture feature of pre-processed image.
3. Construct feature vector.
4. Obtain optimized feature vector.
5. Construct transaction database.
6. Apply IARM using Apriori algorithm.
7. Extract highly correlated image association rules using HCIAR algorithm.

Fig. 2 Proposed image association rule mining technique step inconsistent features are eliminated and only most discriminatory features are chosen. Using these features optimized feature vector for every image is constructed and is provided as an input to the transaction database.

3.4 Formation of Transaction Database and Image Association Rule Mining

Keywords of input leaf images, i.e., green, or dry, and the optimized feature vector obtained in the previous step are provided as an input to the transaction database which is given as an input to the Apriori algorithm. In association rule mining, the expression X→Y represents an association rule where X and Y are the association rules [1]. The other two terms used in finding association rules are support and confidence ((1–2)). Support is a measure to find how frequently a particular item appears in the dataset. Confidence indicates the total number of times the if–then statements are counted true. These parameters are defined as,

$$\text{Support},\ S(X \to Y) = \frac{\text{No. of tuples (transactions) containing both } X \text{ and } Y}{\text{Total no. of tuples (transactions)}} \tag{1}$$

$$\text{Confidence},\ C(X \to Y) = \frac{\text{No. of tuples (transactions) containing both } X \text{ and } Y}{\text{Total no. of tuples (transactions) containing } X} \tag{2}$$

The values of transaction database are provided as an input to the Apriori algorithm. The minimum value for support and confidence is set through experimentation.

3.5 Extraction of Highly Correlated Image Association Rules (HCIAR Algorithm)

The mining of image association rules using Apriori algorithm provides all rules satisfying the minimum support and confidence threshold values. However, sometimes the rules with higher support and confidence values provide conflicting or redundant information, making it uninteresting rule. Thus, to overcome this limitation, we apply HCIAR algorithm described in Fig. 3, on all the image association rules obtained by the Apriori algorithm in the previous step. Using HCIAR algorithm the support and confidence are measured again for the generated rules from the frequent itemsets to obtain feasible outcomes and to find highly correlated rules generated in image mining.

Algorithm 2: HCIAR Algorithm (Extraction of highly correlated image association rule)

Input: m: total no. of rules, support, confidence
Output: Highly Correlated Image Association Rule
For each rule r_i, and $i \leq m$
Compute support (r_i), confidence (r_i)
    If support (r_i) > min_supp threshold and confidence (r_i) > min_conf threshold
        extract Highly Correlated Image Association Rule
End For

Fig. 3 HCIAR algorithm

4 Experimental Results and Discussions

For experimentation purposes, we use tilia-tomentosa green and dry leaves from leaf dataset created by Pedro Silva et al. [17] in 2013 (see Figs. 4 and 5). Initially, the pre-processing of image is being performed in which gaussian noise is being removed from the images. Next, statistical texture features from pre-processed images are being extracted by using Gray Level Co-Occurrence Matrix (GLCM) approach which is organized into feature vectors. Table 1 gives the measure of statistical texture features for each input image.

The generated feature vector for the obtained 10 statistical texture features is shown in Table 2. After performing feature selection and discretization the optimized feature vector is constructed using the most discriminative features as shown in Table

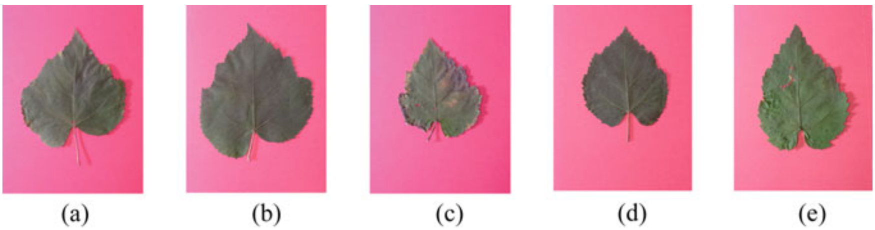


Fig. 4 Sample tilia-tomentosa green leaf

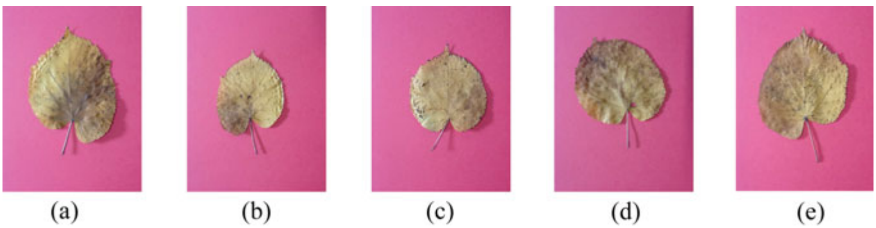


Fig. 5 Sample tilia-tomentosa dried leaf

Table 1 Measure of statistical texture features

Features	Equations
Contrast	$\sum_{i,j} \lvert i-j\rvert^2 p_{i,j}$
Dissimilarity	$\sum_{i,j=0}^{N-1} P_{i,j}\lvert i-j\rvert$
Homogeneity	$\sum_{i,j} \frac{p_{i,j}}{1+\lvert i-j\rvert}$
Angular second moment	$\sum_{i,j=0}^{N-1} P_{i,j}^2$
Max probability	$\max\left(p_{i,j}\right)$
Entropy	$-\sum_{i,j} p_{i,j}\log_2 p_{i,j}$
Energy	$\sum_{i,j} p_{i,j}^2$
GLCM mean	$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j})\mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j})$
GLCM variance	$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2$ $\sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j-\mu_j)^2$
GLCM correlation	$\sum_{i,j}(\frac{(i-\mu_i)(j-\mu_j)p_i p_j}{\sigma_i\sigma_j})$

3. Next, image association rule mining is performed using Apriori algorithm for the generated transaction database. We use 1001 as a keyword for green leaf image and 1002 as a keyword for dry leaf image. Thus, every input image has a keyword, i.e., green, or dry for six optimized feature values. Table 4 shows the transaction database for each input image which is given as an input to the Apriori algorithm for generating image association rules. The min_sup is set to 5% and min_conf is set to 70%. We obtained a total of 22 rules for this set of images.

Some of the image association rules mined are:

0.1→ 0.3, 2, i.e., green leaf (Support = 50%; Confidence = 100%).

0.3→ 0.1, 2, i.e., green leaf (Support = 50%; Confidence = 100%).

According to this rule, images with optimal feature value intervals of 0.3 and 2 are more likely to be green tilia-tomentosa leaf images.

0.005→ 0.04,0.002, i.e., dry leaf (Support = 50% and Confidence = 100%).

According to this rule, images with optimal feature value intervals of 0.04 and 0.002 are more likely to be dry tilia-tomentosa leaf images.

In next step, we perform extraction of highly correlated image association rules overall 22 image association rules attained by Apriori algorithm in prior step using HCIAR algorithm. At this step, we obtain 11 highly correlated rules for this set of images which can be used further for performing the task of image classification.

Table 2 Feature vector

Contrast	Dissimilarity	Homogeneity	Angular second moment	Max probability	Entropy	Energy	GLCM mean	GLCM variance	GLCM correlation
24.081	2.1028	0.5560	0.1193	0.3337	2.01172	0.3454	211.816	4352.703	0.9972
31.527	2.4453	0.5115	0.0976	0.3028	2.11631	0.3124	199.175	5205.804	0.9969
27.513	2.2303	0.5565	0.1263	0.3449	1.97123	0.3554	226.655	2794.210	0.9950
25.555	2.2729	0.4781	0.0466	0.1907	2.12329	0.2160	216.134	4325.348	0.9970
36.584	2.6839	0.4691	0.0443	0.1984	2.41423	0.2105	207.559	4619.284	0.9960
36.042	3.5469	0.3007	0.0018	0.00544	3.0868	0.0426	201.7697	1059.276	0.9820
26.272	2.7790	0.3563	0.0027	0.00682	2.8258	0.0528	208.1817	376.4889	0.9651
33.604	3.5157	0.2935	0.0017	0.00483	3.0513	0.0418	205.9018	673.8481	0.9750
20.285	2.8899	0.3234	0.0020	0.00713	2.9811	0.0457	215.4486	615.8355	0.9835
21.854	3.0039	0.3160	0.0016	0.00555	3.0908	0.040	190.946	973.272	0.9887

Table 3 Optimized feature vector

Dissimilarity	Angular second moment	Max probability	Entropy	Energy	GLCM variance
2.1028	0.1193	0.3337	2.01172	0.3454	4352.703
2.4453	0.0976	0.3028	2.11631	0.3124	5205.804
2.2303	0.1263	0.3449	1.97123	0.3554	2794.210
2.2729	0.0466	0.1907	2.12329	0.2160	4325.348
2.6839	0.0443	0.1984	2.41423	0.2105	4619.284
3.5469	0.0018	0.00544	3.0868	0.0426	1059.276
2.7790	0.0027	0.00682	2.8258	0.0528	376.4889
3.5157	0.0017	0.00483	3.0513	0.0418	673.8481
2.8899	0.0020	0.00713	2.9811	0.0457	615.8355
3.0039	0.0016	0.00555	3.0908	0.040	973.272

Table 4 Transaction database

1001	2	0.1	0.3	2	0.3	4000
1001	2	0.1	0.3	2	0.3	5000
1001	2	0.1	0.3	2	0.3	3000
1001	2	0.05	0.2	2	0.2	4000
1001	2.7	0.04	0.2	2	0.2	5000
1002	3.5	0.002	0.005	3.1	0.04	1000
1002	3	0.003	0.007	3	0.05	400
1002	3.5	0.002	0.005	3	0.04	600
1002	3	0.002	0.007	3	0.04	600
1002	3	0.002	0.005	3	0.04	1000

5 Conclusion

In this paper, we proposed a method for mining images using minimal non-redundant image association rules based on frequent itemsets. In the proposed approach, the image association rules obtained using Apriori algorithm provide conflicting or redundant information, making them uninteresting. Hence, to overcome this limitation HCIAR algorithm is applied to the rules generated from frequent itemsets to obtain highly correlated image association rules. Our results show that mining of images is viable using HCIAR algorithm and results in 11 highly correlated image association rules as opposed to the 22 image association rules obtained using Apriori algorithm. In future, we plan to test the proposed approach on a larger dataset with some modifications in the implementation. Moreover, the generation of highly correlated image association rules can still be explored in-depth for any implications in the field of image mining.

References

1. Gonzalez R, Woods R (1992) In: Digital image processing. 2nd edn. Pearson Addison-Wesley Publications
2. Zaiane OR et al. (1998) Mining multimedia data", CASCON: Meeting of Minds
3. Ordonez C, Omiecinski E (1998) Image mining: a new approach for data mining. College of Computing Technical Reports
4. Zaiane OR, Han J, Zhu H (2000) Mining recurrent items in multimedia with progressive resolution refinement. In: International conference on data engineering
5. Xie J, Wang C (2011) Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. Expert Systems with Applications, Elsevier, pp 5809–5815
6. Han J, Kamber M (2012) Data mining concepts and techniques. Elsevier, Third. Norgan Kaufmann
7. Thirunavu K, Manikandaprabhu P (2013) Function and information driven frameworks for image mining—a review. Int J Adv Res Comput Commun Eng 2(11)
8. Mangat V, Vig R (2014) Dynamic PSO-based associative classifier for medical datasets. IETE Technical Review, Taylor & Francis
9. Deshmukh J, Bhosle U (2016) Image mining using association rule for medical image dataset. In: Proceedings of international conference on computational modeling and security, Elsevier, pp 117–124
10. Ribeiro MX et al. (2008) An associatbased method to support medical image diagnosis with efficiency. IEEE Trans Multimedia 10(2)
11. Abdi MJ, Giveki D (2013) Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules. Engineering Applications of Artificial Intelligence, Elsevier, pp 603–608
12. Tin H (2017) Learning the comparison of image mining technique and data mining technique. In: Proceedings of 4th international conference on multidisciplinary research and practice
13. Sollini M, Antunovic L, Chiti A, Kirienko M (2019) Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. Eur J Nucl Med Mol Imaging 46:2656–2672
14. Abunahel BM, Pontre B, Kumar H, Petrov MS (2020) Pancreas image mining: a systematic review of radiomics. Europ Radiol
15. Neethu J, Wilson A (2014) Retrieval of images using data mining techniques. In: International conference on contemporary computing and informatics
16. Kumar S, Dabas C, Godara S (2017) Classification of brain MRI tumor images: a hybrid. Proc Comput Sci 510–517
17. Silva PFB, Marcal ARS, Almeida da Silva RM (2013) In: Evaluation of features for leaf discrimination. Springer lecture notes in computer science, vol 7950. pp 197–204

Machine Learning Based Crop Detection from Soil Images

Pallavi Srivastava, Aasheesh Shukla, and Atul Bansal

Abstract Recent soil categorization systems are designed to classify soil on the basis of texture, to predict soil moisture content, crop yield and soil ph, etc. These models depend on deep learning, machine learning, and image processing techniques to make the fieldwork less cumbersome for the farmers. The aim of this work is to process the soil images in order to produce a model for prediction of crops based on the soil images to benefit the rural farmers. Properties of soil affect the crop selection. This method is time-saving and cost-effective. It is beneficial to farmers in a way that they would be able to know the group of crops that can be cultivated in that soil from just a soil image. Soil images of three crops are considered in this work. Soil images are processed through different steps like preprocessing, segmentation of region of interest, and feature extraction. For classification, multi-class support vector machines with Gaussian kernel are incorporated with different values of gamma.

Keywords Soil classification · Multiclass · SVM · GLCM

1 Introduction

In the recent years, applications of domains like image processing and computer vision and in the field of agriculture have gained immense attention from researchers. It an interdisciplinary research domain that has various applications like fruit and vegetable grading, soil classification, weed detection, crop yield prediction, leaf disease detection, etc. Soil classification using machine learning and image processing algorithms is one of the most useful automated models. Conventional lab

P. Srivastava (✉) · A. Shukla · A. Bansal
Department of Electronics and Communication Engineering, GLA University, Mathura 281406, India
e-mail: pallavi.sr015@gmail.com

A. Shukla
e-mail: aasheesh.shukla@gla.ac.in

A. Bansal
e-mail: atul.bansal@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_35

methods like elutriation method, pipette method, decantation method, USDA triangle method, and Munsell color chart method are common to use. These methods are labor and time-consuming, this is one of the drawbacks of these methods. Also, the main beneficiaries of these methods are farmers and it must not always be possible for them to test the properties of soil that are in the farming fields. Therefore there is a need for automatic, inexpensive, low labor and time-consuming methods to help the beneficiaries. By applying different feature extraction techniques and machine learning classification tactics, one can build a model to classify soil on different basis which makes it easier for the farmers to know the various properties of soil which will be used for cultivating the crops. Some already proposed models related to soil using image processing and computer vision have discussed different objectives like soil texture (clay, silt, and sand) classification, soil color, soil moisture content, soil oxygen content, etc. Tian et al. [1] presented a general review on different computer vision technology in agriculture automation. Wang et al. [2] proposed digital image processing-based method to predict the utilization rate of nitrogen in Chinese cabbages. Morphological, textural, and color features are extracted from the images. The population is classified using k-means clustering algorithm and BPNN model for cabbage population quality. Srivastava et al. [3] presented an extensive review of various methods for soil classification based on computer vision and deep learning techniques. This paper also discusses methods adopted by authors to create soil image datasets.

Pegalaraj et al. [4] proposed a soil classification model based on munsell chart. Munsell chart is a soil color chart used extensively for soil classification. It has 238 standard soil colors. It is used by soil scientists as manual method to confirm the soil color. This method utilizes the color properties of soil from the soil images. HCV values are calculated using artificial neural network and fuzzy logic system is used to identify the similar munsell chart chip to an input soil image. Swetha et al. [5] proposed a system to predict components of soil in terms of soil, silt, and clay from soil images. Features extracted in this method are local, color, and texture features. Other than this, bag of visual words was also extracted. Also, CNN was used to predict soil, sand, and clay. R2, RMSE, residual prediction deviation (RPD), and RPIQ statistics were used to evaluate the model performance.

A comparative analysis was reported by Wu et al. [6] for defining the soil texture groups. For classification, various machine learning approaches were used. This model was fed a collection of topographical data collected from a watershed in the core regions of the Yangtze River's Three Gorges in southwest China. Artificial neural networks and support vector machines with Gaussian and polynomial kernel functions are used for classification. Under SVM-poly, the classification accuracy was 0.992 for loam, 0.794 for clay, and 0.661 for sand. Han et al. [7] proposed a smartphone-based soil color sensor for soil classification. In this method, visible spectrum and machine vision are incorporated to propose a low-cost and miniaturized soil color classification sensor for 10 different types of soil. Azizi et al. [8] presented a system based on deep learning in order to classify aggregates of any size into specific classes. This model could extract the features on its own from simpler to complex forms of features. CNN was utilized as it is best when working on images. VggNet16, RestNet50, and Inception-v-4 architects are used to train CNN. RestNet50 observed highest accuracy 98.72%. Barman et al. [9] prosed a system to predict soil ph value over a self-created soil image dataset from the soil samples collected from Guwahati, Assam, India. Color features are extracted and value of ph is predicted using logarithmic regression, exponential regression, and quadratic regression. Morais et al. [10] developed an ecologically friendly MIA-based analytical approach for determining the statistical assessment of soil's primary elements (silt, clay, and sand) using digital image processing and multi-variable calibration. Barman et al. [11] established a digital method for predicting soil pH that is both easy and exact. In this procedure, a soiled image is recorded using a smartphone, after which the photos are segmented using the K-Means clustering and HSV attributes are generated using color image processing. The pH of the soil is calculated using a pH meter in a laboratory. On the basis of textural properties of the soil pictures, Barman and Choudhury [12] introduced another soil classification method that used multi-class support vector machine with linear kernel function. The photographs for this experiment were taken with a smartphone camera, and the area studied was West Guwahati. In comparison to traditional laboratory procedures, this technology takes less time to classify the soil type and is more accurate. This system's SVM classification rate is reported to be 95.2%.

2 Materials and Methods

2.1 Soil Sample Collection

The suggested work considers three different crops, namely mustard, potato, and sugarcane. For each crop 10 fields of size 350×300m are considered to collect soil samples. From each of the 10 fields, 3 soil samples are collected. So, there is a collection of total 50 soil samples of each crop taken from 10 different fields of a crop. Samples are collected from district Mau, Uttar Pradesh, India. The sample collecting site is located between the latitudes of 25°56′30"N and 83°33′40"E. Each field is sampled five times, with a 150 m gap between them. Samples are taken from a depth of 2 inches below the field's surface level.

2.2 Soil Image Acquisition

The action of obtaining soil images from soil samples for further processing is known as image acquisition. To acquire soil photos, a Motorola One Power smartphone running Android 10 was used. This smartphone camera has 16-megapixel CMOS device. At the time of capturing image, camera settings are kept in default condition such as exposure time = 1/30 s, F-stop = f/1.8, focal length = 1.12 µm. Images are

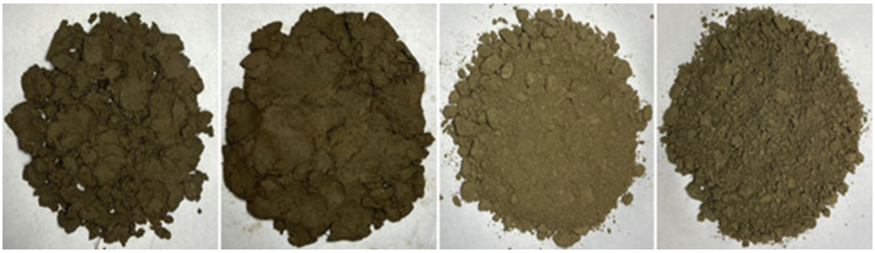


Fig. 1 Some examples of soil samples from soil image dataset

of size 3024 × 4032. The setup was kept inside a closed room, samples of soil were captured in artificial lighting conditions using a table lamp. The soil images were clicked from 20 inches above. Soil samples were taken out on a white paper and arranged in the middle of it to capture the images. Soil samples are taken from fields of three different crops. Here for every crop 10 fields are considered and from every field, 3 soil samples are taken with a separation of 150 m. So, there are 50 samples per crop. At the time of creating soil image, dataset eight pictures are captured per soil sample. Figure 1 shows some images of soil samples. Dataset is created using an android smartphone because farmers and other beneficiaries may not have good quality cameras.

3 Proposed Methodology

3.1 Preprocessing and Segmentation of Soil Images

Data cleaning or data preprocessing is an important step and it takes good amount of time in preprocessing before building the model. Few examples of data preprocessing involve missing value problems, outlier detection, and removing unwanted (noisy) data. Likewise, image preprocessing is a lowest step of abstraction. The objective of preprocessing is to improve the image data by subduing distortions and amplifying image features relevant for further processing and analysis. In this work, for preprocessing the soil images gamma correction is used. Gamma correction is employed to rectify the dissimilarities among the way a camera captures the information, the way a display displays it, and the way human visual system processes light. Human does not respond to light in the similar way a camera captures it. Gamma correction controls the general brightness of an image. Gamma correction is used if the image is bleached out (too dark). In order to reproduce colors precisely, a right value of gamma (γ) should be selected. When the degree of gamma correction changes, the ration of red to green to blue also changes in an image. As the intention to carry out this work is to predict the type of crop based on soil images, the soil dataset is created with a smartphone camera. The image dimension is 3024 × 4032 pixels with an average storage of 4.6 MB. Gamma (γ) value used is 1.2. After applying gamma correction, there is a visible increment in image tonal data values. This completes preprocessing step.

In this work, clustering-based segmentation is incorporated which is unsupervised method. As the name suggests, this method partitions an image into disjoint groups called clusters which have pixels that represent same features. This results in separation of data attributes into clusters where each cluster has similar elements compared to other clusters. Out of all of the clustering methods, k-means clustering is used here. It classifies an image into a number of clusters. The process starts by partitioning the image into k pixels which represents k group centroids. After this, each object is assigned to the group based on the distance between them and the centroid. And at last, when all pixels are assigned to all the clusters, the algorithm can move and reassign the centroids.

3.2 Feature Extraction from Soil Images

The process of extracting different textures from image is called feature extraction. Here, the features extracted from the soil image are presented as feature vectors corresponding to a single feature and then concatenate all the feature vectors into a main feature vector. Features extracted from soil images are gray level co-occurrence matrix. It is a tabulation of how often distinct gray level combinations co-occur in a soil image. The contents of the GLCM are used to calculate texture features. Haralick et al. [13], Soh and Tsatsoulis [14], Clausi [15] to give a measure of the variation in intensity at the pixel. In this paper, the GLCM was used to find the textural pattern, with help of these parameters the statistical texture analysis of the soil images. The GLCM features calculated in this are autocorrelation, contrast, correlation, entropy, IDM (inverse difference moment), etc. Total crops that are considered in this work is three and for each crop, there is a collection of 50 samples. During creation of soil image dataset eight images per sample were captured, so there are 400 images per crop. Therefore, the entire dataset has total of 1200 soil images of three different crops. And total of 22 features is calculated. Then the length of the main feature matrix for 1200 soil images is 1200×22.

3.3 Classification Using Multi-Class Support Vector Machine

When there are only two class labels in a classification problem, filtering the data, applying any classification technique, training the model with filtered data, and predicting the outcomes becomes simple. When there are more than two class occurrences in the input train data, however, analyzing the data, training the model, and predicting relatively correct results becomes more difficult. Multi-class classification can be used to handle these many class instances. As a model prediction, multi-class classification is a strategy that permits categorizing test data into several class labels existing in learned data. One-versus-All (one-versus-rest) and One-versus-One are the two most used multi-class classification approaches. In multi-class classification, the dataset has numerous class labels. As a result, the number of classifier models is determined by the classification method. The One versus. All categorization technique is employed in this paper. N-binary classifier models are built for the N-class examples dataset in One-versus-All classification. The number of class labels in the dataset and the number of binary classifiers constructed must both be equal. k-fold cross-validation is commonly used to solve classification problems, in which the folds are chosen so that each fold has nearly the same percentage of class labels. The proposed multiple classification models is evaluated using fivefold cross-validation in this study.

Classifier used in this work is multi-class support vector machine. SVM is applied with Gaussian kernel with four different values of gamma (γ). The objective is to locate a plane with the greatest margin, or the greatest distance between data points from both classes. Maximizing the margin distance gives some reinforcement, making it easier to classify subsequent data points. Hyperplanes are decision boundaries that aid in data classification. Different classes can be assigned to data points on either side of the hyperplane. The classifier's margin is maximized by using these support vectors. Here, in this work for Gaussian function, which is defined as-

$$k(x_1, x_2) = \exp\left(-\gamma \|x_1 - x_2\|^2\right) \tag{1}$$

where, $\|x_1 - x_2\|$ = euclidean distance between x_1 and x_2. Other parameters are C which is a regularization parameter to control error and γ which is only used for gaussian function.

3.4 Result Discussion

Both C and gamma parameters must be tuned simultaneously for a gaussian kernel. The influence of C becomes minimal when gamma is big. When gamma is small, C has the same effect on the model as it does on a linear model. In this work, values of gamma is changed as 0.5,1,1.5,2 and values of C is changed as 5,10,15,20,25. Results are calculated for every value of gamma and C corresponding with each other.

First, the results are recorded for fivefold cross-validation. At $\gamma = 0.5$, the performance of the model is evaluated for all the values of C. For C = 5 the output accuracy is reported to be 95.25%. For C = 10, accuracy is increased to 96.17%. Then for C = 15 and C = 20, accuracy is 94.12% and 93.16%, respectively. For C = 25, output accuracy is reported to be 92.01%. So, it is clear that for C = 10, this classifier performance is best when rbf kernel with $\gamma = 0.5$ is used. Table 1 shows the performance of classifier at $\gamma = 0.5$. Other parameters that define the performance of the model are also calculated. Recall and precision both values for this model is 96.17%.

Table 1 Performance of classifier at $\gamma = 0.5$ for different values of C for fivefold cross-validation

C	Accuracy(%)	Misclassification rate(%)
5	95.25	4.75
10	**96.17**	**3.83**
15	94.12	5.88
20	93.16	6.84
25	92.01	7.99

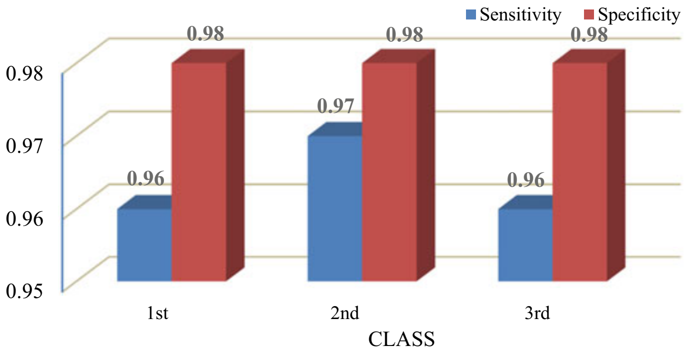


Fig. 2 Class-wise sensitivity and specificity of soil classification model at $\gamma = 0.5$

Also, the value of F1 score is 92.17%. Sensitivity and specificity are calculated class wise. Three crops whose soil are considered for this model are mustard, potato, and sugarcane, number of the class is in sequence to the name given. Figure 2 shows sensitivity and specificity for each of the three classes for $\gamma = 0.5$. In this figure, 1st class is mustard, 2nd class is potato and 3rd class is sugarcane.

At $\gamma = 1$, the performance of the model is evaluated for all the values of C. For C = 5 the output accuracy is reported to be 95.11%. For C = 10, accuracy is increased to 96.33%. Then for C = 15 and C = 20, accuracy is to 95.01% and 94.11%, respectively. For C = 25, output accuracy is reported to be 93.01%. So, it is clear that for C = 10, this classifier performance is best when rbf kernel with $\gamma = 1$ is used. Table 2 shows the performance of classifier at $\gamma = 1$. Recall and precision both values for this model is 96.33%. Also, the value of F1 score is 96.33%. Sensitivity and specificity are calculated class wise. Figure 3 shows sensitivity and specificity for each of the three classes for $\gamma = 1$. In this figure, 1st class is mustard, 2nd class is potato and 3rd class is sugarcane.

Table 2 Performance of classifier at $\gamma = 1$ for different values of C for fivefold cross-validation

C	Accuracy(%)	Misclassification rate(%)
5	95.11	4.89
10	**96.33**	**3.67**
15	95.01	4.99
20	94.11	5.89
25	93.01	6.99

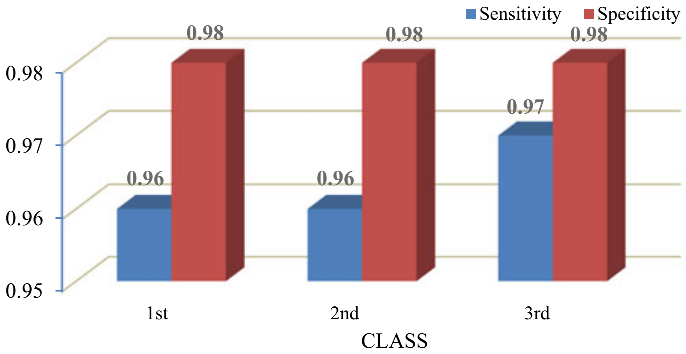


Fig. 3 Class-wise sensitivity and specificity of soil classification model at $\gamma = 1$

Table 3 Performance of classifier at $\gamma = 1.5$ for different values of C for fivefold cross-validation

C	Accuracy(%)	Misclassification rate(%)
5	91.57	8.43
10	92.15	7.85
15	**93.42**	**6.58**
20	92.06	7.94
25	90.51	9.49

At $\gamma = 1.5$, the performance of the model is evaluated for all the values of C. For C = 5 the output accuracy is reported to be 85.11%. For C = 10, accuracy is increased to 89.21%. Then for C = 15 and C = 20, accuracy is to 90% and 89.10%, respectively. For C = 25, output accuracy is reported to be 87.12%. So, it is clear that for C = 15, this classifier performance is best when rbf kernel with $\gamma = 1.5$ is used. Table 3 shows the performance of classifier at $\gamma = 1.5$. Recall and precision both values for this model is 93.42%. Also, the value of F1 score is 93.42%. Sensitivity and specificity are calculated class wise. Figure 4 shows sensitivity and specificity

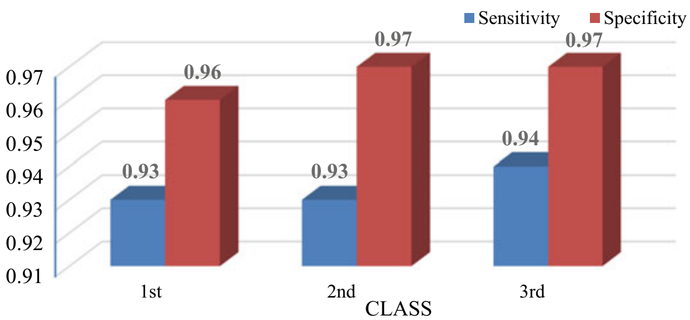


Fig. 4 Class-wise sensitivity and specificity of soil classification model at $\gamma = 1.5$

Table 4 Performance of classifier at $\gamma = 2$ for different values of C for fivefold cross-validation

C	Accuracy(%)	Misclassification rate(%)
5	88.11	11.89
10	89.21	10.79
15	**90.00**	**10.00**
20	89.10	10.90
25	87.12	12.88

for each of the three classes for $\gamma = 1.5$. In this figure, 1st class is mustard, 2nd class is potato and 3rd class is sugarcane.

At $\gamma = 2$, the performance of the model is evaluated for all the values of C. For C = 5 the output accuracy is reported to be 88.11%. For C = 10, accuracy is increased to 89.21%. Then for C = 15 and C = 20, accuracy is 90% and 89.10%, respectively. For C = 25, output accuracy is reported to be 87.12%. So, it is for C = 10, this classifier performance is best when rbf kernel with $\gamma = 2$ is used. Table 4 shows the performance of classifier at $\gamma = 2$. Recall and precision both values for this model is 90%. Also, the value of F1 score is 90%. Sensitivity and specificity are calculated class wise. Figure 5 shows sensitivity and specificity for each of the three classes for $\gamma = 2$.

In this figure, 1st class is mustard, 2nd class is potato and 3rd class is sugarcane. Figure 6 shows comparison in the support vector machines with different values of gamma.

The soil classification model is tested and validated for multi-class SVM with different values of gamma and C. Out of all the classifiers, SVM with $\gamma = 1$ exhibits best performance with output accuracy of 96.33% at C = 10. Next best performing classifier is SVM with $\gamma = 0.5$. Its best output accuracy is 96.17% at C = 10. At third position is SVM with $\gamma = 1.5$. Its best output accuracy is 93.42% at C = 15. And last performing is SVM is with $\gamma = 2$. Its output accuracy is 90% at C = 15.

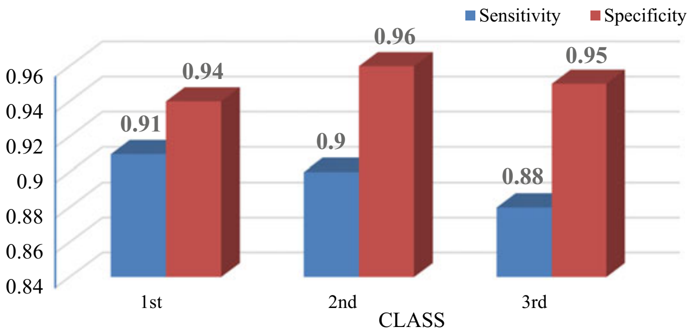


Fig. 5 Class-wise sensitivity and specificity of soil classification model at $\gamma = 2$

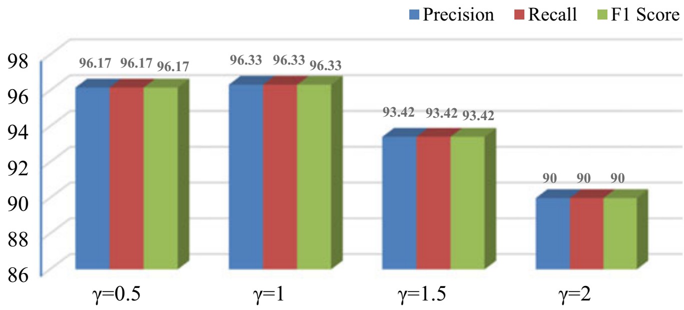
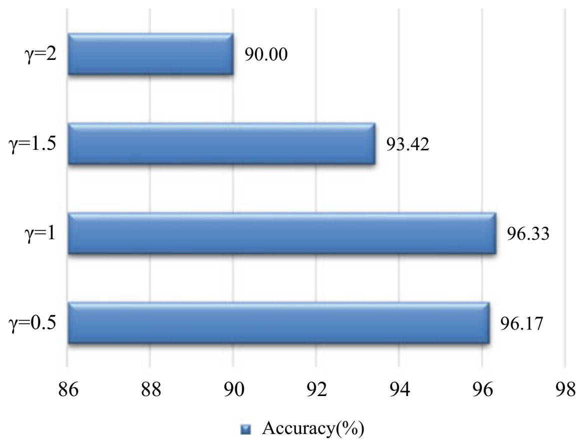


Fig. 6 Comparison of precision, recall, and F1 score for different values of gamma

4 Conclusion and Future Scope

In this paper, a model for prediction of soil on the basis of which crop can be grown into that soil is proposed. The soil samples are collected from 10 fields of size 350 × 300m in Mau, Uttar Pradesh, India. Here, soil of three crops namely, mustard, potato, and sugarcane are taken into consideration. For every crop 10 fields are considered and from each field, five samples are collected. So, there are a collection total 50 soil samples of each crop. The soil image dataset is created using a smartphone camera. It has total of 1200 soil images. Image dataset is created using smartphone camera because there may be no availability of better quality camera with farmers and other beneficiaries. 22 GLCM features are extracted using image processing techniques and fed to classifier. For examining the model performance, multi-class support vector machine is incorporated with one-versus-all method. Gaussian kernel function is used SVM with four different values of gamma. Figure 7 shows comparison of all the classifiers. This work can be extended with more crops into consideration. Also,

Fig. 7 Comparison of accuracy for different values of gamma



other than these machine learning techniques, one can also apply deep learning for the same objective.

References

1. Tian H, Wang T, Liu Y, Qiao X, Li Y (2020) Computer vision technology in agricultural automation—a review. Inf Process Agric 7:1–19. https://doi.org/10.1016/j.inpa.2019.09.006
2. Wang Q, Mao X, Jiang X, Dinstein I (2021) Digital image processing technology under backpropogation neural network and K-means clustering algorithm on nitrogen utilization rate of chinese cabbages. PLoS ONE 16:1–24. https://doi.org/10.1371/journal.pone.0248923
3. Srivastava P, Shukla A, Bansal A (2021) A comprehensive review on soil classification using deep learning and computer vision techniques. Multimed Tools Appl 80:14887–14914. https://doi.org/10.1007/s11042-021-10544-5
4. Pegalajar MC, Ruiz LGB, Sánchez-Marañón M, Mansilla L (2020) A Munsell colour-based approach for soil classification using Fuzzy Logic and Artificial Neural Networks. Fuzzy Sets Syst 401:38–54. https://doi.org/10.1016/j.fss.2019.11.002
5. Swetha RK, Bende P, Singh K, Gorthi S, Biswas A, Li B, Weindorf DC, Chakraborty S (2020) Predicting soil texture from smartphone-captured digital images and an application. Geoderma 376:114562.https://doi.org/10.1016/j.geoderma.2020.114562
6. Wu W, Di LA, He XH, Ma R, Bin LH, Lv JK (2018) A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Comput Electron Agric 144:86–93. https://doi.org/10.1016/j.compag.2017.11.037
7. Han P, Dong D, Zhao X, Jiao L, Lang Y (2016) Original papers a smartphone-based soil color sensor : for soil type classification. Comput Electron Agric 123:232–241. https://doi.org/10.1016/j.compag.2016.02.024
8. Azizi A, Gilandeh YA, Mesri-Gundoshmian T, Saleh-Bigdeli AA, Moghaddam HA (2020) Classification of soil aggregates: a novel approach based on deep learning. Soil Tillage Res 199. https://doi.org/10.1016/j.still.2020.104586
9. Barman U, Choudhury R, Talukdar N, Deka P, Kalita I, Rahman N (2018) Predication of soil pH using HSI colour image processing and regression over Guwahati, Assam , India. 10:805–809. https://doi.org/10.31018/jans.v10i2.1701
10. Augusto P, Morais DO, Souza DMD, Thaís M, Carvalho DM (2019) Predicting soil texture using image analysis. Microchem J 146:455–463. https://doi.org/10.1016/j.microc.2019.01.009
11. Barman,U, Choudhury RD (2019) Predication of soil pH using K mean segmentation and HSV color image processing
12. Barman U, Choudhury RD (2020) Soil texture classification using multi class support vector machine. Inf Process Agric 7:318–332. https://doi.org/10.1016/j.inpa.2019.08.001
13. Haralick R, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans pn Syst Man Cybernet 6:610–621
14. Soh LK, Tsatsoulis C (1999) Texture analysis of sar sea ice imagery using gray level cooccurrence matrices. IEEE Trans Geosci Remote Sens 37:780–795. https://doi.org/10.1109/36.752194
15. Clausi DA (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization. Can J Remote Sens 28:45–62. https://doi.org/10.5589/m02-004

Satellite Image Enhancement Techniques: A Comprehensive Review

Priyanka Darbari and Manoj Kumar

Abstract Satellite images are useful in various domain areas. In satellite images resolution is the major challenging issue because of huge distance, atmospheric changes, clouds, different air devices (drones, aeroplanes). Image enhancement is a technique of image processing for better image visualization, noise removing, other air facts. Image enhancement is applied after image acquisition using different sensors. In this paper, an introduction about satellite image acquisition, image enhancement techniques, survey of different techniques of satellite image enhancement is discussed with their advantages and disadvantages.

Keywords Satellite image acquisition · Image enhancement · Satellite image enhancement · Image fusion

1 Introduction

Satellite images are used in different areas like telecommunication, GPS, Navigation, weather or disaster forecasting, space behavior monitoring, agriculture monitoring, object recognition, sea behavior monitoring, space behavior monitoring, etc. Information about earth surface taken by remote sensing artificial satellite images. These images have lots of noise, distortion, low resolution, etc.

Satellite images are acquired by remote sensors. Remote sensors collect data about the earth's surface using their sensing ability and recoding their emitting or reflecting energy for pre-processing analyzing for better and accurate results or information.

The motive of this paper is to analyze different techniques for processing Satellite images using enhancement for better visualization and how sensor collects data and their resolution types.

This paper is organized into four sections. First is the introduction section followed by the second section of how images are acquired using remote sensors. The third

P. Darbari · M. Kumar (✉)
Department of CEA, GLA University, Mathura, UP 281406, India
e-mail: manoj.kumar@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_36

section is what is image enhancement techniques and their types. The fourth section is a Literature Review of Satellite image Enhancement.

2 Image Acquisition

Remote Sensing Satellite Images is evaluated based on digital Number (DN) which defines details of image brightness, color, wavelength, etc. In Digital Image Processing, the smallest unit of an image is a pixel which arranged in a row and column matrix called a raster image.

Nowadays, Satellites are revolving around the near-polar orbit for gathering data or information about the earth's surface. Sensors are two types: Passive sensors which detect and measure the naturally occurred reflected energy with the help of daylight (Sun's energy), not good at night. Active Sensors which detect and measure their own illuminated energy for investigation purposes, good in both day and night time.

Satellite Image Acquisition is the most important task for making the decision in various fields. The raster data of satellite images is the combination of row and column of pixels. These pixels give all the dimensions and information about the images for better resolution [1].

Figure 1 defines the different resolutions of sensors. Satellites provide many remote sensing imageries. The unique characteristics of Satellites are useful for remote sensing of the Earth's surface. There are different types of satellites are available like thermal sensors, microwave both active and passive sensors, airborne and spaceborne sensors radar is used for measuring electromagnetic distance.

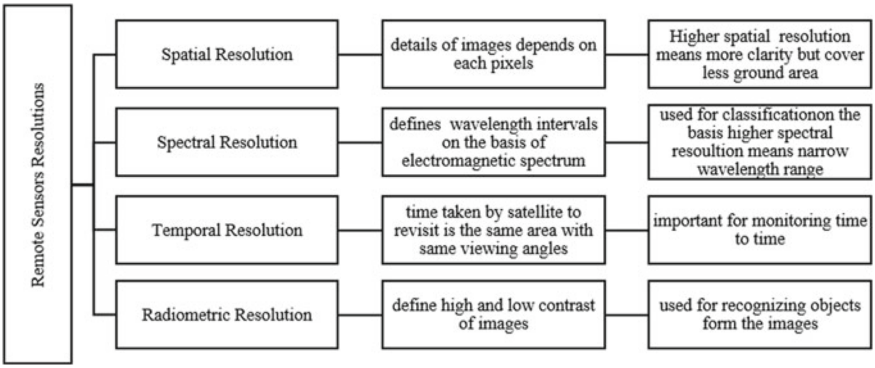


Fig. 1 Image acquisition resolution

3 Image Enhancement

Image enhancement is one of the most important techniques in the image processing research area. The main aim is to enhance the quality and visual appearance of an image or to provide a better transform representation for future processing. In many types of images like medical images, satellite, aerial images and also real-life photographs facing poor and bad contrast and noise. It is necessary to enhance the contrast and remove the noise to increase image quality.

One of the most important stages improves the clarity or digital quality of images for human viewing, removing blurring and noise, increasing contrast, and revealing details. The techniques of enhancement differ from one field to another depending on their objective. The existing techniques of image enhancement can be classified into two categories: Spatial Domain and Frequency Domain Enhancement [2, 3].

With the help of different techniques of Image Enhancement, the author can understand, how to solve noise, distortion kinds of problems from the images. Table 1 define types of image enhancement techniques with types and subtypes.

4 Literature Review of Satellite Image Enhancement

For the better resolution of satellite images, need image enhancement techniques. The objective of the paper is to provide a review of different techniques of satellite image enhancement with their pros and cons. For the literature survey, reviewed 11 quality papers and based on different perceptions result in section is created. With the help of the literature review Table 2, the author can identify the problem of satellite images enhancement and decide the future scope for better results.

On the basis of literature review of different satellite image enhancement techniques, frequency domain image enhancement gives better results but due to different frequency detection for noise distortion blurring ringing, etc. is quite difficult. So that spatial domain is the only way to analyze satellite images. If the spatial domain is used for enhancement, then image fusion of different satellite sensors can be fused and then enhanced their features for better results.

5 Discussion

As per the review Table 2, Satellite image enhancement faces geographical issues such as clouds, fog, sunlight, weather which create shadow or dark regions but it changes from time to time. All these techniques help to recognize advanced execution like segmentation and classification and also help to monitor change in the earth environment for evaluation.

Table 1 Description of image enhancement

Image enhancement techniques					
Spatial domain [4]:					
Description:	Deal with image pixel				
		Types:	**Point process operation** **Spatial filter operations** **Piecewise linear transform**		
			Point process operation:		
				Description: Also called Inverse transformation function perform in single pixels	
				Types of point process operation	
					1. Negative image **2. Image threshold** **3. Log transform** **4. Gamma transform**
					A negative gray level called a **negative image** useful for enhancing the white or gray level of dark images **Image Threshold**—normalized gray level of pixel either 0's or 1's called binary images **Log Transform**—expand the value of low gray level and compress the value of dark gray level **Inverse Log Transform**—expand the value of dark gray level and compress the value of bright gray level **Power law (Gamma) Transform**—a narrow range of dark pixel values into a wider range and wider ranges of bright pixel values to a narrow range
			Spatial filter operations:		
				Description: It is also called neighborhood operation-pixel value is operated on its neighborhood pixels	
				Types of spatial filter:	
					1. Linear filter **2. Non-linear filter**

(continued)

Table 1 (continued)

					Linear filtering—process of mask filter perform in linear way-point to point in an image-also called smoothing filter **Non-linear Filter**—process of mask filter performs in the non-linear way-like ranking of pixels-also called sharpening filter
			Piecewise linear transform		
				Description: It is the type of gray level *transformation* that is used for image enhancement	
				Types of piecewise linear transform:	
					1. Contrast stretching **2. Intensity level slicing** **3. Bit plane slicing**
					Contrast stretching—processing an image in a different region (dark or bright) to make it visible to humans **Intensity level slicing**—highlighting the specific range of intensities for desirable images **Bit plane slicing**—highlighting specific bits of total images which contribute more suitable for detailing of images
Contrast enhancement [2]					
Description:	The ratio of maximum gray level to the minimum gray level				
		Types:	**Histogram equalization(HE)** **Tone mapping**		
			Description: HE calculates the intensity of pixel gray level for increasing the visual appearance Description: Tone Mapping is a technique of high dynamic range image processing and computer graphics to map one set of colors to another set of colors		
			Types of HE:		

(continued)

Table 1 (continued)

					1. BBHE, **2. BPDHE,** **3. AHE,** **4. SR,** **5. CLAHE**
					Brightness Preserving Bi-HE—split images into two independent histograms so that intensities are equally distributed Brightness Preserving **Dynamic HE**—provide means brightness and means intensities due to partitioning of images called sub-histogram **Adaptive HE**—evaluate several histograms and each histogram correspond to a specific section of image for reducing noise and improving contrast Contrast Limited AHE—compute contrast of grayscale of a small set of images called tiles and then using interpolation to combine neighboring tiles for enhancement **Stochastic Resonance (SR)**—use external noise to enhance the contrast of images
Frequency domain [3]					
Description:	Processing of images done on the basis of frequency content				
	Types:	**Image smoothing**: Low pass filter block high-frequency bands			
		Image sharpening: High pass filter block low-frequency bands			
		Bandpass filter: Combination of low pass filter and high pass filter—it removes very low and very high-frequency components or band			
		Periodic noise-reducing: Electromechanical or electrical interference during image acquisition			
		Fourier transform: Separation of different frequency components from the images and then apply filtering (low or high) perform Inverse Fourier transform to get an original image without loss:			
			Types of fourier transform:		
				1. DST-Discrete Sine frequency transform 2. DCT-Discrete Cosine Frequency transform called as phase 3. Orthogonal transform based on magnitude and phases	

Table 2 Review of satellite image enhancement

Literature review of satellite image enhancement techniques					
Title: **Title**:	Overview of satellite image resolution enhancement technique Wavelet transform domain methods for resolution enhancement of satellite images				
	Transaction:	IEEE [5–7], Springer [8]			
			Method:	1. DWT-Interpolation-Bicubic IDWT 2. SWT	
				Advantages	
					High resolution
				Disadvantages	
					DWT method loss high-frequency component due to downsampling in subband
				Results	
					SWT gives the best result compared to DWT
Title: **Title**:	Improve techniques of satellite image enhancement A survey of satellite image enhancement techniques				
	Transaction:	IOP [9] IJAIR [10]			
			Method:	Contrast image enhancement	

(continued)

Table 2 (continued)

				Using 1% top and the bottom intensity value of original image and check dark and light image by adjusting range [0–255] Tool- MATLAB 2018 a, open access seven images of Kolkata	
				Advantages**:**	
					For validation—use histogram in original and contrast image
				Disadvantages:	
					Dark and light images adjustment is difficult for better quality
				Results:	
					Check efficiency and quality of image use entropy
Title:	Analysis of image enhancement techniques used in remote sensing satellite imagery				
	Transaction:	IJCA [11]			
		Method:	Spatial frequency domain Contrast enhancement		

(continued)

Table 2 (continued)

			(LPF-median filter and HPF-subtraction and derivative method) frequency domain and color enhancement technique (apply HE in contrast stretching, coloring and HIS calculation) to find hidden objects due to clouds, tree, buildings, mountains, and flying objects Tools-QuickBird open-source satellite imagery data set. enormous image-ENVI 5.0 software-image enhancement-MATLAB 9.2 tool	
			Advantages:	
				Spatial LPF—median filter generates nice smoothing compared to adaptive and moving average filter **Both HPF**—method generated equal sharpening Color HE gives a more visible image because of increased intensity and RGB high brightness **Gaussian stretch** re-introduces the contrast region **Frequency domain**—discrete Fourier Transform-low pass and high pass filter implemented by ideal low and high pass filter and Butterworth, gaussian filter plot graph based on SNR and PSNR for more qualitative measurement of distortion and replication of the image **Notch (high pass) filter** known as band stop filter sharpen the edge and increase intensity **Color Transform**—pseudo color transform gray level to RGB for finding populated area and identifying hidden elements. HIS transform RGB to HIS color like red green orange magenta to find high intensity
			Disadvantages:	

(continued)

Table 2 (continued)

				HIS also gives high brightness but some objects are still invisible because of high illumination Gray level slicing increase the gray level of the image but discard other information Ideal low and high pass filters are practically impossible to implement, so create a performance matrix
			Results:	
				A high pass filter helps to identify the hidden object and populated area
Title:	An effective method of satellite image enhancement			
	Transaction:	ICCCA [12]		
		Method:	DWT-interpolation-bilinear applied to Low-frequency input image and bicubic applied to High-frequency-IDWT	
			Advantages:	
				For experimental analysis-MSE, PSNR, Entropy is calculated with WZP, CWT, Cycle Spinning, DWT for analysis
			Disadvantages:	
				Low-frequency subbands contain low information consider less information. That is why th e original image is used for bilinear interpolation
			Results:	
				Interpolation half applied to the high-frequency image and a half to the original image for better results
Title:	Image enhancement using dual-tree complex discrete transform based on cycle spinning methodology			
	Transaction:	IEEE (ICCIC) [13]		

(continued)

Table 2 (continued)

		Methods:	DT-CWT—bicubic interpolation	
			DT-CWT—decompose six different subband -to obtain high-resolution image-apply Bicubic interpolation in high-frequency subband- apply spatial shifting in the horizontal and vertical direction to generate N low-resolution images and discard high-resolution image-cycle spinning applied to remove artifact and increase the quality of low-resolution image-again apply DT-CWT using bicubic interpolation in the low-resolution image to generate N high-resolution images from low-frequency image and apply inverse spatial shifting and calculate the average of N resolution image	
			Advantages:	
				Compare with wavelet domain zero paddings, dual-tree complex wavelet transforms, wavelet domain zero-padding cycle spinning, discrete wavelet transform, DWT and SWT, DASR, Low-resolution input image obtained by the sampling of high-resolution image
			Disadvantages:	
				Time-consuming
			Results:	
				High-resolution image using the low-frequency image as an input
Title: **Title**:	Survey of satellite image enhancement An extensive survey on satellite image enhancement			
	Transaction:	IJCAT [14] ICSSIT [15]		
		Method:	Spatial domain-HE, BBHE, BPDHE, AHE, SR, CLAHE Contrast Enhancement DWT based DSR	
			Advantages:	

(continued)

Table 2 (continued)

				Histogram Equalization increase pixels in dynamic range to increase the appearance of an image Brightness Preserving **Bi-HE**—the brightness of an image is much better than HE Brightness Preserving Dynamic Histogram Equalization (**BPDHE**)—provides realistic images and give equal intensity Adaptive histogram equalization (**AHE**)—improve grayscale contrast and color image contrast, reduce noise Stochastic Resonance (**SR**)—increase external noise to provide better contrast Contrast Limited Adaptive Histogram Equalization (**CLAHE**)—remove noise using amplification Contrast Enhancement (**CE**)—brighten the dark or unclear images to improve quality and clarity Adaptive DWT based Dynamic Stochastic Resonance (**DSR**)—perform enhancement in very dark images using internal and external noise and computational complexity is less, enhance image without artifact, ringing, blocking
			Disadvantages:	
				HE is Not suitable for color image **BBHE** take time for processing **BPDHE** does not give clear contrast images **AHE**—Increase unwanted blurring in edges of the image **SR**—applied only low contrast images **CLAHE**—applied only a small region of the image called tiles **CE**—take more computational time
			Result:	
				All techniques give a better structural appearance and improve image quality AHE and CLAHE is good for remote sensing application
Title:	An investigation in satellite image based on image enhancement techniques			

(continued)

Table 2 (continued)

	Transaction:	Taylor and Francis[16]		
		Method:	Image fusion-component substitution like HIS, PCA, Multi-Resolution like wavelet, curvelet, Numerical techniques like Brovey Transform, Statistical Techniques like Fuzzy clustering, Optimize Techniques like DNN, CNN and non-fusion techniques like Spectral mixture analysis, learning base like backpropagation NN, matting techniques like KNN, FIHS for satellite image enhancement	
			Advantages:	
				Image fusion gives a better result like high spectral distortion, low spatial resolution, reduced contrast and sharpness,
			Disadvantages:	
				Non-fusion is not good because of low spatial resolution and high computational complexity
			Result:	
				Fusion is better than non-fusion
Title: **Title**:	Image fusion techniques in remote sensing Image enhancement techniques in remote sensing applications			
	Transaction:	ArXiv [17] Elsevier [18]		
		Method:	Image Fusion Techniques are Hue-Saturation-Intensity (HIS), Principle component analysis (PCA) Wavelet Transform (WT), Brovey Transform (BT)	
			Advantages:	
				BT is easy to use PCA is good for merging a large number of input images and giving a compressed output image HIS is
			Disadvantages:	

(continued)

Table 2 (continued)

				BT is limited to three-band and gives less clear images PCA is not good for fusion because of merging two images then spectral information of the fused image is lost
			Result:	
				HIS technique is widely used with other WT techniques for better result
Title:	Image fusion techniques: a survey Image fusion techniques—a survey			
	Transaction:	Springer [19] IJRASET [20]		
		Method:	Spatial domain: average, min, max, weighted average, brovey, PCA, HIS, filtering Frequency domain: Laplacian/Gaussian pyramid image fusion, DWT, SWT, KHWT, DCT, Curvelet Transform Deep Learning-based Image Fusion-CNN	
			Advantages:	
				Average, max, min, weighted average is easy and efficient and take less computational time **Brovey** is an efficient and fast processing method that gives RGB images with superior contrast **HIS** is a simple and fast processing method that gives high sharpening **PCA** is an easy and less computational time-consuming method that gives high spatial quality images **Laplacian** gives the multi-focus image in better quality **Discrete cosine wavelet** good for real-time image enhancement **Discrete Wavelet Transform** reduce spectral distortion for better quality **Stationary Wavelet Transform** and **curvelet transform** is good for real-time application **Convolutional Neural Network** is a good method for feature extraction and representation
			Disadvantages:	

(continued)

Table 2 (continued)

				Average, max, min, weighted average decrease image quality, produce blur images, not good for real-time image processing **PCA, HIS, Brovey** facing color distortion problem Laplacian affects the Image Fusion results **Discrete cosine** wavelet gives low-quality fusion **Discrete Wavelet Transform** reduce the spatial resolution **Stationary Wavelet Transform and curvelet transform** is time-consuming **CNN** computational cost is high because of the high dataset representation
			Result:	
				PCA, HIS, Bovery are good high-speed methods but reduce color distortion of images **DCT, DWT, SWT, Curvelet Transform** is good for real application but time-consuming The main challenge is the fusion of panchromatic images, hyperspectral images, Multi-Spectral images, because of the different sensors So Deep learning and machine learning algorithm are suitable in the current scenario

DWT with SWT gives high-resolution compared to DWT only. Using mapping of intensity value gives an effective image compared to the conventional method. The method is effective and overcomes the drawbacks of DWT such as lack of directional selectivity, shift variance, redundancy and also using Cycle Spinning methodology to obtain de-noised output resultant image.

These techniques were used for testing different satellite sensor images and also used fusion algorithms for result improvement. In future, these techniques can apply in video images and 3D images. Image fusion technique to enhance the Satellite images for better visualization and classification accuracy. Hybrid system techniques at pixel level give the best result.

6 Conclusion

Currently, image fusion techniques are much better and more effective. During fusion of different satellite images faces many issues, because various techniques have their pros and cons. But it concludes that image fusion techniques can apply differently are of image processing and in the future, Deep Neural Networking based Image Fusion method will improve the quality of images.

References

1. Sowmya DR, Deepa PS, Venugopal KR (2017) Remote sensing satellite image processing techniques for image classification: a comprehensive survey. Int J Comput Appl 161(11):24–37
2. Janani P, Premaladha J, Ravichandran KS (2015) Image enhancement techniques: a study. Indian J Sci Technol 8(22), ISSN (Print): 0974-6846 ISSN (Online) : 0974-5645. https://doi.org/10.17485/ijst/2015/v8i22/79318
3. Maini R, Aggarwal H (2010) A comprehensive review of image enhancement techniques. J Comput 2
4. Singh G, Mittal A (2014) Various image enhancement techniques—a critical review. Int J Innov Sci Res 10(2):267–274, ISSN 2351-8014, Innovative Space of Scientific Research Journals http://www.ijisr.issr-journals.org/
5. Kaur M, Jain KI, Lather V (2013) Study of image enhancement techniques : a review. Int J Adv Resin Comput Sci Softw Eng 3(4)
6. Ramkumar M, Karthikeyan B (2013) A survey on image enhancement methods. Int J Eng Technol (IJET) 5(2):960. ISSN: 0975-4024
7. Ahire RB, Patil VS (2013) Overview of satellite image resolution enhancement techniques. In: 2013 Tenth international conference on Wireless and Optical Communications Networks (WOCN), pp 1–4. https://doi.org/10.1109/WOCN.2013.6616261
8. Rathod M, Khanapuri J, Hiran D (2020) Wavelet transform domain methods for resolution enhancement of satellite images. https://doi.org/10.1007/978-981-15-1081-6_12
9. Sharma L, Sengupta S, Kumar B (2021) An improved technique for enhancement of satellite image. J Phys: Conf Ser 1714:012051. https://doi.org/10.1088/1742-6596/1714/1/012051
10. Ahuja S, Biday S (2018) A survey of satellite image enhancement techniques IJAIR 2(8). ISSN: 2278-7844

11. Bajpai K, Soni R (2017) Analysis of image enhancement techniques used in remote sensing satellite imagery. Int J Compu Appl 169(10):1–11. https://doi.org/10.5120/ijca2017914884
12. Jadhav BD, Patil PM (2015) An effective method for satellite image enhancement. In: International conference on computing, communication and automation, pp 1171–1175. https://doi.org/10.1109/CCAA.2015.7148553
13. Kundeti NM, Kalluri HK, Krishna SVR (2013) Image enhancement using DT-CWT based cycle spinning methodology. In: IEEE international conference on computational intelligence and computing research, pp 1–4. https://doi.org/10.1109/ICCIC.2013.6724213
14. Suganya P, Gayathri S, Mohanapriya N (2013) Survey on image enhancement techniques. Int J Comput Appl Technol Res 623–627. https://doi.org/10.7753/IJCATR0205.1019
15. Kenneth AAJ, Dhanaseelan FR (2018) An extensive survey on satellite image enhancement. In: International Conference on Smart Systems and Inventive Technology (ICSSIT), pp 425–429. https://doi.org/10.1109/ICSSIT.2018.8748631
16. Ablin R, Sulochana CH, Prabin G (2020) An investigation in satellite images based on image enhancement techniques. Eur J Remote Sens 53(sup2):86–94. https://doi.org/10.1080/22797254.2019.1673216
17. Gharbia R, Azar A, Baz A, Hassanien A (2014) Image fusion techniques in remote sensing. ArXiv, abs/1403.5473
18. Simone G, Farina A, Morabito F, Serpico S, Bruzzone L (2002) Image fusion techniques for remote sensing applications. Inf Fusion 3:3–15
19. Kaur H, Koundal D, Kadyan V (2021) Image fusion techniques: a survey. Arch Comput Methods Eng 1–23
20. Suvitha N (2018) Image fusion techniques—a survey. Int J Res Appl Sci Eng Technol 6:288–291

Heart Disease Prediction Analysis Through Ensemble Learning Model

Rohit Shrivastava, P. Vigneshwaran, and Anand Kumar Soni

Abstract Cardiovascular disease is one of the most life-threatening diseases all over the world and under distributed age groups. Mortality rates of heart disease around the world are high compared to any other disease. Thus heart disease prediction gains importance and improvement in prediction is needed to save human lives. Once the disease is early predicted, with proper medications, the human lives can be saved. The delay in detection leaves complexity and the rates of mortality increase. Medical diagnosis industry is gaining pace with the latest technologies such as machine learning, deep learning. Medical treatment industry is already developed with highly effective treatments. Machine learning (ML) is one of the best diagnoses for effective decision-making. The proposed system aims to study cardiovascular disease prediction on UCI repository, Cleveland dataset with a novel ensemble learning strategy. The algorithms used are Decision Tree classifier and Random forest classifier for learning. The proposed ensemble model exploits the use of both models to create novel learning methods. The experimental study proved that the proposed model is a highly optimized model of existing machine learning classification algorithms.

Keywords Heart disease prediction · Ensemble model · Cardiovascular disease · Deep learning · Random forest · Machine learning · Disease prediction

1 Introduction

Data mining is a vast domain, which handles large amounts of data for studying and analysis, the study and deep analysis which in turn is useful for making decisions to

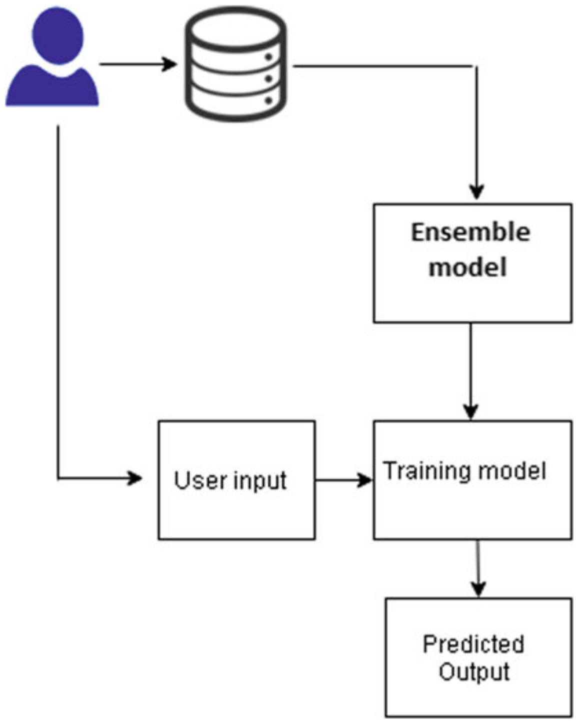
R. Shrivastava (✉) · P. Vigneshwaran · A. K. Soni
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpat, Chennai 603203, Tamil Nadu, India
e-mail: rr5921@srmist.edu.in

P. Vigneshwaran
e-mail: India.vigenesp@srmist.edu.in

A. K. Soni
e-mail: av1555@srmist.edu.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_37

Fig. 1 Overview of heart disease prediction



assist the real-world applications. The most important concepts mostly used under data mining are clustering classifications and regressions.

Machine learning is one of the highly demanded concepts in medical diagnosis, the range of usage includes heart disease prediction, liver disease predictions, tumor cell identifications from MRI scans, and X-ray image analysis and diagnosis (Fig. 1).

The ensemble model uses the novel machine learning technique for classification. The proposed work is approached in two phases. The first phase is to implement existing algorithms such as Decision tree and Random forest and analyze their accuracy metrics. The second phase of work is to propose an ensemble model, which combines Decision tree Algorithm and Random forest model for bringing a new learning method, exploiting the benefits of both algorithms.

The upcoming sections include literature review and work related to heart disease prediction. In Sect. 3, proposed algorithms and ensemble models are elaborated. In Sect. 4, results and evaluations are elaborated. In Sect. 5, this work is Conclusion and further enhancements are elaborated.

2 Related Work

There are several researchers studying heart disease detection through various techniques and some of them, which are more relevant to our work, are addressed below.

Diagnosis of heart disease risk level is addressed in [1] with firefly algorithm for feature selection with convolutional neural networks (CNN) for learning and prediction. This work took an electrocardiogram (ECG) dataset for heart threat prediction. Firefly algorithm gives best-optimized features from the dataset to make learning. Depending on the light intensity and attractive property, the features are selected. The accuracy achieved through this feature optimization and deep learning model is high compared to normal deep learning models. However, the features used in ECG dataset are high, in our model, Cleveland dataset with less features is not required for any feature selection.

Clinical diagnosis as a binary classifier of presence and absence of disease is identified through rough set back propagation neural network was studied in [2]. This study used the Irvine dataset for breast cancer detection. The feature set is reduced through rough set selection, these techniques get only important features from the dataset, the selected features are fed to BPNN for further learning and prediction. The work achieved was high on accuracy.

Coronary Artery disease has been classified by hybrid model as proposed in [3]. The author used Particle swarm optimization for feature subset selection, which selected the best particle as Particle best and gbest and global best features. This gives the optimized feature subset from the Cleveland dataset. The author used a *k*-means clustering algorithm to search for the best feature set. Supervised classification such as multi-layer perceptron and Logistic regression was used for disease classification. This work addressed that logistic regression was outperformed than other learning models.

The work in [4] addressed heart disease detection through novel feature set selection. Feature selection is done through three techniques including fisher score, forward and backward feature selection algorithm. From this technique, only the most important features/attributes from four datasets such as Cleveland, Hungarian, Switzerland, and SPECTF were analyzed. For supervised learning, a kernel-based Support Vector machine algorithm is used with this Switzerland dataset to achieve high accuracy.

The work in [5] addressed the problem of cardiovascular disease prediction through feature selection and learning. For feature selection, the author used factor analysis of mixed data (FAMD) to get important attributes/features from the Cleveland dataset. There are five classifications applied to the extracted features from FAMD. The dataset is normalization done through zero mean and Unit standard deviation then applied to the algorithm. The author proved that Random forest (RF) achieved high accuracy.

There are few points inferred from the existing study, there are many authors who applied feature selection and machine learning for heart disease prediction. These feature subset selection techniques have optimized the results and also improved accuracy, however, the Cleveland dataset has less number of features, not required heavy feature selection techniques. Some of the work addressed was implemented in deep learning algorithms. This motivated us to bring an optimized model based on combining two different supervised learning algorithms. Thus the novel ensemble learning strategy is proposed for learning and detection of heart disease as binary classification.

3 Proposed Work

We used individual decision tree algorithms and random forest algorithms as our preprocessing algorithm before actually getting into the proposed system of combining both of these in the proposed system as the hybrid model.

Heart disease prediction implementation is proposed with Python 3 or anaconda 3 with some important libraries namely scikit-learn, pandas for reading CSV files, matplotlib for plotting graphs. The dataset used is Cleveland Heart disease data from uci.edu. Binary classification is considered for study, thus values such as 0 and 1 are considered for prediction. Machine learning models are generated with decision tree algorithms and random forest algorithms and new ensemble models.

Dataset Details

Cleveland dataset has high range of important attributes such as age of patient, sex, cp, trestbps, cholestrol value, fbs, restecg, thalach, exang, oldpeak, slop, ca, thal, pred_attribute. The sample of collected values is given below as sample (Fig. 2).

The dataset variable names are described below (Table 1).

Data Pre-Processing

The first module considered is data pre-processing, under which pre-process data visualization is handled, this gives ideas to the data analysts for high-level view and understanding the nature of data in a visual way. As the dataset taken has no missing values, thus formatting or cleaning and other preprocessing functions are not mandatory. Thus visualization in different views are handled, heart disease and normal cases are plotted with age-wise in below graph (Fig. 3).

A	B	C	D	E	F	G	H	I	J	K	L	M	N
age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slop	ca	thal	pred_attribute
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	2

Fig 2 Cleveland data view

Table 1 Data set variable description

Variable name	Attribute description
Age	Patient's age
Sex	Patient's sex as 0 or 1
CP	Patient's pain level
Trestbps	Blood pressure value
Chol	Blood cholesterol value
Fbs	Sugar level 120 mg/dl (1 true)
Restecg	Electric. result (1 anomaly)
Thalach	Max heart rate
Exang	Induced angina (1 yes)
Oldpeak	ST values depression
Slope	Peak exercise
Ca	Number of vessels
Thal	Thalassemia
Num	Threat for heart disease (1 or 0)

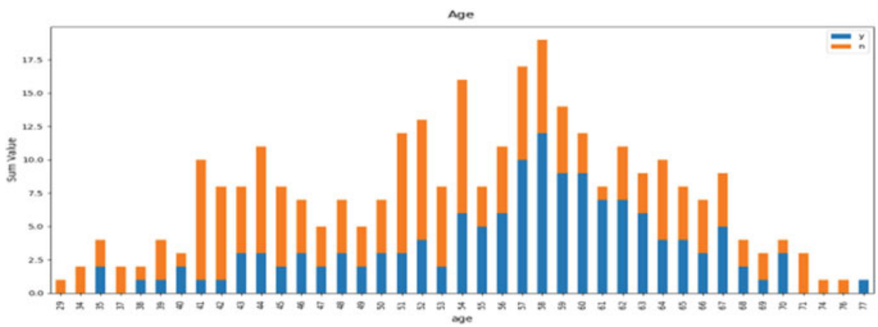


Fig. 3 Heart disease data visualization of c level and data—age-wise

Generally, males have a higher heart disease rate than females as represented below. The number of samples shows a high heart disease threat in males and in female samples, it is less, however (Fig. 4).

Data Splitting

With the Cleveland dataset, data splitting for training the ML and testing the trained ML is done. Around 70% of the whole dataset is considered for training the algo and the 30% of dataset is considered for testing.

The overall architecture of proposed work is shown in Fig. 5, this represents user and dataset input and learning models as mentioned.

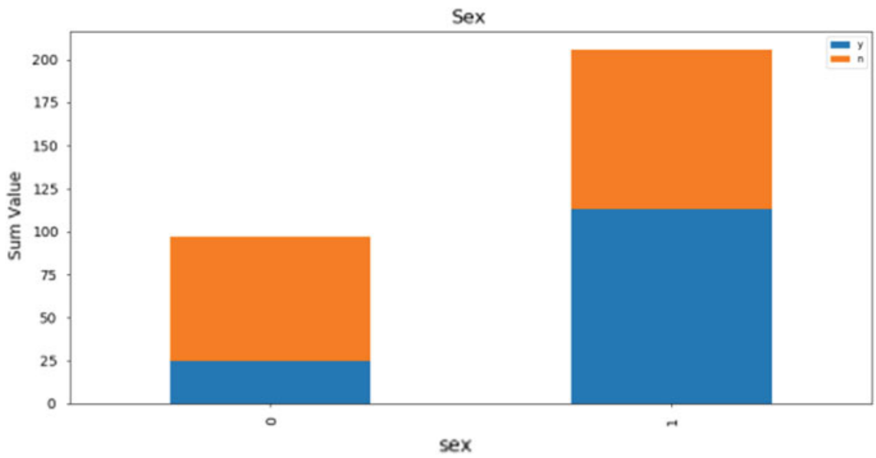


Fig. 4 Heart disease rate visualization of c level and data—sex wise

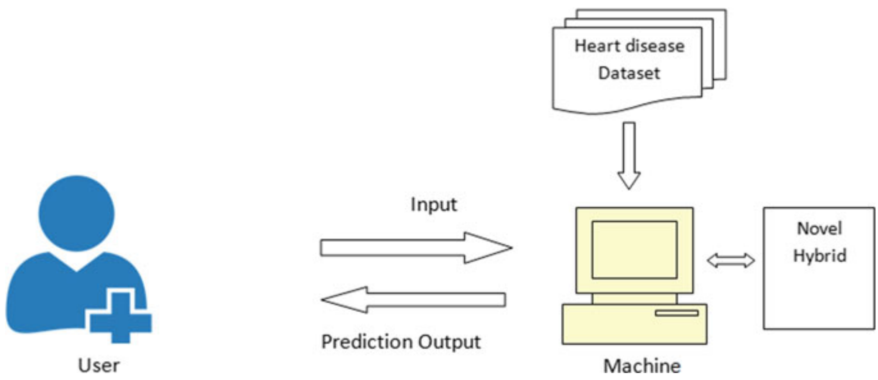


Fig. 5 Architecture of heart disease prediction

Supervised classification algorithms such as Decision Tree (DT), Random Forest (RF) are used for disease prediction. The novel ensemble model based on these two algorithms is proposed as an optimized learning model.

Supervised learning algorithms are discussed in detail with ensemble model below.

Decision Tree Classifier

Decision Tree classifier is one of the effective data structures, which can solve many computation problems. The learning through Decision trees can be done through binary as well as continuous values. The implementation involves splitting the dataset as train and test. The tree generated has a special root node and from which branches are generated. Leaf nodes are generated from branches, this represents the result, in this path the labels are never repeated thus the leaves are decisions made after consequent processing.

Decision Tree has following steps to follow

- Best feature is considered as the root node of a tree.
- Internal nodes are attributes, whereas edges are values of those attributes. External nodes are output of the analysis.
- These steps repeated to get all attributes are allotted in the tree.

The class value to predict for an instance in the decision tree starts only from the root node. Algorithm is classified based on questions starting from root node to end at each terminal node. The compared values with the root attributes are marked as terminal attributes. This comparison arrives at the respective value for the root node.

Random Forest Regression

Random forest algorithm is a supervised machine learning algorithm, which can perform classification and regression. Starts with random samples from Cleveland data in the sampling nodes. The dataset is given as training and testing data along with X, Y attributes, where X refers to all attributes in data and Y refers to Target of prediction.

The samples for random forest trees built by bootstrap, voting process is carried out for every prediction. The same feature is considered many times to vote and get results. The maximum number of features on node splitting could be limited by number. This algorithm reduces overfitting problems.

Hybrid Model

The ensemble model is the combination of supervised classification algorithm Decision Tree Classifier and Random Forest Classifier. The prediction probability of random forest is arrived at and these probabilities are inserted into the train_X data. This probability including train x is given as input for the decision tree to predict the heart disease value. Similarly, the test data probabilities were also predicted through random forest and inserted to test_x data then fed to the decision tree algorithm. This ensemble model gives the optimized results.

Supervised machine learning classification algorithm is taken for this study, the dataset is supervised data with binary values for classification as '0' represents normal and '1' represents abnormal. Figure 6 shows the application home page of heart disease threat prediction. Wherein patients can enter their values and predict the heart disease is presence or normal. This application is designed in Python TKinter as a windows application.

4 Results and Discussions

The first phase is data pre-process in which data visualization is handled. The second phase is implementing the Decision tree algorithm, an arrival evaluation metric for

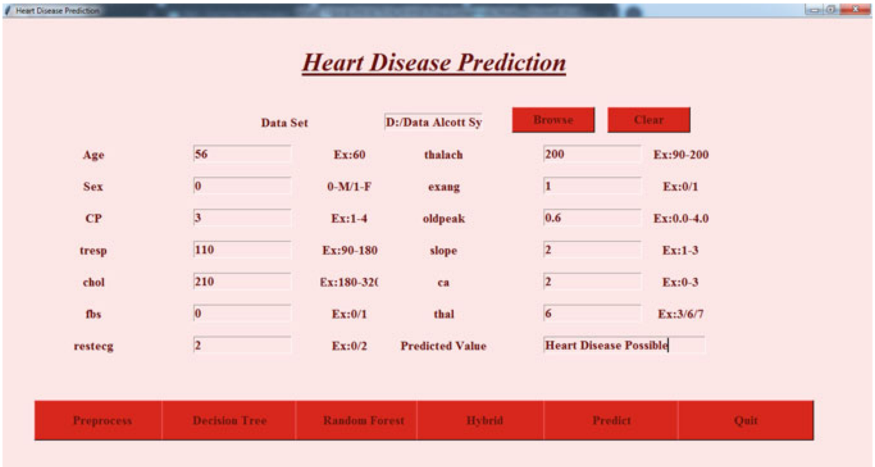


Fig. 6 Application for heart disease prediction

this. The third phase is applying a random forest algorithm, in which this prediction is done and evaluation metrics arrived. The fourth phase is making an ensemble model based on random forest prediction probabilities to decide tree algorithms. Random forest gets 76% accuracy, Decision Tree gets 79% accuracy, Ensemble model gets 76% accuracy. In sensitive field like prediction of Heart Disease, Hybrid algorithm did not just get higher accuracy of around 76% but also gives less errors of R-squared, mean square mean absolute errors, etc. which also contributed to development of efficiency of a model as an overall perspective however existing systems have their own limitations like in knn algorithm accuracy depends on quality of data and is also very slow in prediction with large dataset which is not the case in Random forest algorithm which can handle large dataset with higher dimensionality and also maintain accuracy of missing data combining it with decision tree increases its potential even more as it takes less effort for data preparation (Table 2).

Figure 7, shows Decision tree evaluation metrics with accuracy, error values namely MAE (mean absolute error), MSE (Mean square error), RMSE (Root Mean square error), and R-squared error.

Figure 8, shows Random forest evaluation metrics with accuracy, error values namely MAE, MSE, RMSE, and R-squared error.

Figure 9, shows Ensemble learning model evaluation metrics with accuracy, error values such as MAE, MSE, RMSE, and R-squared error.

Table 2 Experimental results

Algorithm	Accuracy (%)
Decision tree algorithm	79.32
Random forest	76.33
Ensemble learning (decision tree random forest)	76.98

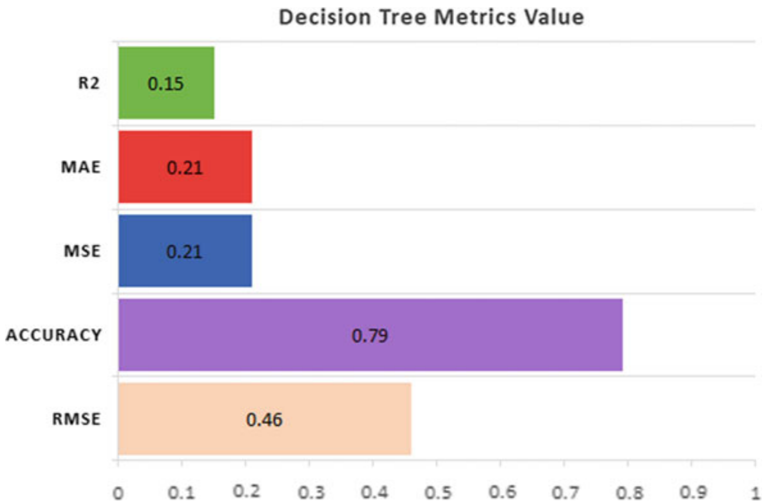


Fig. 7 Evaluation metric using decision tree algorithm

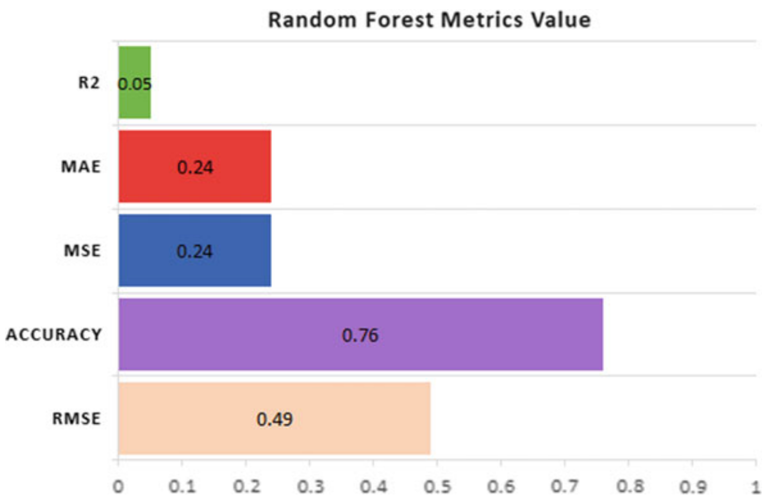


Fig. 8 Evaluation results of random forest

5 Conclusions

Cardiovascular disease is one of the most life-threatening diseases in the age group all over the world. The change in modern lifestyle, lack of physical exercise, usage of alcohol and tobacco, and junk food are the major causes of disease. The medical industry is equipped with highly modernized treatments. However, the medical diagnosis industry needs much efficient technology for automated detections. Artificial

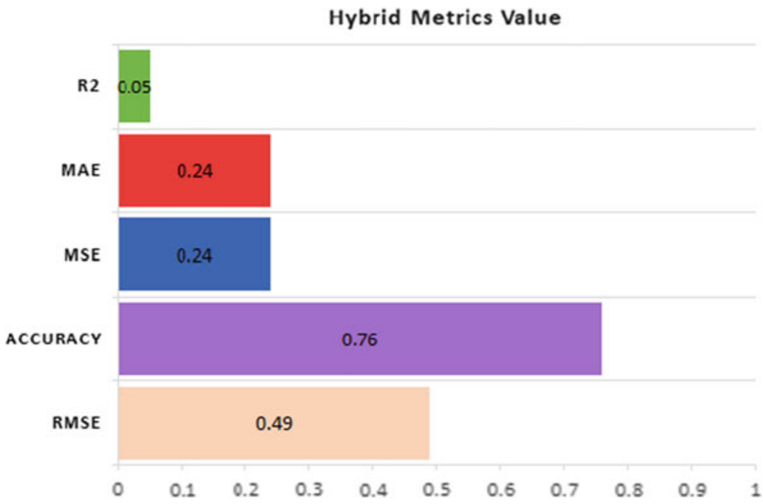


Fig. 9 Results of heart disease prediction using ensemble learning

intelligence is fulfilling the demands of the diagnosis industry. This required a lot of research enhancement to make the process more accurate and reliable. The proposed work detects the threat of cardiovascular disease through machine learning models. The proposed ensemble learning optimized the results of decision tree and random forest.

References

1. Khan Mamun MMR, Alouani AT (2020) FA-1D-CNN implementation to improve diagnosis of heart disease risk level. https://doi.org/10.11159/icbes20.122
2. Singh, Kumar R (2020) Heart disease prediction using machine learning algorithms. In: 2020 international conference on electrical and electronics engineering (ICE3), pp 452–457. https://doi.org/10.1109/ICE348803.2020.9122958
3. Sharma V, Yadav S, Gupta M (2020) Heart disease prediction using machine learning techniques. In: 2020 2nd international conference on advances in computing, communication control and networking (ICACCCT), pp 177–181. https://doi.org/10.1109/ICACCCN51052.2020.9362842
4. Shah SMS, Sher M, Shah FA, Khan I, Ashraf MU, Awais M, Ghani A (2019) Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. Knowl Inf Syst 58(1):139–167. https://doi.org/10.1007/s10115-018-1185-y
5. Devi R, Tyagi HK, Kumar D (2017) Early stage prediction of sudden cardiac death. In: 2017 international conference on wireless communications, signal processing and networking (WiSPNET), pp 2005–2008. https://doi.org/10.1109/WiSPNET.2017.8300112

Diagnosis of Visible Diseases Using CNNs

R. Sandeep, K. P. Vishal, M. S. Shamanth, and K. Chethan

Abstract The recent surge in the deep learning methods for medical image analysis has assisted the progress of intelligent medical imaging-based diagnosis systems that can provide support for human disease detection. In this paper, a low complex Convolutional Neural Network (CNN) based solution is implemented for the skin lesion classification of eight diseases by analyzing characteristic features to distinguish various diseases. Here, the tensor factors are considered for the following disease classification namely Psoriasis, Chicken Pox, Vitiligo, Melanoma, Ringworm, Acne, Lupus, and Herpes. The effectiveness of the solution is compared with VGGNet based solution which classifies only malignant or benign skin cancer. The proposed solution will give classification of skin lesions for eight diseases and provides the classification four times faster when compared to VGGNet. The VGGNet has an accuracy of 71%, whereas the proposed solution will give an accuracy of up to 78%.

Keywords Tensors · Deep learning · Convolutional neural networks

1 Introduction

Our brains make vision seem easy. It is very easy for humans to distinguish between a wolf and a dog, identify the person's emotion from his/her face, and many such complex tasks. But these tasks are challenging for a computer [1]. In the recent years, the field of machine learning and deep learning has made tremendous progress in addressing these onerous problems. There is a model called a deep convolutional neural network (CNN) which can achieve reasonable performance on hard visual recognition tasks, matching or exceeding human performance in some domains. It is observed that without the assistance of a machine, the accuracy of clinical diagnosis of visible disease lies between 65 and 80%. Use of dermoscopic images (shown in Fig. 1) improves diagnostic accuracy of the skin lesions by 49%. The use of artificial intelligence along with human intelligence will improve the accuracy of identifying

R. Sandeep (✉) · K. P. Vishal · M. S. Shamanth · K. Chethan
Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysuru 570002, India
e-mail: sandeep.ece@vvce.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_38

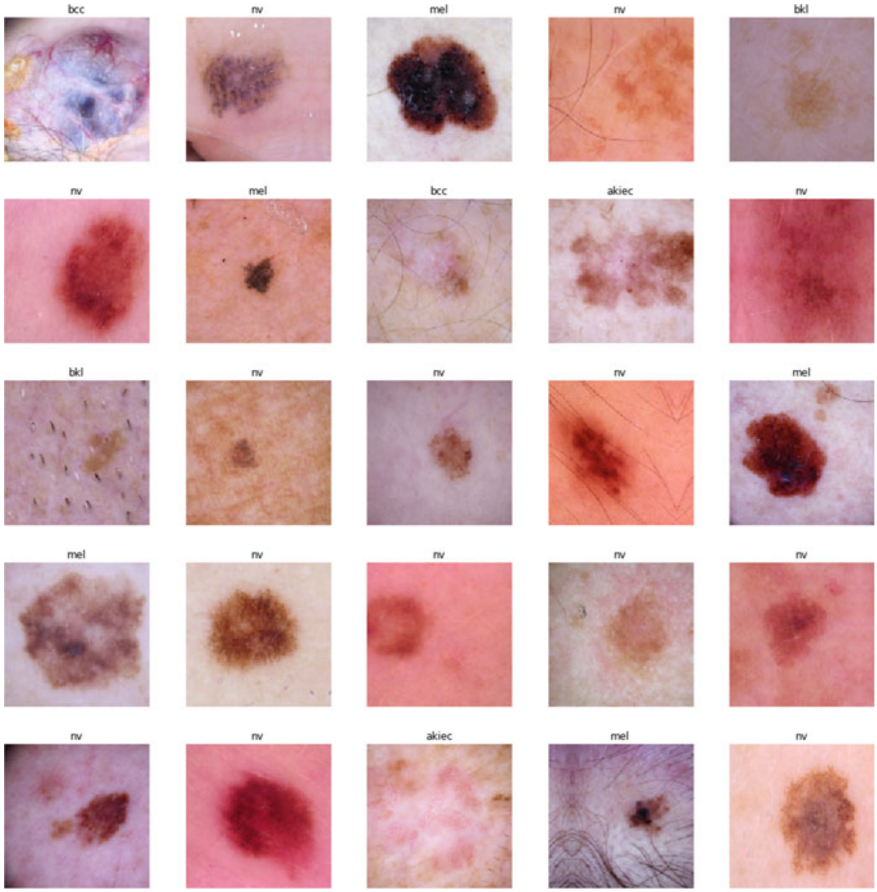


Fig. 1 Random images from the HAM10000 dataset

skin diseases. As seen in Fig. 2, the Inception_v3/v4 has very good trade-off between the accuracy and operations (complexity) when compared to any other CNNs. Hence, we take the Inception_v3 as a base network. We customize and train it for various skin lesions classification. The proposed solution gives 4× less complex solution and provides 6% better accuracy when compared to the VGGNet based solution [2] proposed for identifying only the melanoma and benign diseases.

Shanthi et al. in [3] have proposed automatic diagnosis of skin diseases such as Keratosis, Acne, Urticaria, and Eczema herpeticum using CNN. Their method yielded an accuracy between 98.6 and 99.04% over DermNet dataset. Polat and Koc in [4] have proposed the detection of various skin diseases from dermoscopic images using the combination of CNN and one-versus-all approach with an accuracy of 92.9% using HAM10000 dataset. Wu et al. in [5] have automated the diagnosis of three inflammatory skin diseases using CNN with an accuracy of 95.80% ± 0.09%

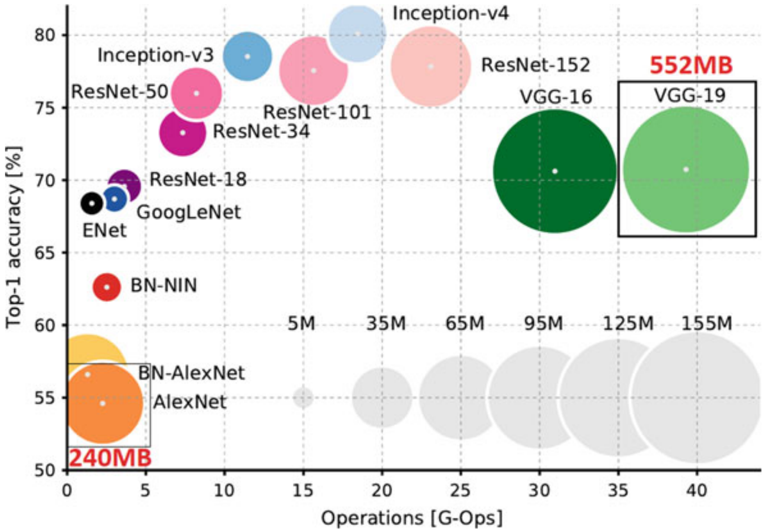


Fig. 2 Complexity versus accuracy of classification using CNNs [10]

using 4740 clinical images. Gupta et al. in [6] have developed a mobile application to analyze and identify the most commonly occurring dermatological diseases using Gaussian mixture model with satisfactory efficiency. Velasco et al. in [7] have proposed smartphone-based skin disease classification using MobileNet CNN with accuracy ranging between 84 and 94%. Zhang et al. was able to achieve an accuracy of 87.25% ± 2.24% in classifying four common skin diseases from dermoscopic images using the combination of deep neural network and human knowledge [8]. Rathod et al. in [9] have proposed the diagnosis of dermatology diseases using CNN and obtained an accuracy close to 70%.

2 Motivation and Problem Definition

One of the major challenges faced by the health ministry of India is the rural health care because more than 70% population of the total population are still living in the rural areas and the health care facilities are still poor. Due to this, the mortality rates are high in the rural areas. According to the National Rural Health Mission Report:

1. There are 6,36,00 villages in India and 70 crore people live in these villages.
2. The reasons for the death of the people are found to be curable and preventable diseases.
3. 66% of the population living in the rural areas have no access to the critical and life-saving medicines.

4. 31% of the population living in the rural areas must travel for more than 30 km to obtain health care.
5. Health centers present in the rural part of India have acute shortage of trained medical professionals.
6. 8% of the primary health centers do not have the doctors.

People living in the rural areas are used to household cures for their ailments and are very unwilling to take medicines. They seek the help of the doctor as a last resort after trying out all the domestic cures suggested by friends, neighbors, and relatives. This leads to complications in the health condition and a possible death. The people are unaware that the awareness and the education about the diseases could solve majority of the problems. Quality of health care services, awareness, accessibility, and affordability are the key concerns of rural health care.

3 Datasets and Challenges

The datasets in the field of dermatology and skin lesion images are either too less and/or not available publicly. This makes the research challenges in these areas. The examples of dermatology related image datasets used in the literature include:

- *Dermnet* is a skin disease atlas with website support that contains over 23,000 skin images separated into 23 classes [11].
- *Dermofit* Image Library is a dataset that contains 1,300 high-quality skin lesion images collected across 10 different classes [12].
- At the beginning of 2017, the international symposium on biomedical imaging (ISBI) released a challenging dataset for skin lesion analysis toward melanoma detection.
- "The HAM10000 dataset is a large collection of multi-sources dermoscopic images of common pigmented skin lesions" [13].

4 Proposed Methodology

The researchers always use ImageNet to showcase their work related to computer vision. The following successive models also display an outstanding performance: QuocNet, AlexNet, Inception (GoogLeNet), BN-Inception_v2, and MobileNet. In this work, we use MobileNet and Inception_v3. In this work, we also give the differences observed while training our network using both the Inception_v3 and MobileNet algorithms.

MobileNet is a small, low-latency, low-power model parameterized to meet the resource constraints of variety of use cases. It can be built for classification, detection, embeddings, and segmentation like other popular large-scale models, such as Inception. MobileNet provides trade-off between latency, size, and accuracy while comparing favorably with popular models from the literature. MobileNet can run with two software namely TensorFlow Mobile and TensorFlow Lite on a mobile device, but TensorFlow Mobile is recommended for better efficiency.

Inception_v3 is trained for the ImageNet Large Visual Recognition Challenge using the data from 2012. This is a standard task in computer vision, where models try to classify entire images into 1000 classes, like “Tiger”, “Dog”, and “Water bottle”. Using this technology (Inception_v3) we form a dataset of the list of the various diseases. We use the following method to form a dataset of the diseases.

4.1 Collection of Dataset

The images are obtained from various sources for each of the following eight diseases: Psoriasis, Chicken Pox, Vitiligo, Melanoma, Ringworm, Acne, Lupus, and Herpes. The dataset comprises 2600 images of about 250–700 images for each disease. They are grouped separately according to their diseases, respectively in the folders named after the diseases.

4.2 Creation of Bottlenecks

The bottlenecks are created by executing `def create_bottleneck_file` code in the python script. The bottleneck is a text file created for every image in which the tensor values are generated to each pixel of the image which then passes through many layers such as convolutional layer, pooling layer, fully connected layer, and normalization layer, thus obtaining the tensor values with more precision of around 100 per image which was in turn reduced from 299×299 tensors per image. These values contain information of the image in numerical format in a text file.

4.3 Training the Network

The network is then trained, where the bottlenecks created in previous step are iteratively learned by the system using two different algorithms namely Inception_v3 and MobileNet. This learning process in default takes place for 4000 steps but it is the choice of the developer of the network to choose the number of steps for learning. Inception_v3 takes approximately twice the time taken by MobileNet to train the network. This trained network is stored in the system as a graphical .pb file (protocol buffers file).

4.4 Creating the Mobile Application

Once the .pb file is generated, this model is then optimized and it is compressed to reduce the size of the application. Next, the Mobile application is built using Android Studio software, where the gradle is built using the optimized and compressed .pb file. This can be installed in the smartphone by activating the developer mode and USB debugging in the smartphone.

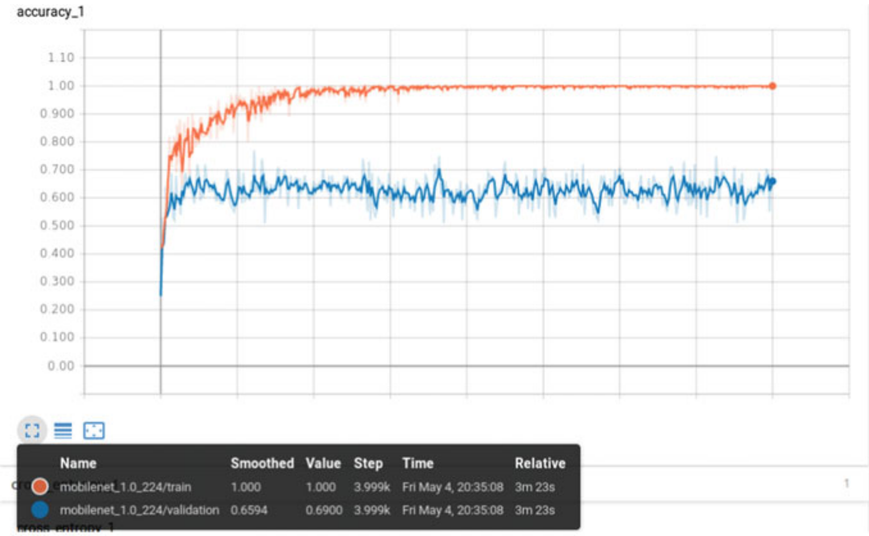
4.5 Comparison Between Inception_v3 and MobileNet

The model trained using Inception_v3 gives an accuracy of 78% whereas the model trained using MobileNet gives an accuracy of 71%. The time taken to train the network is twice in Inception_v3 compared to the training time of MobileNet based network. The size of the .pb file generated using MobileNet algorithm is half the size of the .pb file generated by the Inception_v3 algorithm.

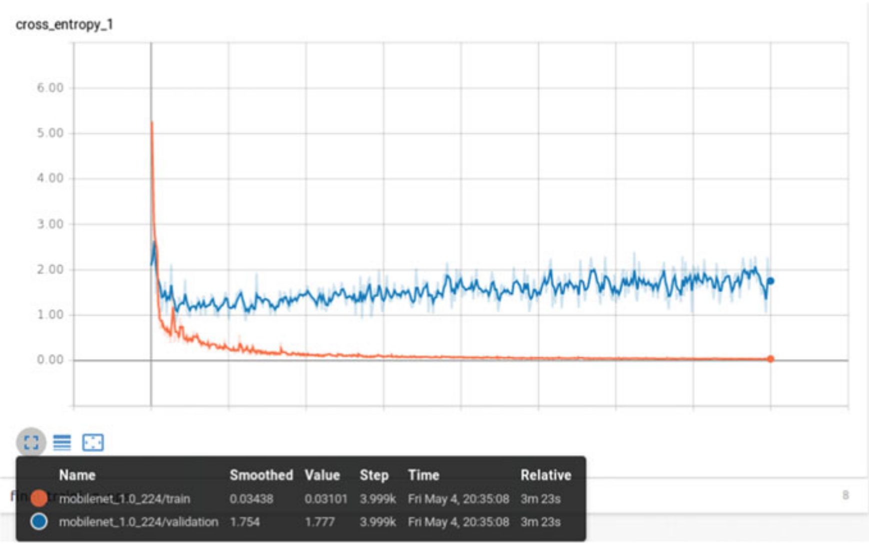
5 Simulation Results and Discussion

The types of software used to perform experiments are TensorFlow, Android Studio, and Ubuntu. We have trained our network using MobileNet and Inception_v3 models for 500, 1000, 2000, 4000, and 8000 steps. In the accuracy graph, we have number of steps on x-axis and accuracy measure on y-axis, respectively. Cross-entropy is the granular measure of computed error. In the cross-entropy graph, we have number of steps on x-axis and cross-entropy measures on y-axis, respectively. Figure 3a, b show the accuracy and the cross-entropy obtained by training the network in 4000 steps using MobileNet algorithm, respectively. The network is trained in 3 min and 23 s. The accuracy is around 69% and the error is around 1.8. Figure 4a, b show the accuracy and the cross-entropy obtained by training the network in 4000 steps using Inception_v3 algorithm, respectively. The network is trained in 5 min and 45 s. The accuracy is around 69% and the error is around 1.0.

Table 1 shows the accuracy and the cross-entropy obtained by training and validating the network in different number of steps using MobileNet and Inception_v3 algorithms. The network training time required for different number of steps in MobileNet and Inception_v3 algorithms is also tabulated. The method is superior if it has highest accuracy and least cross-entropy. In our work, the highest accuracy and the least cross-entropy was obtained in the model trained using Inception_v3 algorithm for 4000 steps. For a smaller number of steps, since the training was insufficient it yielded lesser accuracy value and higher value of cross-entropy. Similar kind of behavior was observed when the model was trained for more than 4000 steps because of the over-training.

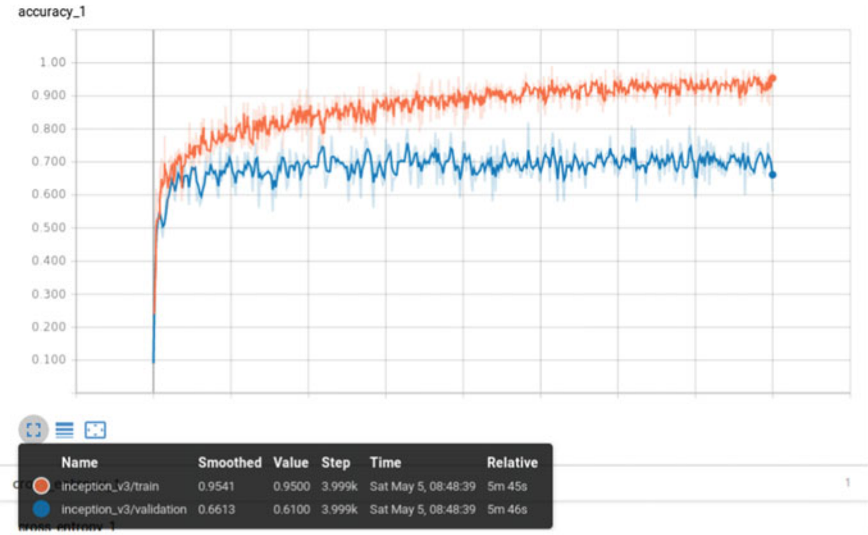


(a) Accuracy for 4000 steps using MobileNet.

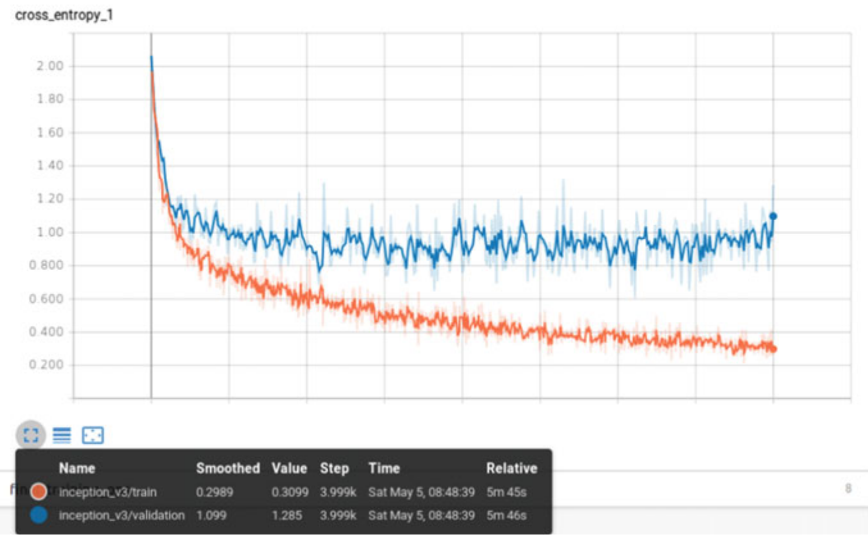


(b) Cross Entropy for 4000 steps using MobileNet.

Fig. 3 Accuracy and cross-entropy obtained by training and validating the network in 4000 steps using MobileNet algorithm



(a) Accuracy for 4000 steps using Inception_v3.



(b) Cross Entropy for 4000 steps using Inception_v3.

Fig. 4 Accuracy and cross-entropy obtained by training and validating the network in 4000 steps using Inception_v3 algorithm

Table 1 Accuracy and cross-entropy (error) obtained by training and validating the network in different number of steps using MobileNet and Inception_v3 algorithms. The network training time is also tabulated

Number of steps	MobileNet algorithm			Inception_v3 algorithm		
	Accuracy (%)	Cross entropy	Training time (s)	Accuracy (%)	Cross entropy	Training time (s)
1000	61	1.25	49	68	0.95	86
2000	64	1.6	102	68	1	173
4000	69	1.8	203	69	1	345
8000	64	2	396	68	1.1	694

6 Conclusion

A solution is proposed for providing the necessary medical attention to the people in rural areas and remote places where the access to doctors is not possible, by creating a smartphone application which can diagnose the visible diseases by comparing the picture taken by the patient of his/her lesion with the database. The characteristic factors are analyzed to distinguish between various diseases by considering the Tensors of the image dataset. A detailed distinguishing characteristic is provided between two different architectures (MobileNet and Inception_v3) which are used to solve this problem.

References

1. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. CoRR abs/1512.03385. http://arxiv.org/abs/1512.03385
2. Lopez AR, I Nieto XG, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. In: 2017 13th IASTED international conference on biomedical engineering (BioMed), pp 49–54. https://doi.org/10.2316/P.2017.852-053
3. Shanthi T, Sabeenian R, Anand R (2020) Automatic diagnosis of skin diseases using convolution neural network. Microprocess Microsyst 76:103074
4. Polat K, Koc KO (2020) Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all. J Artif Intell Syst 2(1):80–97
5. Wu H, Yin H, Chen H, Sun M, Liu X, Yu Y, Tang Y, Long H, Zhang B, Zhang J, et al (2020) A deep learning, image-based approach for automated diagnosis for inflammatory skin diseases. Ann Transl Med 8(9)
6. Gupta C, Gondhi NK, Lehana PK (2019) Analysis and identification of dermatological diseases using Gaussian mixture modeling. IEEE Access 7:99407–99427
7. Velasco J, Pascion C, Alberio JW, Apuang J, Cruz JS, Gomez MA, Molina Jr B, Tuala L, Thio-ac A, Jorda Jr R (2019) A smartphone-based skin disease classification using mobilenet cnn. arXiv preprint: arXiv:1911.07929
8. Zhang X, Wang S, Liu J, Tao C (2018) Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. BMC Med Inform Decis Mak 18(2):69–76

9. Rathod J, Waghmode V, Sodha A, Bhavathankar P (2018) Diagnosis of skin diseases using convolutional neural networks. In: 2018 second international conference on electronics, communication and aerospace technology (ICECA), pp 1048–1051. IEEE
10. Canziani A, Paszke A, Culurciello E (2016) An analysis of deep neural network models for practical applications. CoRR abs/1605.07678. http://arxiv.org//abs/1605.07678
11. Dermnet is the world's free resource and authority on all things skin. https://dermnetnz.org/image-library/
12. Dermofit: A cognitive prosthesis to aid focal skin lesion diagnosis. https://homepages.inf.ed.ac.uk/rbf/DERMOFIT/
13. Tschandl P, Rosendahl C, Kittler H (2018) The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 5(1). https://doi.org/10.1038/sdata.2018.161

Performance Analysis for Accuracy of Various Algorithms to Detect and Classify Brain Tumor Using Data Mining

Smita Deshmukh and Divya Tiwari

Abstract Brain Tumor is referred to as irregular intracranial growing of cells which happens because of abnormalities occurring in body. Detecting the tumorous region manually from MRI, for capturing its exact location and regions is tedious task due to similarity in image pixels which are not distinguishable. Also, tremendous rise in tumor cases makes the manual process of examining the patient reports difficult. Once, the Tumorous region is detected, classification is the next important step which is done to understand the type of it. Detailed diagnosis of location, Region of Interest (ROI), type of tumor, etc. is essential to provide proper treatment to patients. To make the process of detection and classification, automated various work has been done in the field of science and technology. Various technologies like AI, ML, and Data Mining have been utilized for developing automated systems. There are various algorithms available that can be used to achieve this automation but have some limitations. To overcome the limitations of available algorithms many hybrid algorithms are also developed. This paper focuses on comparing the base algorithm and existing other methods with the proposed model for its accuracy as well as other parameters. The proposed model is a hybrid of CNN and LSTM for Detection and for Classification, hybrid of CNN, LSTM, and SVM is used. The experimental results achieved accuracy of 96.06% for detection and 99.34% for classification.

Keywords Brain tumor · LSTM · SVM · CNN · MRI · Segmentation · Detection · Classification

1 Introduction

Our body is a complex structure buildup of multiple cells with each cell performing specific functions. To maintain the health of our body, these cells continuously grow

S. Deshmukh · D. Tiwari (✉)
Terna Engineering College (TEC), Mumbai University, Navi Mumbai, India
e-mail: divyatiwari0301@gmail.com

S. Deshmukh
e-mail: smitadeshmukh@ternaengg.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_39

and divide in controlled and regular manner, and replace old cells with new cells simultaneously. The uncontrolled growth of cells as a result, of cells losing their capability of controlled growth or some abnormalities in the body, is called a Tumor. Such abnormalities when occurring in brain are said to be “Brain Tumor”. Identification of “Brain Tumor” is considered a tedious work because of its location as well as vast variety is known. Till now, atleast 120 various types of Brain Tumors have been identified by the researchers all over the world. At basic level, Brain Tumor is categorized into Primary or Secondary. Brain Tumor is categorized as Primary or Benign when the cells are developed in Brain, such tumor is non-cancerous tumor as it is restricted to its location of growth and does not spread to other body parts. If such cells are grown anywhere in the body but later spreads through blood in brain through a process named Metastasis then it is a Malignant or Secondary Brain Tumor, such tumor is cancerous as it keeps on spreading rapidly becoming hazardous and rapidly reducing the rate of survival of patients who are diagnosed with it. The secondary brain tumors which are raised from different body parts such as blood cells, kidney, colon, lung, etc. are most common. On completion of diagnosis, medical experts work on understanding whether the detected Tumor is within CNS or outside the CNS and then they generally categorize it as either Benign or Malignant, even though this classification is done, it is not always precise [17, 18].

As per the survey conducted by the “American Cancer Society”, it is found that the chances of developing a malignant Brain Tumor are less below 1%, which indicates that 1 in 140 men and 1 in 180 women may be suffering from it [19]. The exact reason for developing Brain Tumor is not found and neither the accurate symptoms defined, thus people may be suffering from it without realizing the danger. Also, various cases have been recorded where patient visits doctor for some brain-related issue but doctors face difficulties in examining the patient and obtaining appropriate conclusion. Based on symptoms doctor mostly prescribe patient to get MRI report done. After MRI report is provided to the doctor their job of detecting issues in brain starts. Even experienced doctors find it difficult to understand MR image because Brain Tumor is a kind of tissue-like “gray matter, white matter and cerebrospinal fluid” but the only difference is that it is unwanted and uncontrolled growth of tissue, which makes it difficult for the human eye to differentiate between normal brain tissue and tumorous tissue. The increasing number of death cases because of failure in timely detection of Brain Tumor is among the major problems faced in recent years. The manual analysis and time consumed in detecting and classifying the Brain Tumor lead to delay in the treatment causing the increased death rate. As a result, it is the need of the hour to provide automated systems to doctors for easy and early identification of Brain Tumor for accurate results and classify the type accordingly. The growing demand, as well as success of automation in almost every field to ease the work and reduce the need of human intervention, has led to development of autonomous systems in every field. Likewise, various automated Brain Tumor Detection and Classification models are too developed, but none of them are used commercially only because they have some limitations which need to be eliminated.

To obtain the internal images of Brain for locating abnormalities, MRI and CT scans are most widely used. MRI is more precise and accurate as compared to CT because of its ability to produce clearer scanned images with more distinguishable features between ROI and entire image [19]. In automated process, MRI scans are provided as image slices as an input for further processes, these include "T1-Weighted, T2-Weighted, Fluid-Attenuated Inversion Recovery-(FLAIR) Weighted MRI". These are high-quality images that are used for image processing. Image processing is one of the technologies developed for obtaining the important features or ROI from images. These features are further used by model for training and learning the important features essential for performing the task of detecting and classifying the tumor based on design of model, thus making the process automated. The trained models are stored in form of weights containing important features and parameters which are used by model to perform operations on unseen or raw images and provide the processed image containing desired output. Using the concepts of Science and Technology to provide solution or a contribution in the field of Medical Science is the main objective of this work. This is proceeded by understanding the working of various existing as well as hybrid algorithms along with the existing models of domain, integrating the concepts based on feasibility and proposing a hybrid model so as to learn its efficiency rate, reliability percent, and limitations to overcome. The proposed model is then compared with base algorithm and other existing models to prove its efficiency and exhibit authenticity of the work done. This comparison and outcome of the proposed model can be used as one of the benchmarks by researchers, thereby helping them in achieving the aim of increasing automation in the Medical domain.

2 Related Works

In last decade, numerous methods is developed for creating an autonomous systems which are capable of Detecting and Classifying Brain Tumor from MRI Brain scans. Precisely extracting the observed abnormality in MRI Reports and accordingly classifying the type identified, is the most important step to be performed for conducting the clinical research work of diagnosis. Development of autonomous applications requires fulfilling the requirement of robust, accurate, reliable, and adaptive techniques. Here overview for understanding performance of various developed automated systems is presented.

"Salma and Rajya (2020)" proposed a method that combines features of optimal wavelet statistical texture with RNN for detecting Brain Tumors efficiently. They first performed the task of removing noisy data and then extracted texture features from the data. They used OGSA to minimize features and then for the purpose of performing classification tasks, the resultant is provided as an input to RNN. In the last step, the abnormal images are processed for segmentation of ROI using MRG. The experimental results achieved max accuracy of 95% for detection and 96.26%

for classification. Determining Tumor type, thickness and size are some of the works which can be included in future [2].

"Jia and Chen (2019)" proposed FAHS-SVM [3] method for segmenting the Tumorous regions of Brain using deep learning techniques. They proposed the concept of separating whole cerebral venous system into MR imagining. They implemented Probabilistic Neural Network for training and validation purposes. The model attained an overall efficiency of 98.51% for identifying the abnormal tissue from the normal tissue of images. They conclude that their model is efficient for primary diagnostic along with support of clinical decision systems [3].

"Pandiselvi and Maheswaran (2019)" proposed a distinctive technique namely "Adaptive Convex Region Contour" [4] (ACRC) for the purpose of identifying Tumor with its actual shape and size. It is followed by (SVM) for classifying the case as "normal" or "abnormal" [4]. Once abnormal case is found it is further processed for segmentation of images. For estimating the original volume of Tumor Rapid Mode Image Matching (RMIM) algorithm is used. The model achieved 36.563 MSE and 19.69 PSNR. The model achieved improved results as compared to considered methods for comparison [4].

"Amin et al. (2017)" proposed an automated method using different kernels for SVM for differentiating cancerous tissues from non-cancerous tissues by considering shape, texture, and intensity of input MR images. In pre-processing step "Brain Surface Extractor [5] (BSE)" method is utilized for purpose of removing skull and other noisy data which is not required for training. To validate their model, they performed validation on three variants of Datasets namely: Harvard, RIDER, and Local [5]. They achieved an overall accuracy of 97.1% and concluded that the model is efficient in identifying tumors accurately faster than the existing methods considered by them [5].

"Nilesh et al. (2017)" investigated Berkeley Wavelet Transformation (BWT) method for segmenting Tumorous regions from MR Images. Further to improve accuracy BWT is followed by SVM classifier for extracting relevant features from the segmented images. They achieved overall accuracy of 96.51% and 0.82 dice coefficient and therefore indicated good similarity between the automatically identified tumor tissues and manually identified tumor tissues. Investigating selective schemes of classifier by integrating different classifiers and feature selection techniques is their future work [6].

"Kaisheng and Wenlian (2020)" worked on analyzing 3D images and focused on improving encoder-decoder network. For enhancing encoder part, they proposed a methodology of adding additional paths and for enhancing decoder they proposed two different up-sampling paths. Evaluation of the model was carried out on BraTS 2019 dataset. On evaluation, it is found that the model achieved DSC of 88.12% [7].

"Hanxiao et al. (2020)" proposed a method using three patterns of distributed-dense connections for enhancing the feature of reusing and propagating CNNs by developing tunnels among the key layers of network. Training of model is done pixel-to-pixel with constricted parameters so as to enhance the performance detection and segmentation. Refining of segmentation results is done by reducing samples (False-Positive). The model achieved DSC of 89.80% and is effective in providing good results at less computational cost [8].

"Muhammad et al. (2020)" proposed a four-phase technique for classifying, localizing, and segmenting tumorous regions. They implemented homomorphic wavelet filter for reducing noisy data. They further implemented NSGA for selecting important features. Evaluation of performance of model is done on BraTS dataset. The model achieved accuracy of 89.49% using Naïve Bayes and 92.50% using ELM. Implementing Quantum Computational algorithms is the future scope of the work [9].

3 Materials and Methods

Below sub-sections cover details about the dataset used and the methods used for preparing dataset suitable for model.

3.1 Dataset

Dataset is obtained by following the instructions given on the BraTS 2018 data request page. All multimodal scans of BraTS are obtained as NIfTI files (.nii.gz). NIfTI is an acronym for Neuroimaging Informatics Technology Initiatives. It is a file format that stores 3D data. The dataset obtained is of two types, i.e., (HGG and LGG) and a validation dataset. Dataset obtained has cases as follows: HGG: 210 and LGG: 75 [14–16]. Each case mentioned five/nii files as below [14]:

- Native (T1).
- Post-Contrast T1-Weighted (T1Gd).
- T2-Weighted.
- T2 (FLAIR) Fluid Attenuated Inversion Recovery.
- Segmented volumes.

Of above .nii files, "Fluid Attenuated Inversion Recovery (FLAIR)" file is used for training the developed model because there is similarity between the Flair sequence and T2-Weighted image, except that the TE and TR, (where TE refers to Time to Echo and TR refers to Repetition Time) are very long [15]. With this, the abnormal tissues become bright and normal CSF fluid is attenuated to make it dark and distinguishable. The sequence obtained is very delicate or sensitive to pathology and hence differentiating CSF from abnormal tissues becomes easy. Middle 65 slices of each MRI scan are selected so that well-defined Brain structure image is obtained for all dimensions. The total size of data used is 55,575. This is further split for training and validation into the ratio of 80:20, respectively.

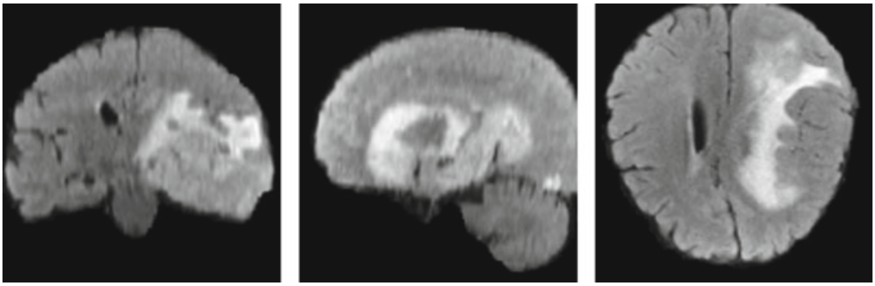


Fig. 1 All three views of brain image (front, side, top)

3.2 Data Pre-processing

Dataset cannot be used directly in the model as there are various anomalies present in the dataset which may alter the performance of the model. To reduce the computational power, the 3D volume of data is converted into 2D slices. These 2D slices are saved in .tif format of size 128 × 128. Instead of processing 3D volumes directly, the volume data of each case is converted to all three side-views (Front, Side, Top) so that when any view of brain image is provided, model is able to process and provide output (Fig. 1).

Once, the dataset is converted into required format and size, normalization process is carried out. Normalization is a process that changes the range of values to bring the values of a dataset in a specific range. For example, if values in a matrix vary with great difference, then performing operation on such matrix provides poor results. So, to obtain better results normalization is performed on the dataset. Min-Max Normalization method is used for the normalizing the dataset. In this technique, the work of transforming the data into linear fashion is carried out on original data. Min and Max value is then obtained from data and each value is substituted according to the following formula [23]:

$$A' = \frac{A - \min(A)}{\max(A) - \min(A)}$$

Here,

A'	Normalized Value.
$\min(A)$	Minimum Value.
$\max(A)$	Maximum Value.

4 Proposed Model

This work comprises of two sections: Detection and Classification. For Detection, a hybrid model is developed by combining features of CNN and LSTM algorithms. In this model feature of Long-short term memory [21] is added for improvising the performance by replacing the last layer of CNN with ConvLSTM. In Classification, a hybrid is developed by combining the features of three widely known algorithms, i.e., CNN, LSTM, and SVM. Here, ConvLSTM is added at the last convolutional layer of CNN, and loss function of SVM is used for including the features of SVM algorithm.

A. Convolutional Neural Network (CNN):

Convolutional Neural Network (CNN) [20] is buildup of neurons containing weights that are learnable as well as biases. Each and every neuron gets multiple inputs and carries "weighted sum" over them passing it through an activation function [20] and providing result. Convolutional Neural Networks consist of many convolutional layers and each layer has its activation function (Rectified Linear Unit). After Convolutional layers, there are FC layers with activation function at the output. These fully connected layers have various parameters due to which it is prone to overfitting. CNNs are basically used for the automation of models since it is capable of extracting features from datasets as well as training the model accordingly [20].

B. Long-Short Term Memory (LSTM):

Long-Short Term Memory (LSTM) [21] algorithm is a variant of RNN, with capability of learning and retaining long-term dependencies [21]. It was developed for avoiding or overcoming the problems of long-term dependency. Retaining learned information for long periods of time is the default behavior of this algorithm. All RNNs are made up of chains containing repeated modules of Neural Networks. The standard RNN consists of repeating module of a single tanh layer [21]. Similarly, LSTMs also consists of such structures but with difference in their repeating module. In this algorithm, instead of a single-layer network, there are four layers that interact with each other in special way [21].

C. Support Vector Machine (SVM):

An algorithm that has ability to classify both types of data, i.e., linear and non-linear is referred to as Support Vector Machine (SVM) [12]. In an n-dimensional space, each and every data item is mapped first, where n denotes the no. of characteristics. Identification of hyperplane is done that distinguishes the data sets into two classes [12]. It maximizes the distance from center for both the classes and minimizes errors that may occur during classification. The marginal distance or distance from center for a class is defined as the distance between the decision hyperplane and its nearest possible instance [12] which is also part of the same class. In other terms, plotting of every data into the space which is n-dimensional (where n is the no. of characteristics or features) is done first along with the value of every feature being the value of specific coordinate. For performing classification, hyperplane is found that uses maximum margin for differentiating the two classes [12].

D. Detection Model Description:

Detection model is developed using CNN as the base algorithm and U-Net architecture. Every layer of the model is connected to or linked to its next layer and parallel layer. The development of this Neural Network model is done using Keras library and TensorFlow as backend. First layer of the model has filter of size 32 with increasing value of product of 2 with previous filter value. The size of filter is increased till the 5th layer of the model and then decreased in the same manner till the last layer. The model has kernel size of 3 × 3 and max pooling layer of size 2 × 2. Dropout function is used for reducing or overcoming the overfitting problem. Every layer has activation function as ReLU while the last layer of the model has SoftMax as the activation function. Batch Normalization is set 32, i.e., the model will take decisions referring to the past 32 images that have been passed through the model. Disk Similarity Coefficient (DSC) function is used for measuring the accuracy of segmented images which shows the detected tumorous regions of the brain and loss function is negative value of disk similarity coefficient. Other parameters used for displaying accuracy of the model are Accuracy, Recall value, and Precision. Diagrammatic view of the detection model is depicted in Fig. 2.

E. Classification Model Description:

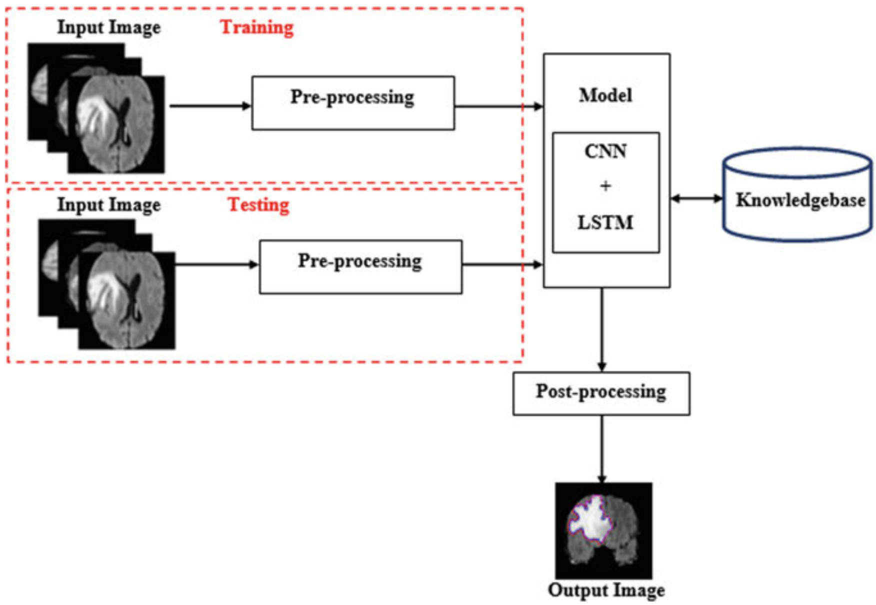


Fig. 2 Detection model

Classification model is developed using combination of CNN, LSTM, and SVM algorithms, of which CNN algorithm is used as Base algorithm. The structure of model is based on Sequential Architecture, so each layer and every layer here is connected to the other adjacent layer in sequential pattern. After performing the convolutional process, the result of a particular layer is given as an input to its next layer. The top-most or the first layer takes MR Images as an input. The size of input image is 128 × 128 and its corresponding truth value is 0 or 1 based on whether the provided image is Benign or Malignant. First layer of the model has filter of size 32 which increases at every layer as product of 2. Each layer has kernel size of 3 × 3 and padding value as same. Max pooling layer of size 2 × 2 is used and Dropout function is used at layers 3, 4, 7 and 8 to reduce the overfitting problem. Sixth layer of the model is replaced with ConvLSTM to include the features of LSTM algorithm in the model. This layer will enable the feature of forgetting and remember layer where the useful features are learned and retained by remember layer while the not learnable features are discarded by forgetting layer. The next layer is flattened layer where the entire data is converted to single-dimensional data. Further two dense layers with activation function as ReLU followed by a layer containing activation function as SoftMax, provides the final output as Benign or Malignant. To include the features of SVM algorithm, Hinge Loss is used as Loss function. To measure Accuracy of model Categorical Hinge and accuracy metrics such as Accuracy, Recall, and Precision Value is used. For optimization of algorithm base learning rate of 1e-4 is set. The trained model is saved in the form of weights which are further used in application and as an initial weight for further training of model. Figure 3 depicts the diagrammatic view of the developed classification model.

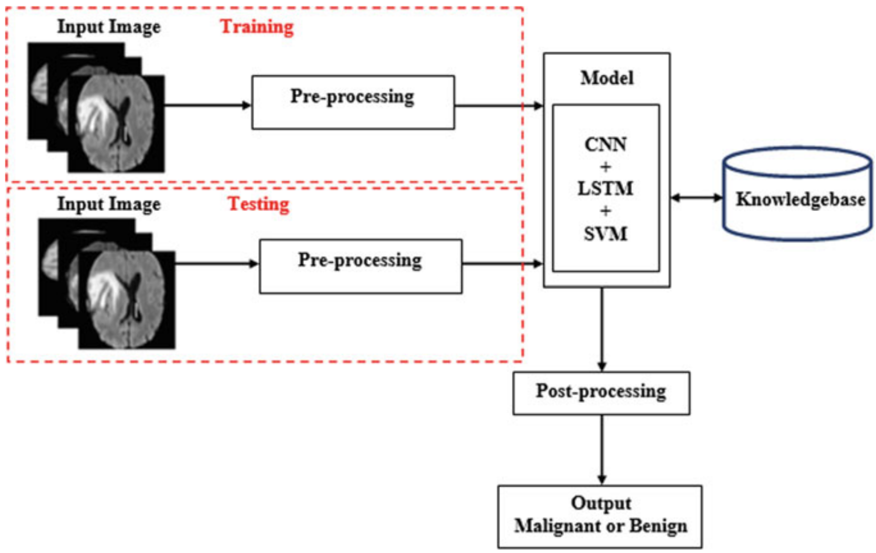


Fig. 3 Classification model

5 Experimental Result

U-Net Architecture of CNN algorithm is combined with features of LSTM algorithm for developing proposed detection model while V-Net Architecture is used in CNN Model. Here set of MR images along with its corresponding segmentation mask is used for training model. Table 1 represents the results attained by Proposed model and CNN model.

Output of model on seen and unseen data is shown in Figs. 4 and 5, respectively.

Figure 4 depicts output of model on seen data. Here the blue contour indicates the ground truth which is segmented mask that is provided while training of model. The contour in red color indicates the prediction done by model based on the characteristics or features it learned during training. In Fig. 4 Input image with its corresponding

Table 1 Detection results of proposed and CNN model

Model	Validation measures			
	Accuracy (%)	Recall (%)	Precision (%)	Dice coefficient (%)
Proposed	96.06	76.29	93.63	70.24
CNN	96.42	76.62	99.57	77.13

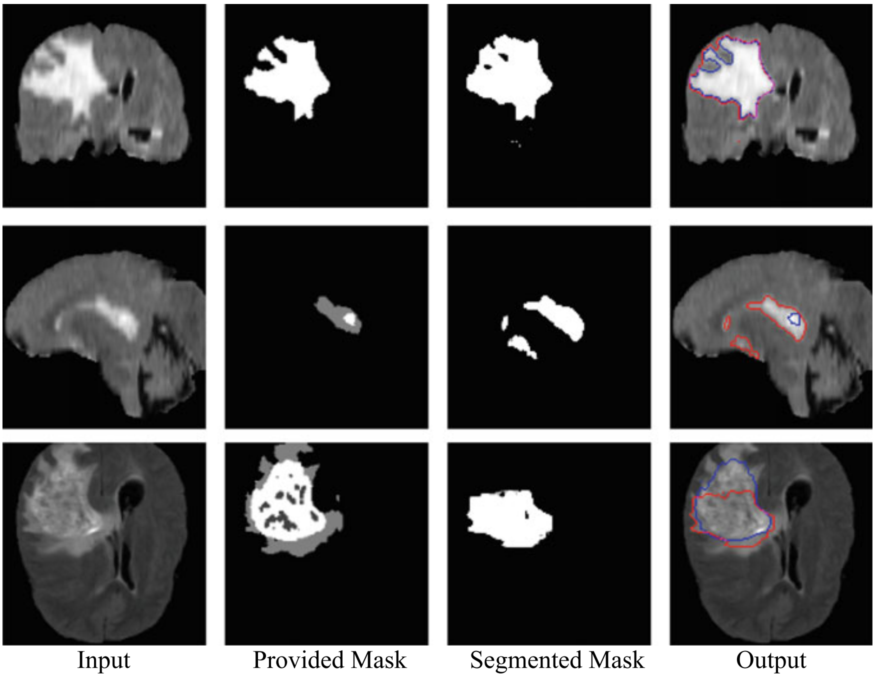


Fig. 4 Output of model on seen data

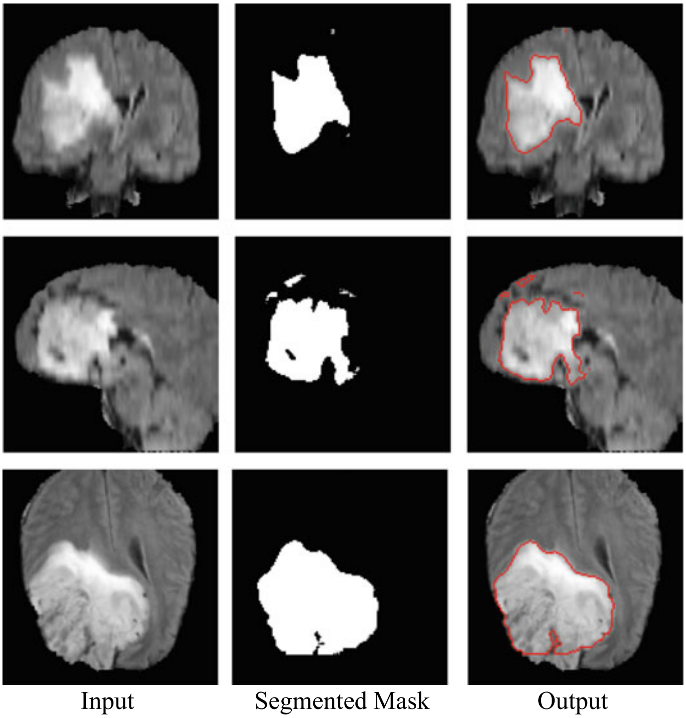


Fig. 5 Output of model on unseen data

mask which was provided while training is shown which is followed by segmentation done by model and its predicted tumorous region along with ground truth. Figure 5 depicts the output of model on unseen data. Here Input image, prediction, and segmented tumorous region predicted by model are shown.

The training of model is done for 100 epochs to achieve the above results. The Proposed Model recorded loss of 29.76% while the CNN Model recorded loss of 22.87%. Figures 6, 7, 8 and 9 illustrates graphs that were generated while training the detection model depicting the Model Accuracy and Model loss of Proposed and CNN Model, respectively.

Sequential Architecture is used as structure for developing Classification model with combination of CNN, LSTM, and SVM algorithms. Labels 0 and 1 for Benign and Malignant were used for training the model. The model is trained for 100 epochs. Table 2 depicts the results of Proposed model and CNN model.

Figures 10 and 11 depicts the classification output of model classifying Input image as Benign or Malignant, respectively.

Proposed Model recorded loss of 01.36% while the CNN Model recorded loss of 0.36%. The possibility of overfitting problem in Proposed Model is very much minimal since the Dropout function used for handling overfitting issue is used at 3, 4, 7 and 8th Layer of Model. Figures 12, 13, 14, and 15 illustrates graphs that were

Fig. 6 Detection model accuracy (proposed)

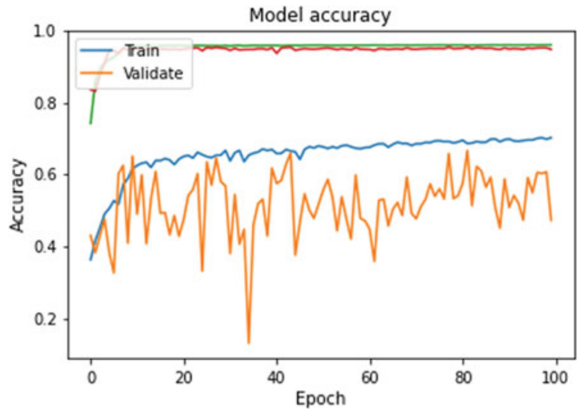


Fig. 7 Detection model loss (proposed)

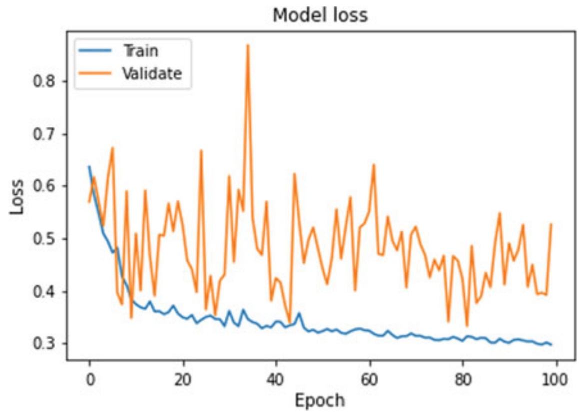
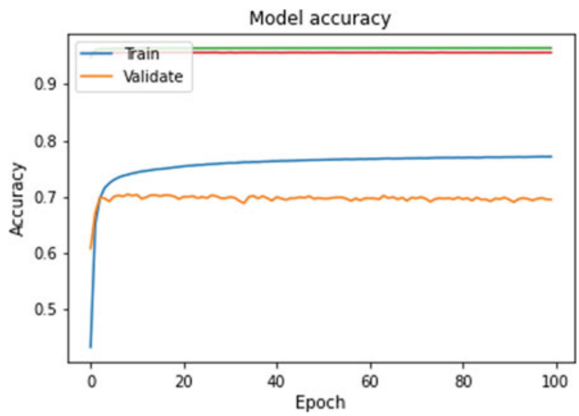


Fig. 8 Detection model accuracy (CNN)



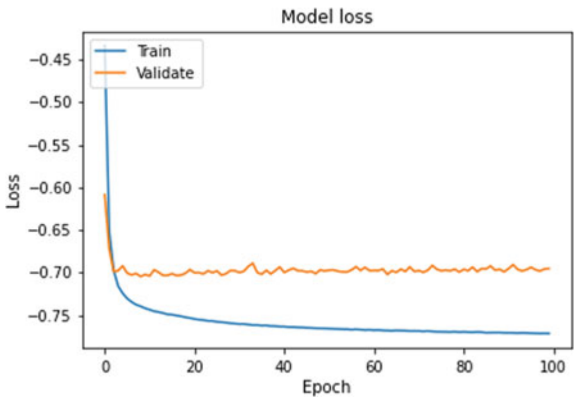


Fig. 9 Detection model loss (CNN)

Table 2 Classification results of proposed and CNN model

Model	Validation measures		
	Accuracy (%)	Recall (%)	Precision (%)
Proposed	99.34	99.34	99.34
CNN	98.83	98.83	98.83

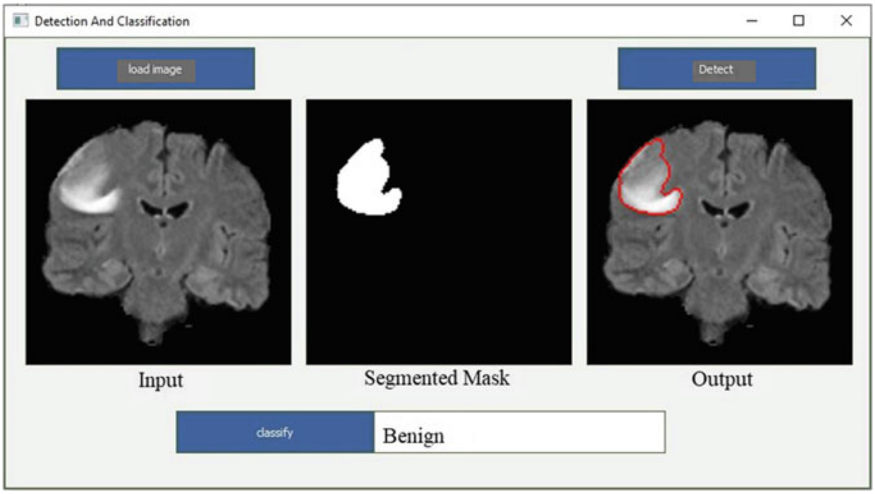


Fig. 10 Classification of detected tumor as Benign

generated while training classification model depicting Model Accuracy and Model loss of Proposed and CNN model, respectively.

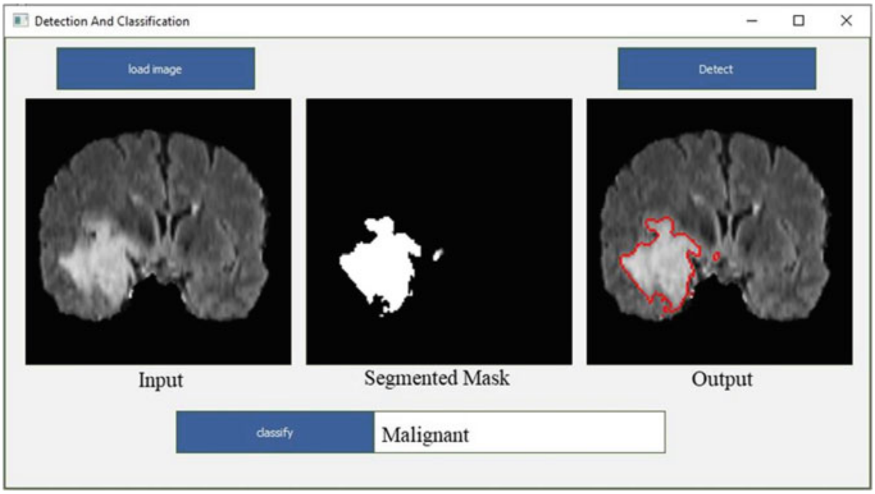


Fig. 11 Classification of detected tumor as Malignant

Fig. 12 Classification model accuracy (proposed)

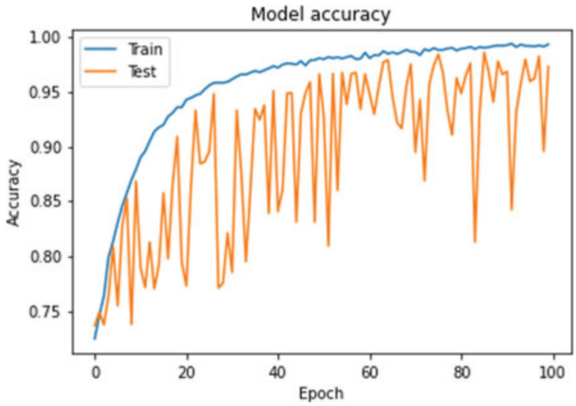
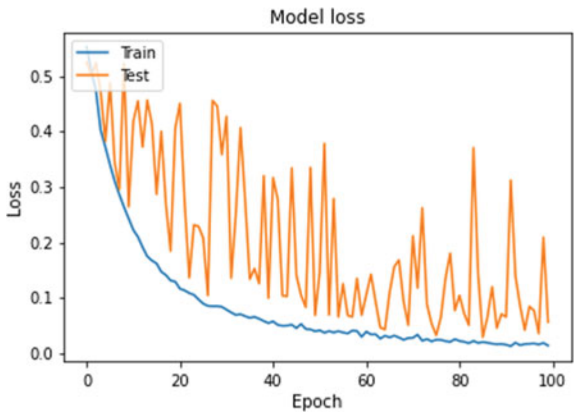


Fig. 13 Classification model loss (proposed)



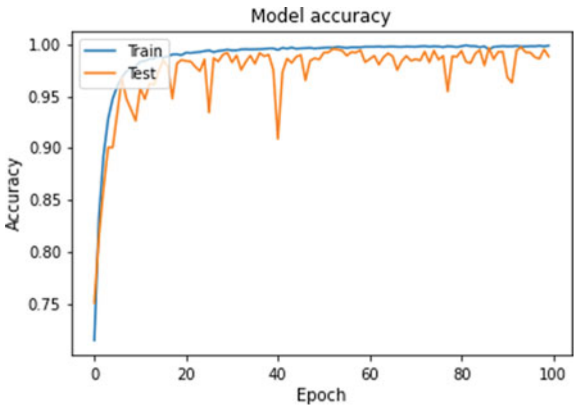


Fig. 14 Classification model accuracy (CNN)

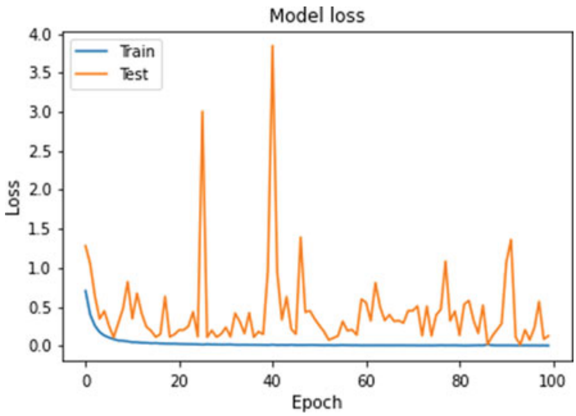


Fig. 15 Classification model loss (CNN)

6 Comparison of Proposed Model

In this section, performance comparison of proposed Detection and Classification model with CNN Model and other existing models is done.

Table 3 shows comparison of Proposed Detection Model with CNN Model and other existing models.

The accuracy attained by the proposed detection model is 96.06% whereas the CNN Model attained accuracy of 96.42%. For proving the efficiency of the proposed Detection model, it is compared with CNN model and published works namely

Table 3 Performance comparison of proposed detection model with CNN and other existing models

Algorithms	Proposed (CNN + LSTM)	CNN	3D CNN [7]	DDU-nets [8]
DSC	70.24%	77.13%	88.12%	89.80%

Table 4 Performance comparison of proposed classification model with CNN and other existing models

Algorithms	Proposed (CNN + LSTM + SVM)	CNN	Naïve Bayes [9]	ELM [9]
Accuracy (%)	99.34	98.83	89.49	92.50

Kaisheng and Wenlian [7] and Hanxiao et al. [8] that used different methods such as 3D CNN and DDU-Nets for performing the same task of Detecting the tumorous regions from MRI Brain scans. The CNN model is developed using V-Net architecture. It is trained for 100 epochs and achieved DSC of 77.13%. In [7] they developed model using Group Normalization and 3D CNN. For enhancing the performance, they added supervise path in encoding and also added two paths in up-sampling of decoding and achieved DSC of 88.12%. In [8] they developed model using Distributed Dense Model (DDU-Nets) and achieved DSC of 89.80%. In comparison of [7] and [8], it is seen that model developed by [8] showed good results by margin of 1.68%. While comparing the Proposed Model with CNN Model it is seen that CNN Model shows higher DSC, whereas the overall comparison too shows that the DSC achieved by DDU-Nets [8] is highest. This can be because the model is trained on High Computing Power with 3D datasets. It can be implied that the proposed model showed good results, on all three sides of Brain Images as the difference of accuracy between the proposed model and CNN model is slightest, i.e., of 0.36%.

Table 4 shows comparison of Proposed Classification Model with CNN Model and other existing models.

The proposed classification model achieved accuracy of 99.34% while the CNN Model achieved accuracy of 98.83%. For proving the efficiency of proposed Classification model, it is compared with CNN model and published works namely Muhammad et al. [9] that used different methods such as Naïve Bayes and ELM (Extreme Learning Machine) for performing the same task of Classifying detected tumors from MRI Brain scans. CNN model achieved accuracy of 98.83%. In [9] they developed model using ELM and achieved accuracy of 92.50% also other model developed using Naïve Bayes is seen to achieve 89.49%. While comparing the Naïve Bayes and ELM method it is seen that ELM method showed higher accuracy by margin of 3.1%. In overall comparison, it is observed that the Proposed Model achieved the highest accuracy of 99.34%, while the accuracy of CNN model is 98.83% which is a slight difference of 0.51%.

Figures 16 and 17 shows pictorial representation of performance comparison of Proposed Model with CNN and other existing models.

By overall analysis it is seen that the training of proposed model and CNN model is done on vast dataset, i.e., HGG: 210 and LGG: 75 case on all three sides of Brain Image, i.e., Front, Side and Top and achieved accuracy of 96.06 and 96.42%, respectively for Detection, and for Classification it is 99.34 and 98.83%, respectively. There is slight difference in accuracy rate. The other existing models are trained for only one side of Brain Image. Hence, the developed model can be concluded to be efficient as we can see good accuracy with combination of CNN-LSTM for Detection and CNN-LSTM-SVM for Classification, respectively.

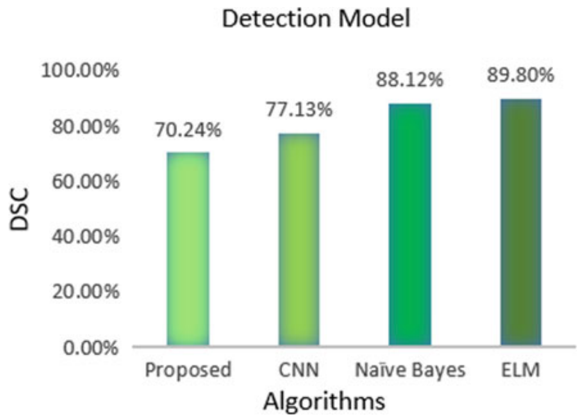


Fig. 16 Performance comparison of detection model

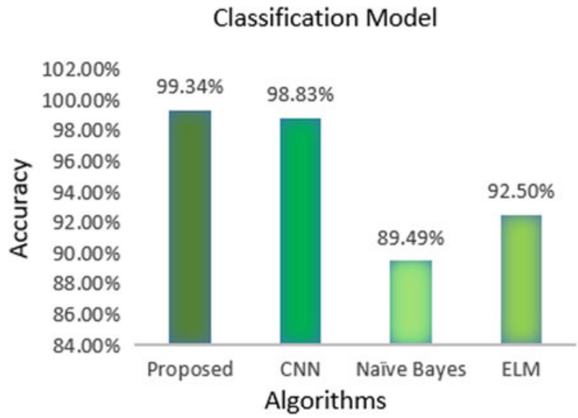


Fig. 17 Performance comparison of classification model

7 Conclusion

The performance of U-Net Architecture for Detection and Sequential Model for Classification is found to be good. The Model is trained and tested for 210 HGG and 75 LGG cases on all three views, i.e., (Front, Side, Top) of MRI Brain Scans and achieved Accuracy of 96.06% for Detection and 99.34% for Classification. Model was trained for 100 epochs for each case and Dropout function was used at various layers to reduce the possibility of overfitting problem. Analysis of the proposed model with CNN and existing models which are developed using other methods also indicate that the method used in proposed model showed good performance. Considering all views of Brain scans to calculate the area of detected tumorous region along with the predicting the survival rate of patient is the future scope of this work.

References

1. Chirodip C, Chandrakanta M, Raghvendra K, Brojo M (2020) Brain tumor detection and classification using convolutional neural network and deep neural network. IEEE
2. Salma S, Rajya D (2020) Combining optimal wavelet statistical texture and recurrent neural network for tumor detection and classification over MRI. Springer
3. Jia Z, Chen D (2020) Brain tumor identification and classification of MRI images using deep learning techniques. IEEE Access
4. Pandiselvi T, Mashwaren R (2019) Efficient framework for identifying, locating, detecting and classifying MRI brain tumor in MRI images. J Med Syst (Springer)
5. Amin J, Sharif M, Yasmin M, Fernandes S (2017) A distinctive approach in brain tumor detection and classification using MRI. Pattern Recogn Lett (Elsevier)
6. Nilesh B, Arun R, Har T (2017) Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. Int J Biomed Eng (Hindawi)
7. Kaisheng L, Wenlian L (2020) Brain tumor segmentation using 3D convolutional neural network. LNCS 11993, pp 199–207. Springer
8. Hanxiao Z, Jingxiong L, Mali S, Yaqi W, Guang-Zhong Y (2020) DDU-Nets: distributed dense model for 3D MRI brain tumor segmentation. LNCS 11993, pp 208–217. Springer
9. Muhammad K, Imran A, Majed A, Robertas D, Rafal S, Amjad R, Syed B (2020) Multimodal brain tumor classification using deep learning and robust feature selection: a machine learning application for radiologists
10. Rahul C, Kamal G, Joshi RC (2018) Convolutional neural network (CNN) for image detection and recognition. In: First international conference on secure cyber computing and computation (ICSCCC). IEEE
11. Angulakshmi M, Lakshmi G (2018) Brain tumor segmentation from MRI using superpixels based spectral clustering. J King Saud Univ Comput Inf Sci
12. Wentau W, Daning L, Jiaoyang D, Xiangyu G, Wen G, Fanfan Z, Xiaojie F, Hong Y (2020) An intelligent diagnosis method of brain MRI tumor segmentation using deep convolutional neural network and SVM algorithm. Comput Math Methods Med
13. Ostrom QT, et al: CBTRUS statistical report: primary brain and other central nervous system tumours diagnosed in the United States in 2012–2016. Neuro-oncology 21, Supplement_5 (2019) Global CEO Survey 2019—Barometer of Corporate Opinion. PwC France Publications. https://www.pwc.fr/fr/publications/dirigeants-et-administrateurs/global-ceo-survey/22ndannual-global-ceo-survey.html
14. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al (2015) The multimodal brain tumor image segmentation benchmark (BRATS) IEEE Trans Med Imaging 34(10):1993–2024. https://doi.org/10.1109/TMI.2014.2377694
15. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al (2017) Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic featuresNat Sci Data 4:170117.https://doi.org/10.1038/sdata.2017.117
16. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A. et al.: Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. arXiv preprint: arXiv:1811.02629 (2018)

Books

17. Taylor LP, Porter Umphrey AB, Richard D: Navigating life with a brain tumor
18. Kaye AH, Laws Jr. ER: Brain tumors an encyclopedia approach, 3rd edn

Websites

19. https://www.cancerresearchuk.org/about-us/cancer-news/press-release/2020-10-06-vital-rethinking-in-cancer-early-detection-needed-to-save-lives
20. CNN simplilearn.com
21. https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/
22. https://bestbraintumorcancer.blogspot.com/2018/10/brain-tumors-temporal-lobe.html
23. https://www.geeksforgeeks.org/data-normalization-in-data-mining/

A Simplified Parameter Adaptive DCPCNN Based Medical Image Fusion

Chirag Agrawal, Sujit Kumar Yadav, Shreyaskar Pratap Singh, and Chinmaya Panigrahy

Abstract Dual-channel pulse coupled neural network (DCPCNN) is extensively applied to image fusion because of properties, viz., simultaneous processing of two images, pulse synchronization of neurons, and global coupling, but manual setting of its parameters adversely influence the fusion results. In this paper, a new simplified parameter adaptive DCPCNN (SPADCPCNN) model-based medical image fusion framework is presented in non-subsampled shearlet transform (NSST) domain, where all the parameters of SPADCPCNN are automatically computed from its inputs and all the neurons have the same linking strength. Firstly, NSST is exercised on source images to get a low-pass sub-band and several high-pass sub-bands, respectively. Secondly, both the low-pass and high-pass sub-bands are fused using the SPADCPCNN to produce the fused low-pass and high-pass sub-bands, respectively. Lastly, the fused image is constructed from the fused sub-bands by employing inverse NSST. Experimental results illustrate the superiority of the proposed method over some of the existing methods with regard to both subjective and objective evaluations.

Keywords Simplified parameter adaptive dual-channel pulse coupled neural network (SPADCPCNN) · Medical image fusion · Transform domain-based (TDB) method · Non-subsampled shearlet transform (NSST) · Objective metric

C. Agrawal
PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, Madhya Pradesh 482005, India
e-mail: 2019234@iiitdmj.ac.in

S. K. Yadav · S. P. Singh
GLA University, Mathura, Uttar Pradesh 281406, India
e-mail: sujit.yadav_cs17@gla.ac.in

S. P. Singh
e-mail: shreyaskar.singh_cs17@gla.ac.in

C. Panigrahy (✉)
Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India
e-mail: chinmaya.panigrahy@thapar.edu

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_40

1 Introduction

Medical image fusion merges the complementary data from various medical images taken at single or different modalities to obtain an enhanced image more suitable for medical diagnosis, machine perception, and assessment of medical problems, where specific features of organs and tissues are preserved in the output image [1]. For instance, magnetic resonance imaging (MRI) images contain high-resolution structural and anatomical information, whereas positron emission tomography (PET) and single-photon emission computed tomography (SPECT) images carry low-resolution functional information [1]. So, a more informative image suitable for noninvasive diagnosis of diseases is acquired by fusing MRI and PET/SPECT images, which will have both the anatomical and functional information.

Generally, two types of medical image fusion methods exist in literature, namely, transform domain-based (TDB) and spatial domain-based methods. Among these types of methods, the TDB methods are commonly used in image fusion because of their efficient extraction of middle-scale edges, large-scale spatial shapes, and fine-scale texture information [1]. The TDB methods mainly contain three steps: decomposition of source images into sub-bands by applying a specific transform, fusion of corresponding sub-bands to get the fused sub-bands, and reconstruction of the fused image from the fused sub-bands by utilizing the respective inverse transform. Some of the TDB methods are accomplished using non-subsampled contourlet transform (NSCT) [2], Laplacian pyramid (LP) [3], non-subsampled shearlet transform (NSST) [1, 4]. The shift-invariant feature of NSCT and NSST make them more suitable for image fusion. However, NSST is preferred over NSCT due to its low computational cost [4].

In the past decade, pulse coupled neural network (PCNN) acquired from cat's visual cortex is widely applied to image fusion because of its properties, such as pulse synchronization of neurons and global coupling [5]. Nevertheless, its manual setting of parameters adversely affects the fusion performance. Moreover, it is unable to process two images simultaneously, which further limits its applicability. In TDB methods, PCNNs are used to devise fusion rules for the decomposed sub-bands. Though many researchers proposed simplified PCNN models [5, 6] to reduce the complexity of traditional PCNN, few of them concentrated on adaptive parameter settings [4]. In [4], authors introduced the parameter adaptive PCNN (PAPCNN) [7] to image fusion, whose all the parameters are automatically evaluated from its inputs. However, like PCNN, it also fails to process two images at a time, leading to the usage of dual-channel PCNN (DCPCNN). Therefore, a parameter adaptive DCPCNN (PADCPCNN) is proposed in [5] to fix the above issue. However, the linking strengths of PADCPCNN model are estimated per neuron, which increased the complexity. Therefore, a simplified PADCPCNN (SPADCPCNN) is proposed in this paper, where all neurons are assigned with the same linking strength to reduce the intricacy. Here, the fixed linking strength along with all the parameters are automatically evaluated from its inputs.

In this paper, a novel SPADCPCNN model-based TDB medical image fusion technique is introduced in NSST domain. The SPADCPCNN is used to fuse both the low-pass and high-pass sub-bands of the source images. The fusion results on MRI-MRI, MRI-PET, and MRI-SPECT image pairs and four objective metrics are considered to compare the performance of proposed method with five existing techniques.

The rest of the paper is as follows: Sect. 2 introduces the new SPADCPCNN model. The proposed medical image fusion framework is discussed in Sect. 3, whereas Sect. 4 examines the experimental results. Lastly, Sect. 5 draws the conclusion.

2 SPADCPCNN Model

The PADCPCNN [5] provides better fusion results by estimating various parameters from its inputs. However, its adaptive linking strength makes the model expensive. Therefore, we modified PADCPCNN and proposed an efficient SPADCPCNN which uses a fixed linking strength for all the neurons.

Figure 1 shows the architecture of the SPADCPCNN model which possesses three segments, viz., the receptive field (RF), information fusion pool (IFP), and pulse generator (PG). The RF acquires the stimuli out-of external inputs and surrounding neurons, whereas IFP fuses the external inputs by computing the activity level of neurons. Here, source images are treated as external stimuli. The PG generates the output pulse. The number of pixels in a source image is equal to the number of neurons in the SPADCPCNN model. Mathematically, the SPADCPCNN model is represented using Eqs. (1) to (8).

$$F_n^A(i, j) = S^A(i, j), \tag{1}$$

$$F_n^B(i, j) = S^B(i, j), \tag{2}$$

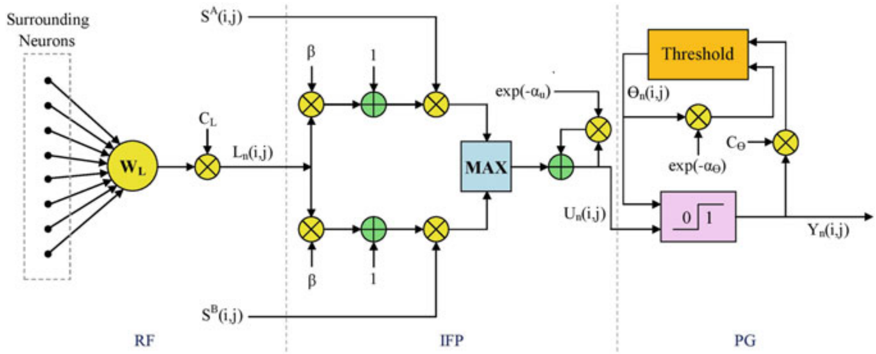


Fig. 1 Proposed SPADCPCNN model

$$L_n(i,j) = C_L \sum_{x=-1}^{1} \sum_{y=-1}^{1} W_L(x+2, y+2) Y_{n-1}(i+x, j+y), \tag{3}$$

$$U_n^A(i,j) = F_n^A(i,j)(1 + \beta L_n(i,j)), \tag{4}$$

$$U_n^B(i,j) = F_n^B(i,j)(1 + \beta L_n(i,j)), \tag{5}$$

$$U_n(i,j) = \mathrm{e}^{-\alpha_u} U_{n-1}(i,j) + \max\left\{U_n^A(i,j), U_n^B(i,j)\right\}, \tag{6}$$

$$Y_n(i,j) = \begin{cases} 1, \text{ if } U_n(i,j) > \theta_{n-1}(i,j) \\ 0, \text{ Otherwise} \end{cases}, \tag{7}$$

$$\theta_n(i,j) = \mathrm{e}^{-\alpha_\theta} \theta_{n-1}(i,j) + C_\theta Y_n(i,j), \tag{8}$$

where $S^X(i,j)|X \in \{A, B\}$ represents the external input of SPADCPCNN at (i, j)th position of the source image, X, where $S^X(i,j) \in [0, 1]$. $F_n^X(i,j)|X \in \{A, B\}$ is the feeding input corresponding to X that store $S^X(i,j)$. The same feeding input is considered for all the iterations. The linking input for the (i, j)th neuron at the nth iteration is denoted by $L_n(i,j)$. The synaptic weights of 8-neighboring neurons of the (i, j)th neuron are specified by a 3×3 matrix, W_L, whose elements are fixed by Eq. (9).

$$W_L = \begin{bmatrix} 0.5 & 1 & 0.5 \\ 1 & 0 & 1 \\ 0.5 & 1 & 0.5 \end{bmatrix}. \tag{9}$$

$U_n^X(i,j)|X \in \{A, B\}$ represents the internal state of (i, j)th neuron corresponding to X after the nth iteration, which represents the non-linear modulation of corresponding feeding and linking inputs. The internal activity of the SPADCPCNN model corresponding to the (i, j)th neuron after the nth iteration is denoted by $U_n(i,j)$**.** The internal state of source images is compared to generate the fused image. $\theta_n(i,j)$ and $Y_n(i,j)$ are the threshold function and external output of the (i, j)th neuron of the SPADCPCNN after the nth iteration, respectively. β is the linking strength of each neuron. C_L, C_Θ, α_u, and α_Θ are four parameters of the SPADCPCNN. C_Θ and C_L represent the amplitudes of threshold function and linking input, respectively. α_u and α_Θ are two exponential decay coefficients. The values of C_L, α_u, C_Θ, α_Θ, and β are estimated using the parameter settings of PAPCNN [7] and PADCPCNN [5], which are determined using Eqs. (10) to (14), respectively.

$$C_L = 1. \tag{10}$$

$$\alpha_u = \log\left(\frac{2}{\sigma_{S^A} + \sigma_{S^B}}\right), \tag{11}$$

where $\sigma_{S^X}|X \in \{A, B\}$ is the standard deviation of S^X.

$$C_\theta = \mathrm{e}^{-\alpha_u} + 1 + 6\beta, \tag{12}$$

$$\alpha_\theta = \ln\left(\frac{2C_\theta}{\left(S^{A'} + S^{B'}\right)\left(\frac{1-\mathrm{e}^{-3\alpha_u}}{1-\mathrm{e}^{-\alpha_u}} + 6\beta\mathrm{e}^{-\alpha_u}\right)}\right), \tag{13}$$

where $S^{X'}|X \in \{A, B\}$ is the Otsu threshold of S^X.

$$\beta = \frac{S^A_{\max}S^{B'} + S^B_{\max}S^{A'} - 2S^{A'}S^{B'}}{12S^{A'}S^{B'}}, \tag{14}$$

where $S^X_{\max}|X \in \{A, B\}$ is the maximum value of S^X.

The proposed SPADCPCNN model receives two images and estimates their internal activities after each iteration, which are efficiently utilized to construct the fused image. The SPADCPCNN model is based on PADCPCNN and PCNN, hence it holds the properties like pulse synchronization of neurons and global coupling. These properties help in extracting the complementary information from two images by properly using the local image information.

3 Proposed Medical Fusion Framework

Medical images are of two types, namely gray-scale and pseudo-color. MRI images are gray-scale images. On the other hand, PET and SPECT images are pseudo-color images that can be viewed as RGB color images [4]. Here, the pseudo-color images are first transformed into YUV color space for better fusion operation, where Y is the luminance component and U and V represent two chrominance components. The various steps of the proposed method to fuse a gray-scale and a pseudo-color medical image are described as follows.

Step 1: The normalized pseudo-color image is first transformed into YUV color space to get the Y, U, and V components.

Step 2: An L level NSST decomposition is applied to the normalized gray-scale image, A, and the Y component of the pseudo-color image, B, to generate the sub-bands $\left\{\mathrm{LS}_X, \mathrm{HS}_X^{l,d}\right\}|X \in \{A, B\}$, where $\mathrm{LS}_X|X \in \{A, B\}$ is the low-pass sub-band of X and $\mathrm{HS}_X^{l,d}|X \in \{A, B\}$ is the high-pass sub-band of X acquired at the lth decomposition level with the direction, d. $l \in [1, L]$ and $d \in [1, D(l)]$, where the vector, D, stores the number of directions at each l.

Step 3: The fused low-pass sub-band, LS_F, is acquired from LS_A and LS_B by initializing the proposed SPADCPCNN model as follows: $F_n^A(i, j) = |LS_A(i, j)|$ and $F_n^B(i, j) = |LS_B(i, j)|$ for $n \in [1, N]$. Here, N represents the total number of iterations. $U_0(i, j) = 0$, $Y_0(i, j) = 0$, and $\theta_0(i, j) = 1$. Here, F is the Y component of the fused image. The value of β is computed by Eq. (14). Equations (10) to (13) are utilized to estimate the free parameters, C_L, α_u, C_Θ, and α_Θ, respectively. After running the SPADCPCNN model up to N iterations, LS_F is computed using Eq. (15).

$$LS_F(i, j) = \begin{cases} LS_A(i, j) \text{ if } U_N^{LS_A}(i, j) \geq U_N^{LS_B}(i, j) \\ LS_B(i, j) \qquad \text{otherwise} \end{cases}, \tag{15}$$

where $U_N^{LS_X}(i, j)|X \in \{A, B\}$ is the internal state of (*i*,*j*)th neuron corresponding to LS_X after Nth iteration.

Step 4: The high-pass sub-bands are also combined by the SPADCPCNN like the low-pass sub-bands. The SPADCPCNN model is loaded with the following specifications $\forall l, d$: $F_n^A(i, j) = \left|HS_A^{l,d}(i, j)\right|$ and $F_n^B(i, j) = \left|HS_B^{l,d}(i, j)\right|$ for $n \in [1, N]$. Here, N represents the total number of iterations. $U_0(i, j) = 0$, $Y_0(i, j) = 0$, and $\theta_0(i, j) = 1$. The value of β is estimated from Eq. (14). Equations (10) to (13) are employed to approximate the parameters, C_L, α_u, C_Θ, and α_Θ, respectively. The SPADCPCNN model is executed up to N iterations for each of the respective high-pass sub-bands of A and B Then, the fused high-pass sub-bands, $HS_F^{l,d}(i, j) \forall l, d$, can be computed using Eq. (16).

$$HS_F^{l,d}(i, j) = \begin{cases} HS_A^{l,d}(i, j) \text{ if } U_N^{A^{l,d}}(i, j) \geq U_N^{B^{l,d}}(i, j) \\ HS_B^{l,d}(i, j) \text{ otherwise} \end{cases}, \tag{16}$$

where $U_N^{X^{l,d}}(i, j)|X \in \{A, B\}$ is the $U_N^X(i, j)$ of $\left|HS_X^{l,d}(i, j)\right|$.

Step 5: The Y component of the fused image, F, is constructed from the fused sub-bands, $\left\{LS_F, HS_F^{l,d}\right\}$, by applying inverse NSST.

Step 6: The Y component of the fused image, F, is first combined with the U and V components of the pseudo-color image and then, inverse YUV transformation is applied to get the normalized fused color image which is finally standardized to generate the fused color image.

The proposed fusion framework is illustrated in Fig. 2. When the second input image is also a gray-scale image, the concept of YUV transformation described in the above-mentioned steps is discarded.

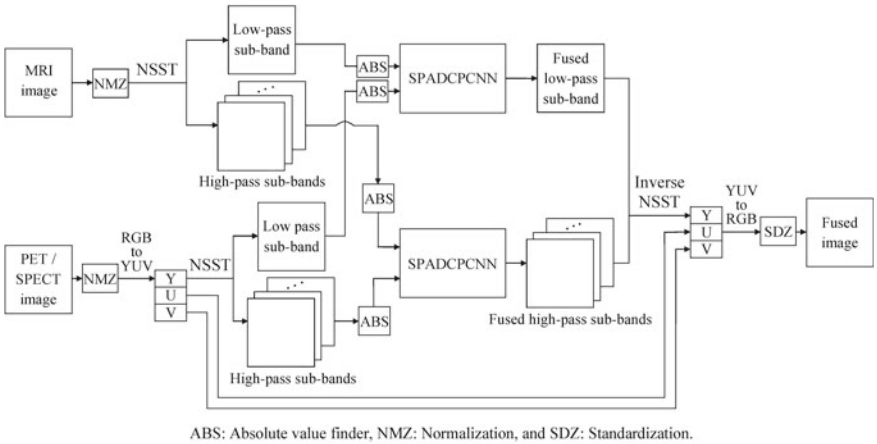


Fig. 2 Proposed medical image fusion method

4 Experimental Results and Discussion

Six medical image pairs [9], viz., two each of MRI-MRI, MRI-PET, and MRI-SPECT image pairs are shown in Figs. 3, 4, and 5, respectively, are used in the experiment, which are of size 256 × 256 pixels. Five state-of-the-art methods, i.e., NSST with weighted PADCPCNN (NSST-WPADCPCNN) [1], NSST-PAPCNN [4], Convolutional neural network (CNN) [8], LP and sparse representation (LP-SR) [3], and NSCT-PCNN [2] are considered to show the competitiveness of the proposed method. These methods are among recent and well-established methods. The performance of different methods is compared in two ways, namely subjective and objective evaluations. The subjective evaluation corresponds to the visual perception of fusion results, whereas the objective evaluation is done using objective metrics. Therefore, four objective metrics, namely, structural similarity (SSIM) [10], Petrovic's metric ($Q^{AB/F}$) [11], correlation coefficient (ρ^F) [12], and entropy ($E(F)$) [13] are used for objective evaluation. $E(F)$ assesses the level of information contained by the fused image, ρ^F depicts the correlation between fused and source images, $Q^{AB/F}$ quantifies the edge information preserved by the fused image that is present in the input images, and *SSIM* counts the similarity among fused and source images. Higher values of these four metrics lead to a better fusion method.

The fusion results of various methods for the MRI-MRI, MRI-PET, and MRI-SPECT image pairs are displayed in Figs. 3, 4, and 5, respectively. The objective assessment of various methods is illustrated in Table 1, where the worst values of each metric are highlighted in bold. It is interpreted from Fig. 3 that the fusion results of NSCT-PCNN and CNN missed some bright regions in the first image pair, whereas introduced some dark regions in the second. The fused image obtained by NSST-PAPCNN for the second image pair contains blurred boundaries. The proposed

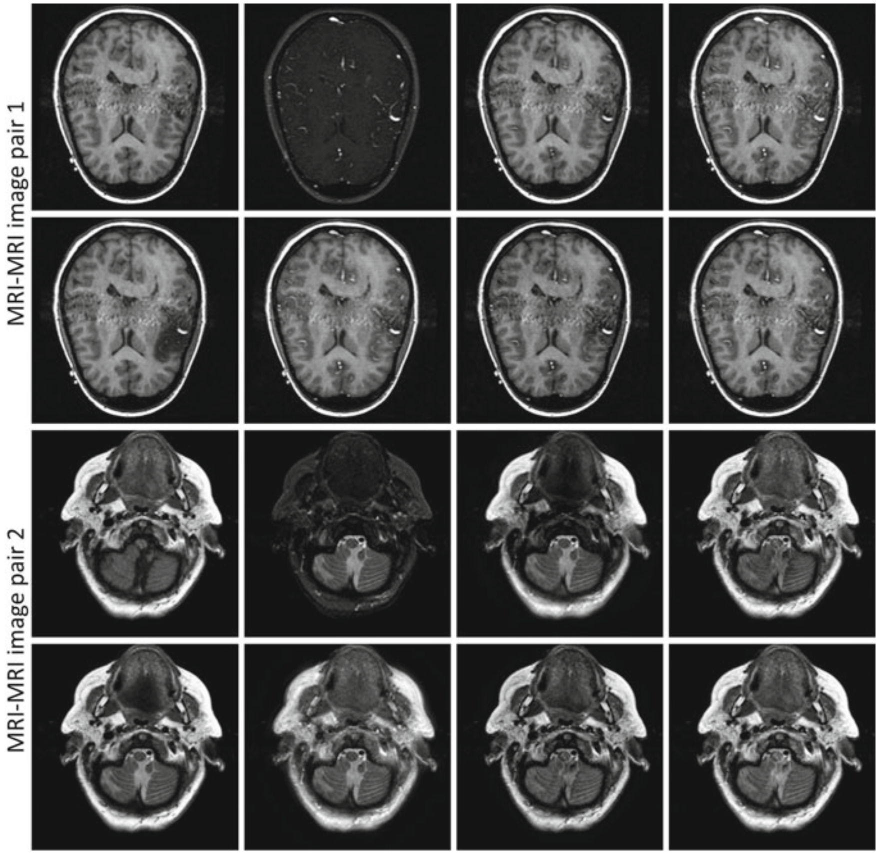


Fig. 3 Fusion results for MRI-MRI image pairs. From west to east and north to south: MRI image 1, MRI image 2, fusion results of NSCT-PCNN [2], LP-SR [3], CNN [8], NSST-PAPCNN [4], NSST-WPADCPCNN [1], and Proposed method

method along with LP-SR and NSST-WPADCPCNN provides better-fused images for the MRI-MRI image pairs.

The color variation of the functional information in the PET images is improperly extracted by NSCT-PCNN and CNN for the first image pair that is manifested in Fig. 4. CNN failed to preserve the structural and anatomical information in the fusion result of the second image pair. LP-SR and NSST-WPADCPCNN provide better fusion results for the MRI-PET image pairs, which are closely followed by the proposed method and NSST-PAPCNN.

It can be concluded from Fig. 5 that the proposed method gives the best fusion results by well preserving the structural/anatomical details of the MRI images and functional details of the SPECT images. Among the existing methods, NSST-PAPCNN produces better results. NSCT-PCNN, LP-SR, CNN, and NSST-WPADCPCNN introduced darker regions into the fused image of the first image

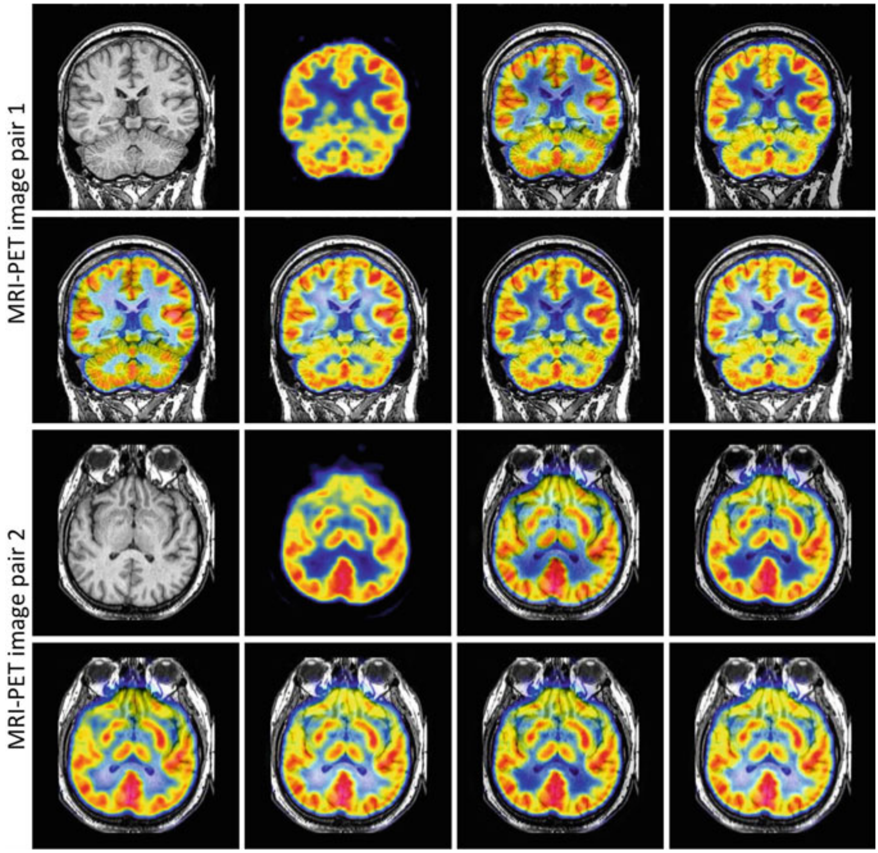


Fig. 4 Fusion results for MRI-PET image pairs. From west to east and north to south: MRI image, PET image, fusion results of NSCT-PCNN [2], LP-SR [3], CNN [8], NSST-PAPCNN [4], NSST-WPADCPCNN [1], and Proposed method

pair. NSCT-PCNN and CNN missed the small yellowish functional part on the east region of the second image pair. For this image pair, LP-SR introduced some darker regions.

Table 1 reveals that the proposed method never generates the worst value for any of the mentioned metrics. Moreover, it is the only method that exhibits such trend. In maximum cases, NSCT-PCNN provides the worst objective results, which is closely followed by CNN and LP-SR. The objective assessments of NSST-PAPCNN, NSST-WPADCPCNN, and the proposed method are satisfactory.

It can be concluded from the objective and subjective evaluations of various methods that the proposed method provides satisfactory fusion results, whose performance is comparable to some of the state-of-the-art methods. Though the proposed method does not always generate the best results, it never produced the worst results. Simultaneously, the fusion results of the proposed method are at par with

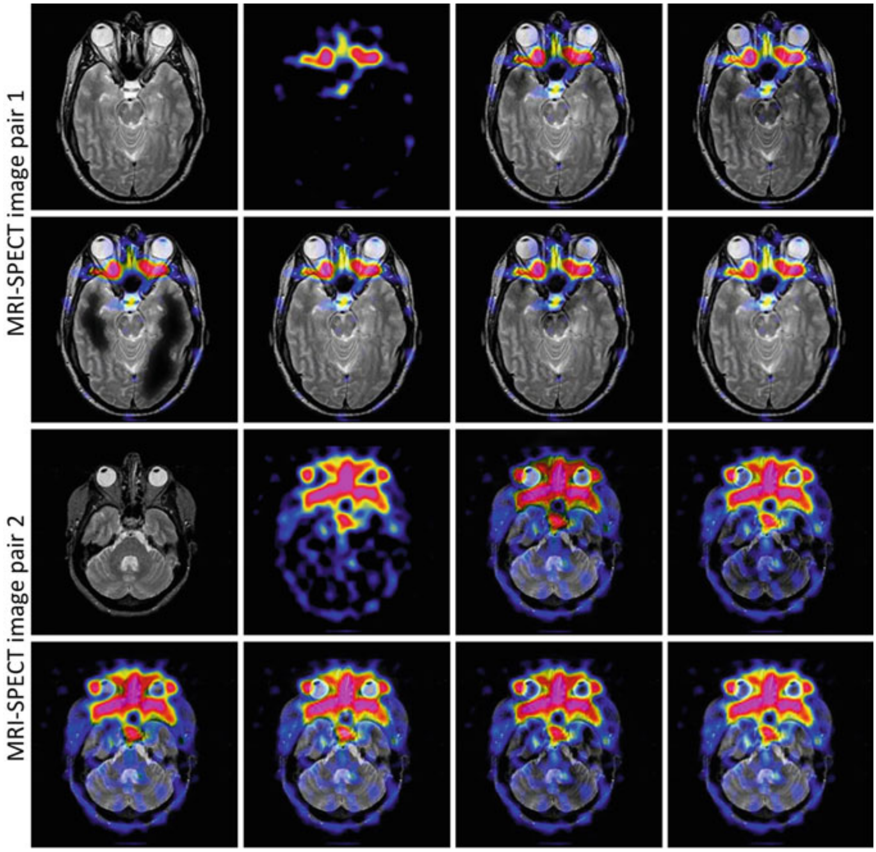


Fig. 5 Fusion results for MRI-SPECT image pairs. From west to east and north to south: MRI image, SPECT image, fusion results of NSCT-PCNN [2], LP-SR [3], CNN [8], NSST-PAPCNN [4], NSST-WPADCPCNN [1], and Proposed method

the discussed approaches, which even beats few of them. Therefore, as a potential method, the proposed method can be employed to fuse medical images.

5 Conclusion

This paper presents a new SPADCPCNN model-based medical image fusion framework in NSST domain, where pseudo-color source images are properly handled using the YUV color space. The SPADCPCNN model is employed to fuse both the low-pass sub-bands and high-pass sub-bands of the source images. Fusion results of MRI-MRI, MRI-PET, and MRI-SPECT image pairs using five state-of-the-art methods and four objective metrics signify the competitiveness of the proposed method. Future scope

Table 1 Objective evaluation of various methods

Image pair	Objective metric	NSCT-PCNN [2]	LP-SR [3]	CNN [8]	NSST-PAPCNN [4]	NSST-WPADCPCNN [1]	Proposed method
MRI-MRI image pair 1	$E(F)$	6.1394	5.6529	**5.4534**	6.6468	6.3811	6.5766
	ρ^F	0.9143	0.9272	**0.8999**	0.9281	0.9220	0.9272
	$Q^{AB/F}$	0.8527	0.8763	**0.8411**	0.8649	0.8713	0.8682
	SSIM	0.6856	**0.6846**	0.6866	0.6872	0.7049	0.6966
MRI-MRI image pair 2	$E(F)$	5.8659	**5.3127**	5.3569	5.9898	5.5720	5.4404
	ρ^F	**0.7814**	0.7993	0.8002	0.8248	0.8136	0.8201
	$Q^{AB/F}$	0.7315	0.8407	0.8338	**0.7070**	0.8385	0.8418
	SSIM	**0.6266**	0.7264	0.7293	0.6527	0.7087	0.7393
MRI-PET image pair 1	$E(F)$	6.0226	**5.2229**	5.2891	5.8570	5.9556	6.0566
	ρ^F	0.8086	0.8451	**0.7998**	0.8293	0.8295	0.8344
	$Q^{AB/F}$	0.8850	**0.8685**	0.8893	0.9013	0.8980	0.8911
	SSIM	0.6731	0.6640	0.6911	0.6900	**0.6599**	0.6784
MRI-PET image pair 2	$E(F)$	6.1034	**5.0614**	5.2943	5.5346	5.6796	5.7224
	ρ^F	**0.8241**	0.8576	0.8294	0.8438	0.8455	0.8487
	$Q^{AB/F}$	**0.8825**	0.8937	0.8866	0.9037	0.9059	0.8959
	SSIM	**0.6595**	0.6919	0.7125	0.7139	0.6858	0.7044
MRI-SPECT image pair 1	$E(F)$	5.0339	4.5203	4.6237	**4.4683**	4.8083	4.9675
	ρ^F	0.6204	0.6898	**0.6120**	0.6578	0.6517	0.6688
	$Q^{AB/F}$	0.9075	0.9091	**0.8870**	0.9188	0.9136	0.9088
	SSIM	0.6816	0.6868	**0.6686**	0.7032	0.6911	0.6936
MRI-SPECT image pair 2	$E(F)$	4.7714	**4.2710**	4.2905	4.7481	4.5354	4.7252

(continued)

Table 1 (continued)

Image pair	Objective metric	NSCT-PCNN [2]	LP-SR [3]	CNN [8]	NSST-PAPCNN [4]	NSST-WPADCPCNN [1]	Proposed method
	ρ^F	**0.7330**	0.7647	0.7467	0.7754	0.7703	0.7780
	$Q^{AB/F}$	**0.7344**	0.8049	0.7697	0.8165	0.8027	0.8035
	SSIM	**0.7083**	0.7389	0.7555	0.7344	0.7432	0.7453

lies in investigating more image fusion applications, like, infrared-visible, multi-exposure, multi-focus, and remote sensing image fusion. The proposed method can be extended to work for above image fusion domains. The various PCNN models can be applied to other image processing applications, such as image enhancement and segmentation.

References

1. Panigrahy C, Seal A, Mahato NK (2020) MRI and SPECT image fusion using a weighted parameter adaptive dual channel PCNN. IEEE Signal Process Lett 27:690–694
2. Xiao-Bo Q, Jing-Wen Y, Hong-Zhi X, Zi-Qian Z (2008) Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain. Acta Autom Sinica 34(12):1508–1514
3. Liu Y, Liu S, Wang Z (2015) A general framework for image fusion based on multi-scale transform and sparse representation. Inf Fusion 24:147–164
4. Yin M, Liu X, Liu Y, Chen X (2018) Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. IEEE Trans Instrum Meas 99:1–16
5. Panigrahy C, Seal A, Mahato NK (2020) Fractal dimension based parameter adaptive dual channel PCNN for multi-focus image fusion. Opt Lasers Eng 133:106141
6. Huang C, Tian G, Lan Y, Hao Y, Cheng Y, Peng Y, Che W (2019) A new pulse coupled neural network (PCNN) for brain medical image fusion empowered by shuffled frog leaping. Front Neurosci 13:210
7. Chen Y, Park SK, Ma Y, Ala R (2011) A new automatic parameter setting method of a simplified PCNN for image segmentation. IEEE Trans Neural Networks 22(6):880–892
8. Liu Y, Chen X, Peng H, Wang Z (2017) Multi-focus image fusion with a deep convolutional neural network. Inf Fusion 36:191–207
9. The whole Brain Atlas. http://www.med.harvard.edu/AANLIB/home.html. Last accessed 31 May 2021
10. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
11. Xydeas CS, Petrovic V (2000) Objective image fusion performance measure. Electron Lett 36(4):308–309
12. Seal A, Bhattacharjee D, Nasipuri M, Rodríguez-Esparragón D, Menasalvas E, Gonzalo-Martin C (2018) PET-CT image fusion using random forest and à-trous wavelet transform. Int J Numer Methods Biomed Eng 34(3):e2933
13. Panigrahy C, Seal A, Mahato NK, Krejcar O, Herrera-Viedma E (2020) Multi-focus image fusion using fractal dimension. Appl Opt 59(19):5642–5655

Retinal Blood Vessel Segmentation Using Attention Module and Tversky Loss Function

Paresh Chandra Sau

Abstract Retinal blood vessel segmentation plays significant role in the diagnosis of fundus diseases like Diabetes retinopathy, Glaucoma, macular degeneration, etc. Numerous algorithm of above said problem has been cited in the literature. Some more attention is required to overcome the shortcomings of these algorithms like database invariability, anti-noise interference ability. This paper proposed an algorithm based on well-known U-Net supported by attention module which will be able to segment the thin blood vessel up to few pixel spans giving more attention to the gradient changes. Proposed method reduced the requirement of memory requirement as number of learnable parameters reduced and it requires less number of training datasets as compared to U-Net. Proposed method has been tested on DRIVE and STARE datasets and achieve the AUC as 0.980 and 0.974 respectively.

Keywords Attention module · Diabetes retinopathy · U-Net · Thin vessel

1 Introduction

Human eyes can be treated as mirror of anatomical changes. The fundus image enriches pathological information which can help to diagnosis of eye-related diseases like diabetic retinopathy, glaucoma, macular degeneration. Fundus image caries many features, among which retinal vessel tree play very important role, especially to diagnosis of diabetic retinopathy. For accurate diagnosis of retinopathy, blood vessel segmentation should be accurate. But due to the pathology present over fundus images, it is very challenging task to segment vessel trees accurately especially when tiny vessels are tending to capillary. Manual annotation by ophthalmologist is very time-consuming and tedious. The advancement of artificial intelligence has been exploited by researchers to segment blood vessels using image processing techniques which is faster and with more accuracy.

P. C. Sau (✉)
GLA University, Mathura, India
e-mail: Paresh.chandra@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_41

The paper is organized in following structure. Section 1 gives the introduction to the purpose of this article. This part mainly introduces the importance of the early diagnosis of fundus-related diseases. The state of art and proposed algorithm of previous researchers has been briefly discussed in Sect. 2. Section 3 introduces the detailed methodology of present work. In Sect. 4, result of proposed algorithm has been discussed. Finally, in section, conclusions have been drawn and future work has been proposed. The article has been ended with references.

2 Related Works

In literature, blood vessel segmentation processes are broadly divided into two classes namely supervised and unsupervised algorithms.

Unsupervised algorithms are based on the extraction of manual features like matched filter-based [1–4], model-based [5–8], based on vessel tracking [9, 10], and mathematical morphology-based [11–14]. Zhao et al. [14] proposed a hybrid region growing technique segment infinite active contour.

Supervised methods are those in which the networks are learned the features from annotated data (ground truth data) and input raw data. Supervised algorithms again can be sub-classified into shallow learning base algorithms and deep learning algorithms. Shallow learning uses hand-crafted features for classification and segmentation. Soares et al. [15] proposed wavelet transformation response based on two-dimension Gabor filter and pixel intensity feature vectors. He employed Bayesian classifier for vessel segmentation. AdaBoost classifier has been employed on features based on local pixel intensity, geometry, and spatial properties by Lupascu et al. [16]. Fraz et al. [17] developed classifier based on features consisting of gradient, morphology, Gabor filter, and line strength. In deep learning-based algorithm exploited the convolutional networks in which features are automatically learned during training. These algorithms are again classified into two categories. In first-class, algorithms are based on deep networks. In literature, number of different networks have been reported to improve the accuracy of the segmentation of blood vessels. Fu et al. [18, 19] proposed fully convolutional network followed by fully connected convolutional random fields (CRFs) where blood vessels are treated as boundary detection problems. Wang et al. [20] proposed a novel algorithm based on combination of Dense U-Net for training and random transformation for image augmentation. Other frequently used networks architectures are AlexNet [21], VGGNet [22], GoogLeNet [23] and CDNet [24].

When segmentation is treated as image classification task, the loss function used for training plays a very important role in performance. Simple most loss function reported in literature is the cross-entropy loss function. The cross-entropy loss function is suitable for the classification when images have equal probable among different classes. But most medical images have some sort of class imbalances. To overcome this problem, cross-entropy loss function has been modified in different ways to improve the accuracy of segmentation. Brosch et al. [25] proposed weighted binary cross-entropy loss function for lesson segmentation of brain to train deep convolution encoder networks. Same loss function was also used by Ronneberger et al. [26] in his U-Net for training. Milletari et al. [27] proposed novel loss function Dice loss function using Dice score coefficient which is measure of overlap.

3 Methodology

3.1 Attention Module

In this work attention module is incorporated along with UNet architecture. Consider $x^k = \{x_i^k\}_{i=1}^n$ is the activation mapping of a given layer $k \in [1, 2, 3, \ldots, K]$ where x_i^k is pixel-wise feature vector of length F_1 which is equal to number of channels. Attention units produce coefficient $\alpha^k = \{\alpha_i^k\}_{i=1}^n$ for each x_i^k which will preserve salient feature regions. The output of attention module thus can be represented as $\hat{x}^k = \{\alpha_i^k x_i^k\}_{i=1}^n$. Each feature vector is scaled down by corresponding coefficient of attention. Attention can be used in two different forms: (i) multiplicative form is faster and more memory efficient. This type of attention is used by Luong et al. [28]. (ii) additive attention produces better performances where the feature vector has large dimensions. In this work, additive attention has been used to accommodate large dimensional feature vectors as shown in Fig. 1.

Attention modulated can be formulated as follows:

$$q_{\text{att},i}^k = \psi^T \{\sigma_1 (W_x^T x_i^k + W_g^T g + b_{xg})\} + b_\psi$$
$$\alpha^k = \sigma_2 \{q_{\text{att}}^k (x^l, g; \Theta_{\text{att}})\}$$

where $\sigma_1(.)$ represent ReLU function and σ_2 represent normalization function. $\Theta_{\text{att}}(.)$ represent the attention module parameter set. The dimensions of various terms are as follows: $W_g \in R^{F_g \times F_{\text{int}}}$, $\Psi \in R^{F_{\text{int}} \times 1}$, bias term $b_\Psi \in R$ and $b_{xg} \in R^{F_{\text{int}}}$.

Attention module parameters are trained by back-propagation updates and this does not require auxiliary loss function to optimize. Deep supervision [Lee et al.]

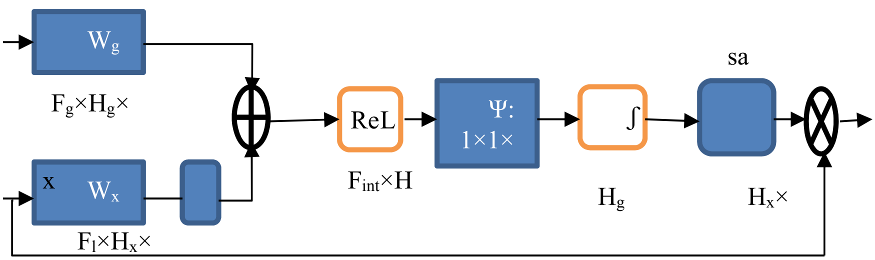


Fig. 1 Schematic diagram of attention module

suggest that intermediate feature maps are semantically discriminative at each image scale. This guarantees that attention module has the ability to influence the large range of image foreground content. Present task of retinal blood vessel segmentation has only two classes, background and blood vessel. Hence, two-dimensional attention module is used.

3.1.1 Attention Signal and Attention Grid

Attention signal g encodes global information from large spatial context and it is obtained by activation mapping. For the classification task, activation mapping is done just before softmax or cross-entropy layer. In this work, max-pooling operations have been employed to infer the global context without explicitly using global pooling. For this purpose, attention grid mechanism is proposed to reduce the tensor size by max-pooling operation. Consider that the input size of image is $F_k \times H_x \times W_x$. If r number of max-pooling layers are employed, then image size will be reduced by the factor $2r \times 2r$ and new image size reduces to $F_g \times H_g \times W_g = F_k \times H_x/(2r) \times W_x/(2r)$. Attention map can be generated by either downsampling or upsampling the coarse feature to match spatial resolution of $\times^1$. By doing so, attention mechanism will be more flexible to have a focused region. Downsampling is done using bilinear techniques and downsampling is done using max-pooling.

3.1.2 Attention Module Backward Pass

In this section, we shall discuss, how signal will propagate in backward direction through attention module during learning process. In U-Net architecture, attention modules (gating process) are used just before the concatenations merge. In this case. The update rule for each parameter of (k-1)th layer can be obtained as follows.

$$\frac{\partial(\hat{x}_i^k)}{\partial(\Theta^{k-1})} = \frac{\partial(\alpha_i^k f(x_i^{k-1};\Theta^{k-1})}{\partial(\Theta^{k-1})} = \alpha_i^k \frac{\partial(f(x_i^{k-1};\Theta^{k-1})}{\partial(\Theta^{k-1})} + \frac{\partial(\alpha_i^k)}{\partial(\Theta^{k-1})} x_i^k$$

3.2 Convolutional Neural Network

For the segmentation, classification, and localization of biomedical images, the state-of-art methods are Convolutional Neural Networks (CNNs) [29–34]. These networks outperform the conventional method of medical image segmentation. The performance of these networks become superior due to following fact. (a) It allows the sharing of learned features across each pixel, (b) Convolution kernels exploit the structural information buried in medical images contaminated by noise, and finally (c) the image features are learned optimization techniques like stochastic gradient

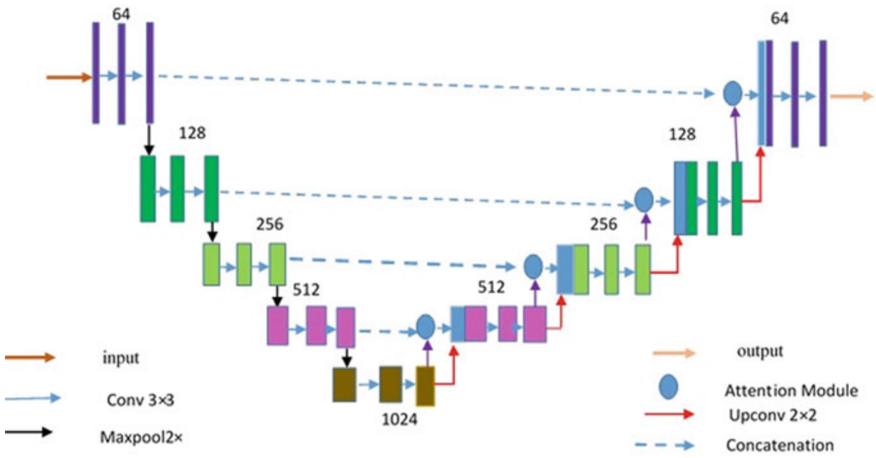


Fig. 2 U-NET with attention module

decent (SGD). Though all the measures are taken care of to have smaller false-positive predictions, it is a challenging task for the researchers for the image having highly imbalanced foreground–background ratios, especially in medical images. To overcome these problems, researcher approaches different methods. Soomro et al. [35] uses image enhancement techniques to reduce false-positive predication. Yan et al. used a three-stage deep learning model for the same purpose. In this work, attention module is incorporated with well-known UNet to improve the false-positive predication for blood vessel segmentation. Attention module reduces the use of multistage networks, thus reducing extra parameter learning as depicted in the proposed model in Fig. 2.

3.3 Segmentation in Attention Module

Attention module is used to segmentation of blood vessels of fundus images on top of the 2-D U-Net to get the advantage of using GPU. Image features are extracted in multiple image scales. Course feature maps contain the contextual information and highlighted the category of blood vessel pixels. During the deconvolution process, the extracted course feature maps are merged with finer-level dense prediction through skip connection. The attention module is incorporated into U-Net architecture to highlight the salient features which are passed through skip connection. The sigmoid activation function has been employed for normalization of attention module.

3.4 Loss Function

Loss function is key aspect of classification problem. The blood vessel segmentation task can be viewed as classification problem of two classes namely background class and foreground or vessel pixel class. The major problem of these medical images like retinal images has highly imbalanced the background to foreground ratio. Due to this, simple binary cross-entropy loss function does not work properly. To improve the performance, this paper used Tversky loss function, a hybrid of Focal loss and Tversky similarity index, which has been employed [36]. Proposed loss function is actually the combination of weighted binary loss function with Tversky similarity index.

Binary cross-entropy loss function with weight factor can be written as

$$L_{wbce} = \sum_{j} w_j y_j \log(t_j) + (1 - w_j)(1 - y_j)\log(1 - t_j) \tag{1}$$

where $w_j \in (0, 1)$ weight factor depends on foreground–background ratio, $y_j \in [0\ 1]$ is the predicted class of blood vessel and $t_j \in [0\ 1]$ the original ground truth pixel.

The Dice Score Coefficient (DSC) is the overlap index used for the medical image application which is represented as

$$\mathrm{DSC}_x = \frac{\sum_{i=1}^{N} y_{ix} g_{ix} + \delta}{\sum_{i=1}^{N} y_{ix} + g_{ix} + \delta} \tag{2}$$

where y_{ix} be the predicted class probability x and g_{ix} represent the ground truth probability of class x and belongs to (0, 1). N represents total number of pixels in the image and finally, δ is smoothing function to avoid zero in denominator [37]. The dice loss for the class c can be expressed as [38]

$$\mathrm{DL}_x = \sum_{x} 1 - \mathrm{DSC}_x \tag{3}$$

The main disadvantage of Dice Loss function is that it treats equally for false positive and false negative prediction which reduces the efficiency as most of the biomedical images have the highly class imbalance. To overcome this problem, Tversky similarity Index is introduced with two controlling parameters α and β.

$$\mathrm{TI}_x = \frac{\sum_{i=1}^{N} y_{ix} g_{ix} + \delta}{\sum_{i=1}^{N} y_{ix} g_{ix} + \alpha \sum_{i=1}^{N} y_{i'\bar{x}} g_{ix} + \beta \sum_{i=1}^{N} y_{ix} g_{i\bar{x}} + \delta} \tag{4}$$

Therefore, Tversky loss function [36] can be represents as

$$\mathrm{TL}_x = \sum_x 1 - \mathrm{TI}_x \tag{5}$$

To improve Tversky loss function for low-intensity pixels another control parameter γ has been introduced to get Focal Tversky loss function [39].

$$\mathrm{FTL} = \sum_x (1 - \mathrm{TI}_x)^{\gamma} \tag{6}$$

The γ takes the value in the range of (0 3). In this paper, gamma value has been taken as 1.33.

4 Experimental Setup

The U-Net in association with proposed Attention module was evaluated retinal blood segmentation task.

4.1 Database

Proposed algorithm has been evaluated on publicly available two datasets DRIVE and STARE. DRIVE data contains total of 40 images of size 565 × 584 and is divided into two parts. First part contains 20 images for training and second part contains 20 images for testing. Each image is provided with manual annotation of two evaluators. The training images are used to increase the number dataset by data augmentation algorithm. During this process, each image was resized to 512 × 512. Images of test datasets are further divided into two parts. First part contained 10 images and was used for validation purposes whereas second part contains remaining 10 images used for testing the algorithm. For this work, ground truth of first evaluator has been taken. Similar treatment has been adopted for the other publicly available datasets STARE.

4.2 Performance Matrices

To analyze the performance of proposed algorithm quantitatively following standard parameters are calculated.

$$\text{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Table 1 Analysis of performance of proposed method on database

Database	Se	Sp	ACC	AUC
Drive	0.872	0.984	0.965	0.980
Stare	0.824	0.971	0.963	0.974

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}$$

$$\text{AUC} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

where TP represents all the pixel that belongs to vessel marked in ground truth and also detected as vessel pixels, TN represents all the pixel that belongs to background pixel marked in ground truth and also detected as non-vessel pixels, FP represents all the pixel that belongs to non-vessel marked in ground truth but detected as vessel pixels and FN represents all the pixel which belongs to vessel marked in ground truth but detected as non-vessel pixels. Sensitivity indicates the ratio of correctly detected blood vessel pixels to the total number of vessel pixels which is the sum of TP and FN; whereas specificity represents the ratio of correctly classified non-vessel pixels to the number of background pixels which is the sum of FP and TN. The accuracy indicates the fraction of correctly classified pixels over entire images pixels. Finally, AUC represents the area under the curve of ROC which actually indicates the performance of the overall algorithm.

4.3 Result Evaluation

The performance of the proposed work has been represented in this section. The result of this work has been analyzed on two challenging databases DRIVE and STARE. Our obtained result is presented in Table 1. The table shows that the proposed algorithm achieves AUC of 0.980 and 0.974 and Accuracy of 0.965 and 0.963 respectively.

Further, the results have been analyzed for the comparison with other State-of-Art methods showing clear improvement in performance matrices. Comparison is shown in Table 2 and Fig. 3.

5 Conclusion

In this paper, attention module base U-Net has been employed for the purpose of retinal blood vessel segmentation. The DRIVE and STARE data sets were tested on

Table 2 Comparison with state of art with proposed work

Database	Drive				Stare			
Method	Se	Sp	ACC	AUC	Se	Sp	ACC	AUC
Soomro et al. [40]	0.739	0.956	0.848	0.844	0.748	0.962	0.947	0.855
Soomro et al. [35]	0.802	0.974	0.959	0.948	0.801	0.969	0.961	0.945
Wang et al. [33]	0.798	0.973	0.951	0.974	0.791	0.972	0.953	0.970
Yan et al. [41]	0.765	0.981	0.954	0.975	0.758	0.984	0.961	0.981
Sau et al. [42]	0.841	0.984	0.964	0.981	…	…	…	…
This work	0.872	0.984	0.965	0.980	0.824	0.971	0.963	0.974

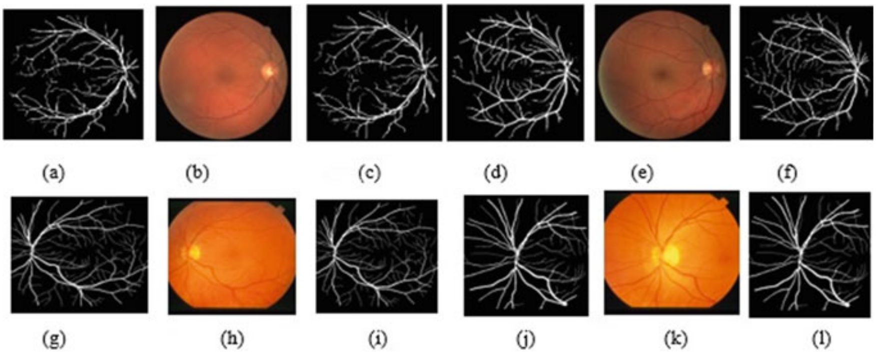


Fig. 3 Images of fundus images and with segmented results: **b** and **e** are original fundus from DRIVE datasets and **a**, **d** and **c**, **f** are their ground truth and segmented image respectively. Similarly, **h** and **k** are original fundus from STARE datasets, and **g**, **j** and **i**, **l** are their ground truth and segmented image respectively

the algorithm and obtained AUC as 0.980 and 0.974 respectively. The proposed algorithm, while applied on datasets, achieves accuracy as 0.965 and 0.963 respectively whereas specificity achieves as 0.984 and 0.971 respectively. Analysis of this result reveals that though it is able to detect the thin vessels but fails to detect very thin vessels of the span of one or two pixels. So attention is to be paid to additional prior knowledge to overcome this problem which may be the future task.

References

1. Kande GB et al (2010) Unsupervised fuzzy based vessel segmentation in pathological digital fundus images. J Med Syst 34(o. 5):849–858
2. Chakraborti T et al (2014) A self-adaptive matched filter for retinal blood vessel detection. Mach Vis Appl 26(1):55–68
3. Yang X et al (2014) Accurate vessel segmentation with progressive contrast enhancement and canny refinement. In: Proceedings of Asian conference on computer vision, pp 1–16

4. Roychowdhury S et al (2015) Iterative vessel segmentation of fundus images. IEEE Trans Biomed Eng 62(7):1738–1749
5. Yu Z, Sun K (2010) Vessel segmentation on angiogram using morphology driven deformable model. In: Proceedings of 3rd international conference on biomedical engineering and informatics (BMEI), vol 2, Oct 2010, pp 675–678
6. Tagizaheh M, Sadri S, Doosthoseini AM (2011) Segmentation of coronary vessels by combining the detection of centerlines and active contour model. In: Proceedings of 7th Iranian conference on machine vision and image processing (MVIP), Nov 2011, pp 1–4
7. Zhu T (2010) Fourier cross-sectional profile for vessel detection on retinal images. Comput Med Imaging Graph 34(3):203–212
8. Lam BSY, Hong Y (2008) A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields. IEEE Trans Med Imaging 27(2):237–246
9. Delibasis KK, Kechriniotis AI, Tsonos C, Assimakis N (2010) Automatic model-based tracing algorithm for vessel segmentation and diameter estimation. Comput Methods Programs Biomed 100(2):108–122
10. Li H, Zhang J, Nie Q, Cheng L (2013) A retinal vessel tracking method based on Bayesian theory. In: Proceedings of IEEE 8th conference on industrial electronics and applications (ICIEA), June 2013
11. Rodrigues J, Bezerra N (2016) Retinal vessel segmentation using parallel grayscale skeletonization algorithm and mathematical morphology. In: Proceedings of 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), Oct 2016, pp 17–24
12. Kumar SJJ, Ravichandran CG (2017) Morphological operation detection of retinal image segmentation. In: Proceedings of international conference on intelligent sustainable systems (ICISS), Dec 2017, pp 1228–1235. 232–235
13. Nisha KL, Sreelekha G, Savithri SP, Mohanachandran P, Vinekar A (2017) Fusion of structure adaptive filtering and mathematical morphology for vessel segmentation in fundus images of infants with retinopathy of prematurity. In: Proceedings of IEEE 30th Canadian conference on electrical and computer engineering (CCECE), Apr/May 2017, pp 1–6
14. Zhao Y, Rada L, Chen K, Harding SP, Zheng Y (2015) Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. IEEE Trans Med Imaging 34(9):1797–1807
15. Soares JVB et al (2006) Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. IEEE Trans Med Imaging 25(9):1214–1222
16. Lupascu CA et al (2010) FABC: retinal vessel segmentation using AdaBoost. IEEE Trans Inf Technol Biomed 14(5):1267–1274
17. Fraz MM et al (2012) An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans Biomed Eng 59(9):2538–2548
18. Fu H et al (2016) Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In: Proceedings of IEEE 13th international symposium on biomedical imaging, pp 698–701
19. Fu H et al (2016) Deep vessel: retinal vessel segmentation via deep learning and conditional random field. In: Proceedings of medical image computing and computer-assisted interventions conference, pp 132–139
20. Wang C, Zhao Z, Ren Q, Xu Y, Yu Y (2019) Dense U-net based on patch-based learning for retinal vessel segmentation. Entropy 21:168
21. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Conference on advances in neural information processing systems, pp 1097–1105
22. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: ICLR
23. Szegedy C, Liu W, Jia Y, Sermanet Y, Reed P, Anguelov S, Erhan D, Vanhoucke D, Rabinovich V (2015) Going deeper with convolutions. In: CVPR
24. Wang Y, Jodoin P-M, Porikli F, Konrad J, Benezeth Y, Ishwar P (2014) Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) workshops, pp 387–394

25. Brosch T, Yoo Y, Tang LY, Li DK, Traboulsee A, Tam R (2015) Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 3–11
26. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: MICCAI. Springer, Berlin, pp 234–241
27. Milletari F, Navab N, Ahmadi SA (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth international conference on 3D vision (3DV), Oct 2016. IEEE, pp 565–571
28. Luong MT, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:150804025
29. Yang D, Ren M, Xu B (2019) Retinal blood vessel segmentation with improved convolutional neural networks. J Med Imaging Health Inf 9(6):1112–1118
30. Guo S, Wang K, Kang H, Zhang Y, Gao Y, Li T (2019) BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation. Int J Med Inf 126:105–113
31. Son J, Park SJ, Jung K-H (2019) Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. J Digit Imaging 32(3):499–512
32. Hu K, Zhang Z, Niu X, Zhang Y, Cao C, Xiao F, Gao X (2018) Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. Neurocomputing 309:179–191
33. Wang C, Zhao Z, Ren Q, Xu Y, Yu Y (2019) Dense U-net based on patch-based learning for retinal vessel segmentation. Entropy 21(2):168
34. Dharmawan DA, Li D, Ng BP, Rahardja S (2019) A new hybrid algorithm for retinal vessels segmentation on fundus images. IEEE Access 7:41885–41896
35. Soomro TA et al (2019) Impact of image enhancement technique on CNN model for retinal blood vessels segmentation. IEEE Access 7:158183–158197. https://doi.org/10.1109/ACCESS.2019.2950228
36. Hashemi R, Mohseni Salehi SS, Erdogmus D, Prabhu SP, Warfield SK, Gholipour A (2018) Tversky as a loss function for highly unbalanced image segmentation using 3d fully convolutional deep networks. CoRR abs/1803.11078
37. Wong KCL, Moradi M, Tang H, Syeda-Mahmood T (2018) 3d segmentation with exponential logarithmic loss for highly unbalanced object sizes. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 612–619
38. Milletari F, Navab N, Ahmadi S-A (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth international conference on 3D vision (3DV). IEEE, pp 565–571
39. Zhu W, Huang Y, Tang H, Qian Z, Du N, Fan W, Xie X (2018) Anatomynet: deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation. arXiv preprint arXiv:1808.05238
40. Soomro TA, Hellwich O, Afifi AJ, Paul M, Gao J, Zheng (2018) Strided U-Net model: retinal vessels segmentation using dice loss. In: Proceedings of digital image computing: techniques and applications (DICTA), pp 1–8
41. Yan Z, Yang X, Cheng K-T (2018) Joint segment-level and pixelwise losses for deep learning based retinal vessel segmentation. IEEE Trans Biomed Eng 65(9):1912–1923
42. Sau PC et al (2020) Retinal blood vessel segmentation using focal Tversky loss function and U-Net. Int J Adv Sci Technol 29(7):3693–3702

Performance Evaluation of 2D CNN Optimizers for Lung and Colon Cancer Image Classification

Adnan Zafar and Mohammad Nadeem

Abstract The paper presents a convolutional neural network (CNN) based approach to classify lung cancer and colon cancer image dataset using two popular optimization algorithms: Adam and RMSprop. The lung cancer and colon cancer are very common among cancer-related diseases as around 25% of cancer patients suffer from either one. Survival rate may be increased if we identify the cancer in early stages. Deep learning, a field of machine learning has revolutionized the field of medical science. In the present study, separate models are built for the lung cancer and colon cancer using CNN to predict the class of the disease more accurately. The image dataset taken was available in the form of histopathological images. Paper consists of four models lung cancer using Adam, lung cancer using RMSprop, colon cancer using Adam, colon cancer using RMSprop, and performance (accuracy) of Adam and RMSprop is compared. The results are particularly useful for cancer prediction from images and can be generalized to other image classification problems as well.

Keywords Convolutional neural network · Deep learning · Lung cancer · Colon cancer · Optimizer

1 Introduction

Lung cancer is the most lethiferous cancer touching all over the world. All over the world lung cancer causes 19% of cancer-related death [1]. Lung cancer arises when cells split up in the lungs unmanageably. It will help tumor to spread. When lung cancer occurs, it reduces the ability to breathe. Colon cancer is also a vital cancer that is affecting the world in a very rapid manner. It starts from the large intestine. Colon cancer is also known as colorectal cancer. The term colorectal cancer is formed by combining the term colon cancer and rectal cancer. The colon is the concluding part of digestive region. Data collected from 2015 to 2017, the American citizens from every five, two are diagnosed during their lifetime.

A. Zafar (✉) · M. Nadeem
Department of Computer Science, AMU, Aligarh, UP 202002, India
e-mail: aligadnanzafar@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_42

Various researchers have worked on cancer image datasets and still, the studies are going on. That is why nowadays we are in a position to predict the lung cancer and colon cancer more accurately so that we can save lives of many people if cancer is predicted in early stages. LUNA [2] made available the data which carries 888 CT scans of lungs for the use of researchers. Another type of data that is used in research is histopathological image which is publically available on kaggle [3]. LIDC-IDRI [4] in the collaboration of several academic centers and medical imaging companies created a dataset containing 1018 cases. These datasets are used in deep learning for making the prediction of the disease. This shows that deep learning is playing a vital role in the field of medical sciences to enhance the safety of the people.

As deep learning is the sub-branch of machine learning. The idea behind deep learning is to build learning algorithms that mimic a human brain. It contains several types of deep neural networks out of which we have used convolutional neural network (CNN).CNN is a feed-forward neural network [5]. A huge repute achieved by Convolutional neural networks (CNN) in the field of image classification in last recent years. The convolutional neural network is usually used for classification problems. We will follow CNN technique to classify the lung cancer and colon cancer in this paper. Deep learning is very effective image classification technique that helped a lot in the field of medical science [6].

The learning rate technique RMSprop is suggested by Geoff Hinton. The learning rate in RMSprop is divided by an exponentially decomposing average of squared gradients in RMSprop. 0.001 is the best default leaning rate while using RMSprop suggested by Geoff Hinton. Another learning technique is ADAM, it also keeps the data of exponentially decomposing average of squared gradients which are already finished as done by RMSprop. It also has exponentially decomposition of already done gradients. It works better in comparison to other learning methods.

Since 1955, the computers have been able to do medical diagnoses, proposed by Lusted [7]. The first computer-aided diagnosis is applied to diagnose lung cancer by Lodwick et al. [8]. This research keeps on going some researchers worked on Support Vector Machine classification [9], some worked on Artificial Neural Network classification [10]. Some researchers also worked on convolutional neural networks with a great accuracy [11]. In last some years with the unbroken development of technology in the field of medical sciences, the study and the care of lung cancer has given a great advancement to medical sciences. Today low dose (LD-CT), is a powerful approach broadly used for early tracking of lung cancer, and on the other hand, the radiation dose is reduced and image quality is improved by IMR based full iterative CT recombination technology [12]. Researchers of paper [13] and some other also presented complete iterative CT recombination technique based on IMR cannot decrease the dose of radiation, but noise ratio also decreases with respect to signal of images. At today's time, sparse literature on the benefit of IMR in lung cancer has been outlined at home and widely. This study used the combination of LD-CT and IMR approach for the evaluation of peripheral lung cancer. In ULD-CT, the image quality is compared with IMR, PIR and FBR [14]. The low dose spiral CT examination is found to be more accurate in the determination of peripheral lung cancer and simple to execute, and can get the idea of peripheral lung cancer timely

[15]. The experiment in paper [16] shows an important effect of single hole Thoracic lobectomy for the cure of lung cancer. In paper [17] researchers worked on CT scan images of lung cancer and got an accuracy of 84% giving best in comparison of previous results.

Guidelines play a major role in curing and minimizing the outcome. Comparisons are made between western and Asian guidelines for colon cancer management. Surgical method for colon cancer is different in the guidelines of western and Asian countries. This field requires further research. Another work [18] compares the diagnostic performance between MRI scan and CT scan and reached the conclusion that MRI presents higher performance than CT scan for detecting EMVI, that is why CT is an alternate application in local stage of colon cancer. Researchers of paper [19] determine the accuracy of CT staging in colon cancer in the determination of unfavorable confirmed factors. But this study failed in the detection of 19.3% of primary cancer and researchers concluded that CT scan for verification of colon cancer is not sufficient.

Proposed work tries to answer which optimization algorithm works better for cancer image classification. The results are particularly useful for cancer prediction from images but can be generalized to other image classification problems as well. The rest of the paper is organized as follows. Section 2 provides a brief summary of related works. Section 3 describes the dataset used in present study. The methodology adopted is outlined in detail in Sect. 4. Section 5 presents the results of the study and concluding remarks are put in Sect. 6.

2 Related Work

Convolutional neural network is a powerful clustering technique that divides the data into their respected groups. They are complex feed-forward neural networks. Because of its high accuracy convolutional neural networks are used in image classification. Some researchers also worked on support vector machine generally called SVM which is a supervised classification technique for dividing the data into two groups. The technique used by SVM is kernel trick which transforms the given data. It simply draws a hyper-line or hyper-plane between two different groups. The received precision is about 73.7% using SVM classifier which is applied to histopathological image dataset of colon cancer in the paper [9] and predicts colon cancer.

In paper [11] researchers achieved an accuracy of 96.33% using convolutional neural network which is applied to histopathological image dataset. They made a model which predicts whether person has lung cancer and colon cancer or not. A computed tomography scan usually known as CT scan is a medical imaging procedure help radiology to acquire detail of the body image and CT scan can display cancer nodule even they are small. Using CT Scan image dataset the researchers of the paper [15] got an accuracy of 93.9% with CNN approach to get whether a person has lung cancer or not. Paper [20] also represents a CNN approach to predict lung cancer with an accuracy of 97.2%, in that lung dataset of histopathological image dataset is used.

If anyone uses a pre-trained model then it is a very effective approach to work RSnet-50. RSnet-50 contains 50 layers deep convolutional neural network in which pre-trained version of network is used. Researchers in paper [21] have worked on colon dataset of histopathological image dataset using RSnet-50 which gives an accuracy of 93.91% for making the prediction of colon cancer. Researchers of paper [22] make a comparative study of colon cancer classification in order to get maximum prediction accuracy and to serve the medical science. The paper [23] shows an approach to detect and classify nuclei in routine colon cancer with the use of histology images.

3 Dataset Description

Researchers assembled data used in the paper and data became available from kaggle with the title LC25000 Lung and colon histopathological image dataset [3]. There are two subfolders in our dataset colon_set contains colon_aca and colon_n and lung_set comprises lung_aca, lung_n, lung_scc. So there are total five classes colon_n, colon_aca, lung_aca, lung_n, lung_scc containing 5000 images to each class. The pixel of each image is 768 × 768 pixels and the format of each image is jpeg. There are 15,000 cancerous images (colon_aca, lung_aca, lung_scc) and 10,000 non-cancerous images (colon_n, lung_n). Sample images of dataset are given in Fig. 1. Explanation and the distribution of data are given in Table 1.

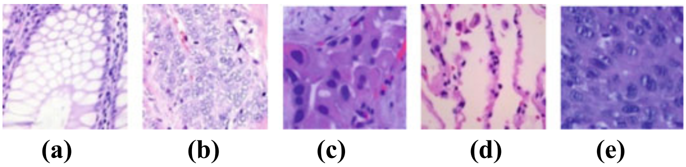


Fig. 1 Sample images of **a** colon benign tissue, **b** colon adenocarcinoma, **c** lung adenocarcinoma, **d** lung benign tissue, **e** lung squamous cell carcinoma

Table 1 Distribution of images into categories with description

Data	Description	Number of images
Colon_n	Colon benign tissue	5000
Colon_aca	Colon adenocarcinoma	5000
Lung_scc	Lung squamous cell carcinoma	5000
Lung_aca	Lung adenocarcinoma	5000
Lung_n	Lung benign tissue	5000

4 Methodology

In this paper, we have four sections that mean we made four different models that are LUNG cancer using ADAM, Lung cancer using RMSprop, Colon cancer using ADAM and Colon cancer using RMSprop. The paper predicts whether a person is cancerous or not. The proposed model is divided into several stages as given below:

4.1 Pre-processing

It is also known as rescaling images into needed size. The used data set have the image size 768 × 768 pixels which are transformed to 256 × 256 pixels. There is a need to transform our dataset to remove noise, gather useful information, and increase the performance, pre-processing is a necessary task [24]. OpenCV package was used for pre-processing.

4.2 Data Division

After pre-processing we divide our dataset into three categories: train, validate and test. After that, we start constructing model and compiling the model. The next phase is to train with different epochs for different models. The epochs size and used optimizer table are given in Table 2. Further, we tried to visualize the accuracy and loss of training dataset and validation dataset. We saved the model so that the model can be used in further estimation of cancer or to test cancer. The architecture of the used model is given in Fig. 2.

As Fig. 2 shows, total of five layers/channels is used that are conv2D layer, batch-normalization layer, max-pool layer, dropout layer. The used kernel size, input shape, pool size, and dropout size are (3, 3), (256, 256, 3), (2, 2) and 0.2 respectively.

Table 2 Number of epochs used for each dataset for two optimizers

Data	Optimizer	No of epochs used
Colon_set	Adam	50
Colon_set	RMS-prop	50
Lung_set	Adam	100
Lung_set	RMS-prop	100

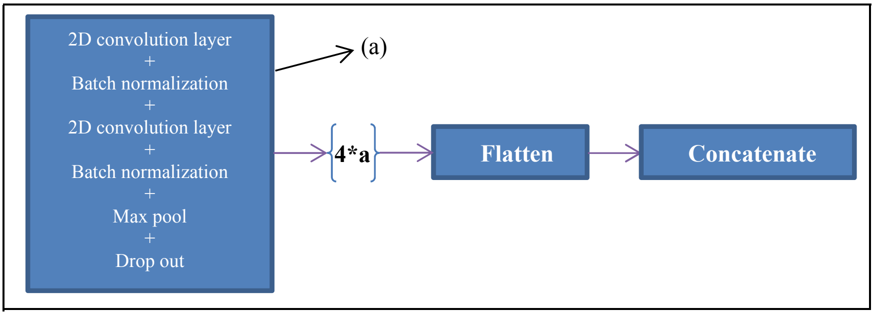


Fig. 2 The architecture of the used model

5 Result and Discussion

This project is done using histopathological image dataset. In this dataset, there are 25000 images (cancerous and non-cancerous). Kaggle has this dataset openly available [3]. The dataset was divided into two sets that are colon set and lung set. The percentage of training, validation, and testing is 80%, 16%, and 4% respectively. For results, we used different optimizers ADAM and RMSprop which are applied on both colon set and lung set separately. And we make a comparison between the accuracy generated by two different optimizers. The results can be generalized to other image classifications using CNN network with the most effective optimizer means Adam will give better results in comparison to RMSprop.

5.1 Obtained Accuracies

Accuracy is the method for the exploration of clustering models. Generally, accuracy is defined as the ratio of correct prediction made by the model to total number of predictions. The blue line represents the training accuracy and orange line represents validation accuracy. Figure 3 represents the testing accuracies of all the models. For simplicity, we create a Table 3 of accuracy so that it became easy to understand the results of accuracy. Histogram in Fig. 4 represents various accuracies.

5.2 Loss Values

Loss is the punishment for making a wrong prediction. If model predicts correctly then loss is zero otherwise loss has some value. The blue line represents the training loss and orange line represents validation loss. Figure 6 represents the graph of loss.

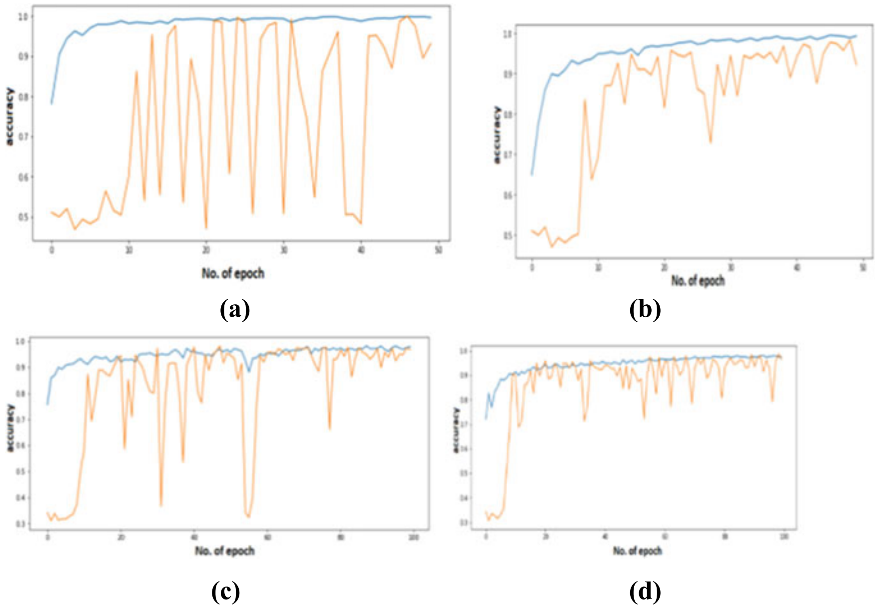


Fig. 3 Training accuracy and validation accuracy plots; **a** colon cancer using ADAM, **b** colon cancer using RMSprop, **c** lung cancer using ADAM, **d** lung cancer using RMSprop

Table 3 The accuracies obtained on lung and colon cancer datasets

Cancer	Optimizer	Training accuracy (%)	Validation accuracy (%)	Testing accuracy (%)
Colon	Adam	99.92	93.12	95.39
Colon	RMS-prop	99.55	92.19	93.22
Lung	Adam	98.35	96.88	96.83
Lung	RMS-prop	97.23	96.56	94.36

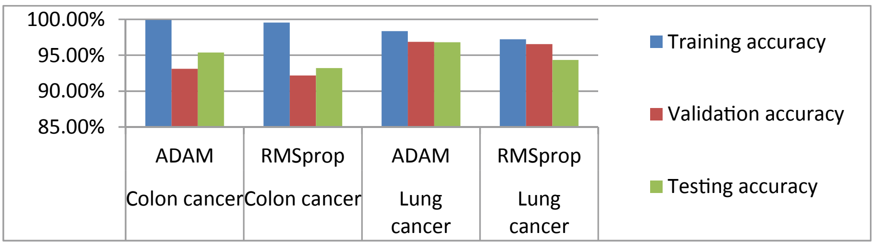


Fig. 4 Graphical representation of obtained accuracies

Table 4 Loss values for lung and colon cancer datasets

Cancer	Optimizer used	Loss value (%)
Colon	ADAM	27.75
Colon	RMSprop	33.55
Lung	ADAM	9.60
Lung	RMSprop	18.12

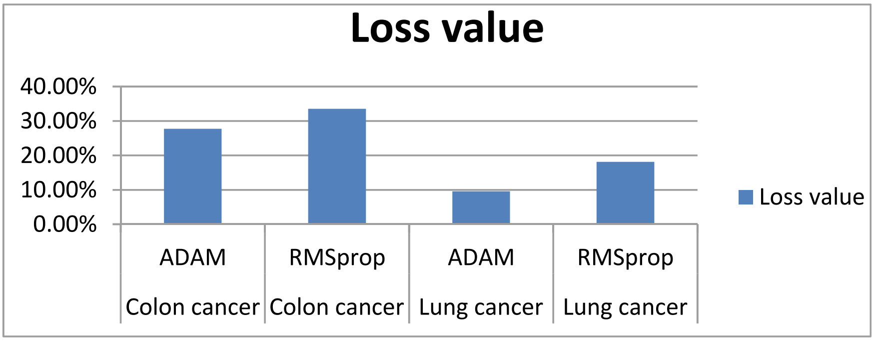


Fig. 5 Graphical representation of loss value obtained

For simplicity, the loss is given in Table 4 and also in Fig. 3. Histogram in Fig. 5 represents testing loss.

It can be observed easily that ADAM optimization algorithm performed significantly better than its counterpart in correctly classifying the cancer image datasets. Also, the difference in performance is more evident in loss values. As shown in Table 4, ADAM optimizer is almost 50% more efficient than RMSprop. Given these results, it can be said that ADAM algorithm is a better choice to be incorporated in a 2D convolutional neural network for classifying cancer image datasets. Also, after careful examination, the results can be extended to other image datasets.

In [9], total accuracy of 73.7% was there. In paper [11] researchers got an accuracy of 96.33%. The paper [15] results in total accuracy of 93.9%. In article [16] an accuracy of 96.2% is received. Paper [20] resulted in an accuracy of 97.2%. Research of paper [21] gives an accuracy of 93.91%. And in this paper we got an accuracy of approximately 99% and compare the optimizers' accuracy which is not done by any other research will help in future for generalized image classification of any of the available datasets.

6 Conclusion

The study presents a deep learning approach to classify images, the approach is known as CNN approach. Here we used histopathological image dataset which is

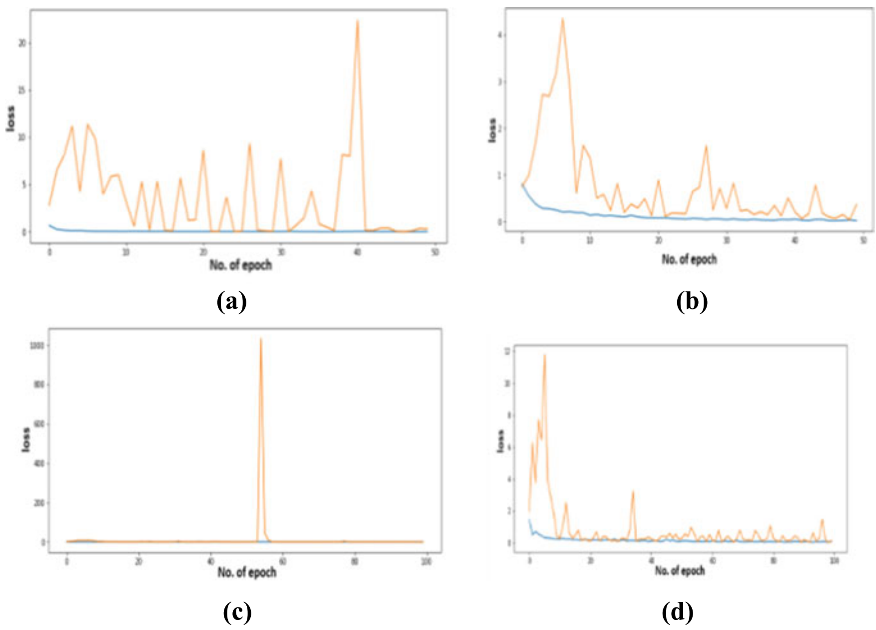


Fig. 6 Training loss and validation loss plots; **a** colon cancer using ADAM, **b** colon cancer using RMSprop, **c** lung cancer using ADAM, **d** lung cancer using RMSprop

openly available on kaggle [3]. The training accuracy given by the models is 99.92%, 99.55%, 98.35%, and 97.23% for colon using Adam, colon using RMSprop, lung using Adam, and lung using RMSprop respectively. A comparison between optimizers became successful. Here we got better results by using Adam optimizer in comparison to RMSprop optimizer. Hence we concluded that Adam optimizer is more efficient than RMSprop optimizer for making the prediction of cancer as our research shows that loss is also less in Adam optimizer. This work will help us in future to determine lung and colon cancer more accurately and the optimizers which gave better accuracy must be used in future in the models which we are using. In future we can make more powerful and effective models using 3D-CNN, YOLO, and many other techniques which can be applied to different image datasets.

References

1. Malik PS, Raina V (2015) Lung cancer: prevalent trends & emerging concepts. Indian J Med Res 141(no. Jan 2015, Supplement):5–7. https://doi.org/10.4103/0971-5916.154479
2. Data - LUNA16 - Grand Challenge. https://luna16.grand-challenge.org/Data/. Accessed 06 June 2021
3. Lung and Colon Cancer Histopathological Images | Kaggle. https://www.kaggle.com/andrewmvd/lung-and-colon-cancer-histopathological-images. Accessed 30 Apr 2021

4. LIDC-IDRI - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI. Accessed 06 June 2021
5. Liu T, Fang S, Zhao Y, Wang P, Zhang J (2015) Implementation of training convolutional neural networks [Online]. Available: http://arxiv.org/abs/1506.01195
6. Shen D, Wu G, Suk H-I (2017) Deep learning in medical image analysis 19:221–248. https://doi.org/10.1146/ANNUREV-BIOENG-071516-044442
7. LB Lusted 1955 Medical electronics N Engl J Med 252 14 580 585 https://doi.org/10.1056/NEJM195504072521405
8. GS Lodwick TE Keats JP Dorst 1963 The coding of roentgen images for computer analysis as applied to lung cancer Radiology 81 185 200 https://doi.org/10.1148/81.2.185
9. Y Xu 2013 Multi-label classification for colon cancer using histopathological images Microsc Res Tech 76 12 1266 1277 https://doi.org/10.1002/jemt.22294
10. SA Deppen 2014 Accuracy of FDG-PET to diagnose lung cancer in areas with infectious lung disease: a meta-analysis JAMA J Am Med Assoc 312 12 1227 1236 https://doi.org/10.1001/jama.2014.11488
11. Masud M, Sikder N, Al Nahid A, Bairagi AK, Alzain MA (2021) A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors (Switzerland) 21(3):1–21. https://doi.org/10.3390/s21030748
12. H Kim CM Park YS Song SM Lee JM Goo 2014 Influence of radiation dose and iterative reconstruction algorithms for measurement accuracy and reproducibility of pulmonary nodule volumetry: a phantom study Eur J Radiol 83 5 848 857 https://doi.org/10.1016/j.ejrad.2014.01.025
13. H Yuki 2016 Clinical impact of model-based type iterative reconstruction with fast reconstruction time on image quality of low-dose screening chest CT Acta Radiol 57 3 295 302 https://doi.org/10.1177/0284185115575537
14. X Huang J Jia Y Li Z Wang 2016 Complex nonlinear dynamics in fractional and integer order memristor-based systems Neurocomputing 218 296 306 https://doi.org/10.1016/j.neucom.2016.08.078
15. S Suresh S Mohan 2020 ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis Neural Comput Appl 32 20 15989 16009 https://doi.org/10.1007/s00521-020-04787-w
16. Shakeel PM, Burhanuddin MA, Desa MI (2020) Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. Neural Comput Appl 1–14. https://doi.org/10.1007/s00521-020-04842-6
17. S Bhatia Y Sinha L Goel 2019 Lung cancer detection: a deep learning approach Adv Intell Syst Comput 817 699 705 https://doi.org/10.1007/978-981-13-1595-4_55
18. Liu LH, Lv H, Wang ZC, Rao SX, Zeng MS (2019) Performance comparison between MRI and CT for local staging of sigmoid and descending colon cancer. Eur J Radiol 121:108741. https://doi.org/10.1016/j.ejrad.2019.108741
19. LM Fernandez 2019 How reliable is CT scan in staging right colon cancer? Dis Colon Rectum 62 8 960 964 https://doi.org/10.1097/DCR.0000000000001387
20. BK Hatuwal HC Thapa 2020 Lung cancer detection using convolutional neural network on histopathological images Int J Comput Trends Technol 68 10 21 24 https://doi.org/10.14445/22312803/ijctt-v68i10p104
21. Khalid Bukhari SU, Syed A, Arsalan Bokhari SK, Hussain SS, Armaghan SU, Hussain Shah SS (2020) The histological diagnosis of colonic adenocarcinoma by applying partial self supervised learning. https://doi.org/10.1101/2020.08.15.20175760
22. Ben Hamida A et al. (2021) Deep learning for colon cancer histopathological images analysis. Comput Biol Med 136:104730. https://doi.org/10.1016/J.COMPBIOMED.2021.104730
23. K Sirinukunwattana SEA Raza YW Tsang DRJ Snead IA Cree NM Rajpoot 2016 Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images IEEE Trans Med Imaging 35 5 1196 1206 https://doi.org/10.1109/TMI.2016.2525803

24. Sudeep KS, Pal KK (2017) Preprocessing for image classification by convolutional neural networks. In: 2016 IEEE International conference on recent trends in electronics, information and communication technology, RTEICT 2016—Proceedings, Jan 2017, pp 1778–1781. https://doi.org/10.1109/RTEICT.2016.7808140

A Research Study on Brain Tumor Detection Techniques

Nisha Joseph, D. Murugan, and Divya Mohan

Abstract Even though there are remarkable advancements in brain tumor detection in medical technology, it remains the most tedious and complicated issue for doctors. For brain tumor detection, radiologists commonly use Magnetic Resonance Imaging (MRI). The MRI provided both the standard and abnormal anatomy of the brain. Also, MRI does not require dissection. Computer-aided image analysis is the promising solution for detecting the diseases such as a tumor, cancer earlier stages. Automatic computerized classification for detecting the tumor from the MRI image is essential. It will help detect the tumor, reduce the number of hours worked, lessen errors, and help decide the treatment plan. This work depicts the techniques proposed in contemporary literature by briefing the novel facts of the research.

Keywords Brain tumor · Automatic detection · Magnetic resonance imaging · Segmentation · Classification accuracy

1 Introduction

In the human body, the brain is an important organ. The functions of the brain are still partially known to us. So, if there exists an abnormality in it, the effects of variations are unknown too and keep varying from person to person. The anomaly which is unknown to us is called a tumor. According to the survey taken brain tumor leads to the highest death rate in the world. The brain tumor symptoms are hormone changes, blood clotting, loss of vision, weakness, mood swings, etc. The two phases in tumor are benign tumors and malignant tumors. When surgeons remove benign tumors from the human body, there will be no threat. While malignant tumor comes back again when it is removed and hence classified as cancerous. The prediction

N. Joseph (✉) · D. Murugan
Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu 627012, India
e-mail: nishajoseph226@gmail.com

D. Mohan
Department of Computer Science and Engineering, Satyabhama University, Chennai, Tamil Nadu, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_43

of a patient-facing tumor is to detect and classify the tumor in emerging stages. It enabled researchers to propose the detection and classification of tumors in their works in recent years using image processing technology advancements. Overall, many techniques are involved in detecting and categorizing the tumor, namely, Image segmentation, histogram equalization, feature selection, extraction and classification, image enhancement, and many more.

2 Brain Tumor Detection

The steps involved in tumor classification are illustrated in Fig. 1. The important steps involved after image acquisition are preprocessing, segmentation, feature extraction, and classification.

A. Image Preprocessing

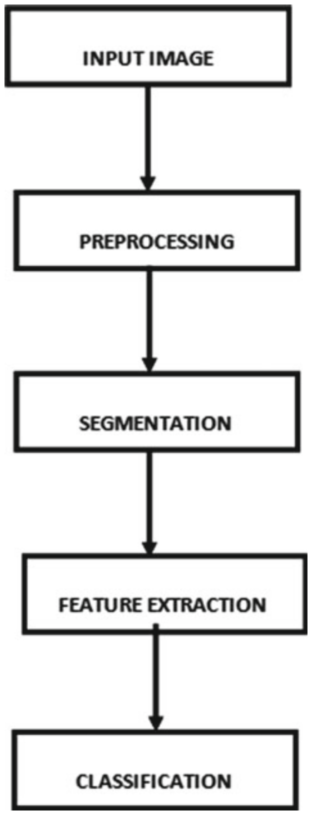


Fig. 1 Work flow of brain tumor detection system

Digital Image preprocessing is an essential process in all image-based applications. We preprocess to ensure the following.

1. It is higher-level processing, which will be helpful for segmentation and feature extraction.
2. We use it to remove the unwanted portion such as the marks, labels, and the noise present in the image, which may affect the classification.
3. For enhancing the quality of the image.
4. To remove all noises present.

B. **Image Segmentation**

The researchers segment the image for splitting an image into regions to extract the area of interest. We do the standard to separate the image's essential details from the remainder of the image to be observed or recognized as objects separately. The three main categories in segmentation are pixel-based, edge-based, and region-based methods. The researchers classify the pixel-based into thresholding and clustering.

C. **Feature Extraction**

Nassiri et al. [1] highlight extraction of the deterioration of images centered on the unique characteristics used to classify various patterns. The researchers perform dimensionality reduction in the feature extraction process, as they can use it to find the exciting parts of the image as a precise feature vector. This helps handle large databases. So the feature vector size can be reduced, which enables the speedy completion of image matching and retrieval.

D. **Classification**

The researchers classify to categorize the type of tumor into its respective class. There are mainly two labels, namely the benign tumor and the malignant tumor. They choose the classifiers depending upon the type of problem to be solved. Some of the classifiers are Support Vector Machine classifier, Neural network, K-Nearest neighbor classifier, Bayesian classifier, and many more.

3 Materials and Methods Used for Literature Survey

Zhang et al. [2] introduced cancerous brain detection which uses wavelet entropy, which recognizes neurotic brains from ordinary brains with MRI help. The Eigenvalue Proximal Bolster Vector Machine (GEPSVM) classifier is, for the arrangement and also used to implant the RBF portion.

Wang et al. [3] suggested a scheme on neural Networks for Network-based intrusion detection. They used computerized images for the investigation and based the survey on prior procedures' quality and restrictions. The MR brain images are too an issue; HPA Technique achieves the more productively for, the more classification strategies.

Zhou et al. [4] introduced a process for recognizing brain tumors using Magnetic Resonant Images. They analyzed the curative treatment of patients. A computer-based technique with wavelet entropy for the component space approach and enhancing mind determination for the NMR images. They used the Naive Bayes classifier arrangement strategy.

Laishram et al. [5] proposed a method for improving the magnetic resonance image. They used this approach for upgrading human brain images' hidden features captured by an MRI. At first, in the edge identification approach, image segmentation was done by Particle Swarm enhancement consolidating Fuzzy C-Means Clustering (PSOFCM) and at the second stage for recognizing the delicate edges. They did the edge discovery algorithm. It concludes that PSOFCM based vigilant edge location strategy produces better results than the shrewd edge recognition.

Oo and Khaing [6] suggested a system for detecting tumors using MRI. The researchers performed the Skull stripping, tumor region count, brain image test, tumor area assurance, image shifting, and morphological operation in this proposed technique. A morphological disintegration algorithm identified the tumor. For the separation of tumor from brain image, preprocessing step plays an essential role for fragmentation. They found the volume of the tumor to get the correct consequence of the tumor region. The researchers used the Frustum model to find the tumor volume.

Ibrahim et al. [7] introduced a scheme for classifying brain tumors using a neural network. There are three phases in the proposed neural network method: preprocessing, dimensionality reduction, and ordering. The scientists used backpropagation neural network classifier to assemble the inputs as normal or anomalous brain images.

Islam et al. [8] suggested a method for brain tumor classification. Brain tumor surface uses a multi resolution-fractal because of its intricate appearance in MRI called the multi-fragmentary Brownian movement (mBm), followed by a delivery multifractal brain tumor division methodology. Besides, the AdaBoost algorithm uses the weights to segment classifiers.

Sivaramakrishnan and Karnan [9] introduced a scheme that uses fuzzy algorithm. Tumor analysis was a pivotal task. In this strategy, the researchers did a brain's practical recognition of tumor by fuzzy C-means and histogram. This work provides good results.

In 2013, Ali et al. [10] advised a scheme for brain tumor detection using clustering. Morphological operations are used to eliminate the affected area in brain.

Patil SB et al. [11] suggested a scheme to remove the affected part of the brain from the Magnetic Resonant scan image, and area of the tumor is also measured. They performed Global threshold and watershed-based segmentation where they group the pixels only based on intensity. The segmentation output looks white color patch on the affected part of the brain.

Machhale et al. [12] suggested a system for recognizing cancerous brain images. Brain tumor treatment is usually given based on the patients' radiological appearance.

Xiao et al. [13] proposed estimating lateral ventricular deformation (LaV) and tumor features. The scientists applied the extracted highlights for segmenting tumors from the MR images. They used deformation of the lateral ventricle (LaV) for attribute removal. The researchers further used autonomous segmentation strategies to evaluate the lateral ventricular deformation feature. This segmentation technique wrongly allocates the pixel of non-CSF to the CSF bunch. Finally, in the technique, a worldwide mask is applied to remove the unwanted pixel.

Gopal and Sukanesh [14] proposed a technique by combining wavelet co-occurrence texture features and statistical features. The researchers used a vector machine classifier to aid in categorizing the tumor. Kalbkhani et al. [15] proposed a GARCH (Generalized Auto Regressive Conditional Heteroscedasticity), statistical model, to classify MR images into its category as normal and abnormal.

Sindhumol et al. [16] introduced Spectral Clustering Independent Component Analysis (SCICA) for improving the classification of brain tumors based on angle-based feature extraction from Magnetic Resonance Image. Because of its spectral distance, the MR image is split into various clusters. They analyzed the Independent component to examine the data that are clustered and performed the classification of SVM.

Navarro et al. [17] gave a new procedure for reducing dimensions based on feature selection and classifiers on various HMRS modalities. The entropy selection algorithm quickly generates relevant subset of spectral frequency.

Sumitra and Saxena [18] proposed a method with neural network. PCA is employed for extracting features. Mean, median, difference, interrelationship values having the highest and lowest intensity values are the elicited features. In this method, the type of tumor is grouped into normal, non-malignant, and malignant. The classification accuracy of this proposed technique is 73%.

Jayachandran and Dhanasekharan [19] introduced a technique for detecting brain tumors using computative classifiers that uses the hybrid algorithm. At first, they reduced the noise, and later they performed attribute extraction, feature reduction, and categorization. They used an anisotropic filter to reduce the noise. In this work, FSVM classifier is used to categorize the input as a standard or atypical image. This technique provides a classification accuracy of 92.80%.

Padma Nanthagopal and Sukanesh Rajamony [20] suggested a computerized software system for tumor identification. They used SVMto choose the maximum repeated values and synchronized appearance of the wavelet estimation tumor location. This technique provides a classification accuracy of 92.90%.

4 Observations

Out of several methods studied for classification, neural networks and support vector machines are commonly used. Since these classifiers can process the image faster and provide the classification results within a low computation time than the other methods, we chose it for classification purposes. Simultaneously, Shree and Kumar [21] provided the best classification accuracy, having 100% accuracy for all modalities. Table 1 shows the survey of existing techniques and their accuracy details.

Table 1 Survey of the existing work techniques and its accuracy

Author name	Technique	Modalities of input	Accuracy (%)
Shree and Kumar [21]	Noise removal, DWT, PNN classifier	All modalities	100
Mohsen et al. [22]	DWT feature extraction method	T2-weighted	96.9
Mathew and Babu Anto [23]	DWT and SVM	T1, T2, and FLAIR	86
Shil et al. [24]	DWT and SVM	T2-weighted	100
Usman and Rajpoot [25]	Random forest classifier, DWT	T1c	88
Amin et al. [26]	SVM	T1	97.1
Cheng et al. [27]	GLCM, BoW, SVM	T1-weighted	91
Kermi et al. [28]	The region and boundary based segmentation method	T1, T1C, T1-weighted, T2-weighted, FLAIR, T2	83.04 for FLAIR 93.63 for T2
Bahdure et al. [29]	BWT and SVM	T2-weighted	96.51
Praveen and Agrawal [30]	SVM, multi-layer perceptron	T1-weighted	96.6
Kaur et al. [31]	CEEMDAN, Hilbert transformation method	T1 and T1C	100
Sornam et al. [32]	Wavelet, Zernike extraction, ELM algorithm	T1 and T2-weighted	72

5 Conclusion

The advancement and automation in machine learning are increasing nowadays and provide great attention to brain tumor detection systems' research. The automated systems will be advantageous in the medical research field. Hence this survey is taken from the most recent research work about automated systems and provides a review of tumor segmentation with classification. Also, it recognizes tumor region in the brain. From the survey taken, we found that there are still many issues and detachment in the medical field, and also there are issues regarding the computation time and robustness. In conclusion, most of the work is carried out based on SVM and NN classification techniques for recognizing brain tumors.

References

1. Nassiri MJ, Vafaei A, Monadjemi A (2007) Texture feature extraction using Slant-Hadamard transform. Int J Appl Sci Eng Technol 3:4
2. Zhang Y, Wang S, Sun P, Phillips P (2015) Pathological brain detection based on wavelet

entropy and Hu moment invariants. Bio-med Mater Eng 26(s1):S1283–S1290
3. Wang S, Zhang Y, Yang X, Sun P, Dong Z, Liu A et al (2015) Pathological brain detection by a novel image feature-fractional Fourier entropy. Entropy 17(12):8278–8296
4. Zhou X, Wang S, Xu W, Ji G, Phillips P, Sun P et al (2015) Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier. In: International conference on bioinformatics and biomedical engineering, pp 201–209
5. Laishram R, Kumar W K, Gupta A, Prakash KV (2014) A novel MRI brain edge detection using PSOFCM segmentation and canny algorithm. In: 2014 International conference on electronic systems, signal processing and computing technologies (ICESC), pp 398–401
6. Oo SZ, Khaing AS (2014) Brain tumor detection and segmentation using watershed segmentation and morphological operation. IJRET Int J Res Eng Technol 3(3):367–374
7. Ibrahim WH, Osman AAA, Mohamed YI (2013) MRI brain image classification using neural networks. In: 2013 International conference on computing, electrical, and electronics engineering (ICCEEE), pp 253–258
8. Islam A, Reza SM, Iftekharuddin KM (2013) Multifractal texture estimation for detection and segmentation of brain tumors. IEEE Trans Biomed Eng 60(11):3204–3215
9. Sivaramakrishnan A, Karnan DM (2013) A novel based approach for extraction of brain tumor in MRI images using soft computing techniques. Int J Adv Res Comput Commun Eng 1845–1848. ISSN, No. 2319-5940
10. Ali S, Abood LK, Abdoon RS (2013) Brain tumor extraction in MRI images using clustering and morphological operations techniques. Int J Geograph Inf Syst Appl Remote Sens 4(1):12–25
11. Patil SB (2016) Detection of brain tumor based on segmentation using region growing method. Int J Eng Innov Res 5(2):173–176
12. Machhale K, Nandpuru HB et al (2015) MRI brain cancer classification using hybrid classifier (SVM-KNN). In: International conference on industrial instrumentation and control (ICIC). IEEE, pp 60–65. 978-1-4799-7165-7/15
13. Xiao K, Lei Liang A, Guan HB, Hassanien AE (2013) Extraction and application of deformation based feature in medical images. Neurocomputing
14. Gopal PN, Sukanesh R (2013) Wavelet statistical feature based segmentation and classification of brain computed tomography images. IET Image Process 7:25–32
15. Kalbkhani H, Shayesteh MG, Zali-Vargahan B (2013) Robust algorithm for brain magnetic resonance image classification based on GARCH variances series. Biomed Signal Process Control 8:909–919
16. Sindhumol S, Kumar A, Balakrishnan K (2013) Spectral clustering independent component analysis for tissue classification from brain MRI. Biomed Signal Process Control 667–674
17. Navarro DH et al (2015) Feature and model selection with discriminatory visualization for diagnostic classification of brain tumor. Neurocomputing 73(4–6):622–632
18. Sumitra N, Saxena RK (2013) Brain tumor classification using back propagation neural network. Int J Image Graph Signal Process 45–50
19. Jayachandran A, Dhanasekaran R (2013) Brain tumor detection and classification of MRI using texture feature and fuzzy SVM classifiers. Res J Appl Sci Eng Tech 6(12):2264–2269
20. Padma Nanthagopal A, Sukanesh Rajamony R (2013) Classification of benign and malignant brain tumor CT images using wavelet texture parameters and neural network classifiers. J Vis 16:19–28
21. Shree NV, Kumar TNR (2018) Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. Brain Inf 5(1):23–30
22. Mohsen H, El-Dahshan E-SA, El-Horbaty E-SM, Salem A-BM (2018) Classification using deep learning neural networks for brain tumors. Future Comput Inf J 3(1):68–71
23. Mathew AR, Babu Anto P (2017) Tumor detection and classification of MRI brain image using wavelet transform and SVM. In: 2017 International conference on signal processing and communication (ICSPC). IEEE, pp 75–78
24. Shil SK, Polly FP, Hossain MA, Shareef Ifthekhar Md, Uddin MN, Jang YM (2017) An improved brain tumor detection and classification mechanism. In: 2017 International conference on information and communication technology convergence (ICTC). IEEE, pp 54–57

25. Usman K, Rajpoot K (2017) Brain tumor classification from multi-modality MRI using wavelets and machine learning. Pattern Anal Appl 20(3):871–881. https://doi.org/10.1007/s10044-017-0597-8
26. Amin J, Sharif M, Yasmin M, Fernandes SL (2017) A distinctive approach in brain tumor detection and classification using MRI. Pattern Recogn Lett. https://doi.org/10.1016/j.patrec.2017.10.036
27. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, Wang Z, Feng Q (2015) Enhanced performance of brain tumor classification via tumor region augmentation and partition. PloS One 10(10):e0140381
28. Kermi A, Andjouh K, Zidane F (2018) Fully automated brain tumor segmentation system in 3D-MRI using symmetry analysis of brain and level sets. IET Image Process 12(11):1964–1971
29. Bahadure NB, Ray AK, Thethi HP (2017) Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. Int J Biomed Imaging
30. Praveen GB, Agrawal A (2015) Hybrid approach for brain tumor detection and classification in magnetic resonance images. In: Communication, control and intelligent systems (CCIS). IEEE, pp 162–166
31. Kaur T, Saini BS, Gupta S (2017) Quantitative metric for MR brain tumor grade classification using sample space density measure of analytic intrinsic mode function representation. IET Image Process 11(8):620–632
32. Sornam M, Kavitha MS, Shalini R (2016) Segmentation and classification of brain tumor using wavelet and Zernike based features on MRI. In: IEEE International conference on advances in computer applications (ICACA). IEEE, pp 166–169

Soft and Cloud Computing

A Systematic Review on the QoS-Aware Service Provisioning in Fog Computing

M. K. Dhananjaya, Kalpana Sharma, and Amit Kumar Chaturvedi

Abstract Modern Internet of Things (IoT)-based technologies and services have demanded the processing of vast amounts of data in a short period of time. Surveillance and object detection has risen to the top of the priority list, but Cloud Computing is unable to reduce network latency time to meet response time requirements. Fog computing could tackle this issue. Fog computing is an ideal illustration of how to arrange assets and administrations outside of the cloud, at the organization's edge, and nearer to end gadgets. Fog-cloud computing frameworks have a promising stage for asset provisioning. We will propose engineering for Quality of Service (QoS) mindful Fog administration provisioning in our examination, which will empower the execution of IoT application undertakings on a group of haze hubs to be booked. This paper likewise evaluates the impact of certain security concerns and potential arrangements.

Keywords Internet of Things (IoT) · Object detection · Fog computing · Quality of Service (QoS) · Cloud computing

1 Introduction

Fog computing is another worldview that broadens the Cloud stage model by giving computing assets on the edges of an organization. It is normally portrayed as a cloud-like stage having comparative information, calculation, stockpiling, and application services, yet is essentially extraordinary in that in light of its decentralized nature. Furthermore, Fog frameworks are fit for handling enormous highlights making the Fog stage particularly sensible for time and region tricky submissions. For instance, Internet of Things (IoT) contraptions is needed to rapidly handle an outsized measure

Present Address:
M. K. Dhananjaya (✉) · K. Sharma
Computer Science Department, Bhagwant University, Ajmer, India
e-mail: dhanunitte@gmail.com

A. K. Chaturvedi
CSE Department, Government Engineering College, Ajmer, India

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_44

of information. This wide determination of usefulness-driven measures of data nearby, work on-premise are totally advantageous and may be presented on mixed hardware. This application expands various security issues concerning data, virtualization, disconnection, organization, malware, and observing. As of now, Internet applications running on cell phones create a colossal measure of information that will be communicated to Cloud Computing for preparation. In any case, one basic restriction of a Cloud is the network with end gadgets. Fog computing beats this impediment and supports the needs of time-delicate applications by conveying calculation, correspondence, and capacity administrations along with the Cloud to Things (C2T) continuum, engaging expected new applications, similar to keen urban communities, increased reality (AR), and PC game (VR).

In any case, the reception of Fog-based computational assets and they are joining with the Cloud presents new difficulties in asset the executives, which needs the execution of inventive procedures to guarantee consistency with the Quality of Service (QoS) necessities of uses. During this unique situation, one significant inquiry is the best approach to plan the QoS necessities of uses for Fog and Cloud assets. One potential methodology is to segregate applications showing up at Fog into Classes of Service (CoS). Thus, this presents a gathering of CoS for Fog applications which fuses, the QoS prerequisites that best describe these Fog applications. In addition, this proposes the execution of a commonplace AI characterization strategy to segregate Fog registering applications as a component of their QoS prerequisites. The appropriation of a methodology for AI-based classification comprises a first venture toward the definition of QoS provisioning instruments in Fog registering. For contemporary cloud-based frameworks, all assistance solicitations and asset requests are broken down and handled inside Data Centers (DCs). Since ordinary distributed computing includes preparing, calculation, and capacity of the information just inside DCs, the gigantic information traffic created from the IoT gadgets is expected to encounter a tremendous organization bottleneck, which thus high assistance inertness and low Quality of Service (QoS). The fundamental goal of this paper is to present an incredible QoS provisioning framework (QoPF) for administration arranged IoT. The QoPF framework is anticipated to grow the compound assistance superiority in the IoT application layer by adjusting administration dependability and an adequate expense of computational time.

The predominant proposed procedures in haze figuring come up short on the versatile and savvy conduct. They just consider the immediate correspondence of end gadgets ideally with close haze hubs. Further, it is very mind-boggling to execute versatile and shrewd learning-based errand planning during this design. Therefore, we cannot exploit heterogeneity of haze hubs encompassed by IoT gadgets. To work with shrewd and flexible task arranging courses of action during this vibrant and mixed situation, we want a splendid layer between end gadgets and Fog Center points that ought to have three guideline capacities: (1) The capacity to describe whether or not the advancing toward deals ought to be served by a cloud or a Fog Center, (2) The possibility to plan the approaching undertaking to suitable haze hub among the accessible haze gadgets, and (3) For task planning this layer ought to have the usefulness that broadly executes versatile and clever learning-based assignment booking.

2 Related Work

Cloud computing might be a typical answer for IoT framework execution issues. Notwithstanding, during a cloud climate, the proceeded with age of enormous measures of information causes network blockage and focused overhead on the focal worker. Mist hubs in FC (Fog Computing) take care of this issue by preparing a few or the entirety of the information prior to sending it to the cloud. Consequently, FC components significantly lessen information traffic and weight on the cloud [1]. Maintaining these benefits as a primary concern, a few investigations have endeavored to circulate Deep Learning overhead by applying mist processing systems. Regardless, when all haze hubs work their best steady with accessible assets, the weights of cloud and organization traffic are diminished, and more DL applications are frequently obliged in an FC climate. Thus, we expected to treat one among the most difficulties for executing DL in Fog Computing framework, the decision of ideal layers notwithstanding lopsided accessible asset in mist hubs [2].

The examination centers around unique help provisioning in edge mists according to a hypothetical point of view. The model inside the examination catches the restricted limit of mist hubs, the obscure appearance cycle of solicitations, the benefit of sending solicitations to the far-off cloud, and along these lines the expense of stacking a substitution administration on a haze hub. During a re-enactment study, the creators in [3] recommend an applied edge work for haze asset provisioning. They present the idea of "mist cell," which might be a product segment running on haze gadgets that control and screen a specific social affair of IoT gadgets [4]. Utilizing this and other related thoughts, the model coordination of IoT gadgets utilizing a various leveled cloud/haze asset control and supply a fitting asset provisioning answer for disseminating assignments among them. To adjust sent application in dispersed frameworks, Wen et al. highlight the significance of mist arrangement. The creators in [5] foster strategies for contemplating mist organization frameworks with consideration on challenges in unwavering quality, versatility, and security for mist coordination. Another exertion for administration arrangement is that crafted by the creators in [6] called "Hazy.“ Foggy might be a structure for persistent and programmed application sending inside the haze. It works with dynamic asset provisioning and programmed IoT application sending in mist design. The creator in [7] proposes a help situated middleware that means to convey administrations over haze hubs for versatility, what is more with the assistance of SDN, achieves QoS mindful organization by booking streams among administrations. Most existing plans for haze frameworks, for example, offloading, load adjusting, or administration provisioning, just think about a couple of destinations (e.g., QoS, cost), accepting that different targets have no impact on the issue. Another exploration course will be to foster plans that think about different targets (e.g., QoS, transmission capacity, energy, and cost) simultaneously [8].

Saurez et al. recommend Fog-Lets, a program design that works with dispersed indoctrination through Fog Nodes. Foglets offers APIs to spatio transient information consideration for taking care of and recuperating application-delivered information on the local center points. Over and done with the Foglets API, Foglets measures are set for a certain geospatial locale and Foglets accomplishes the submission parts on the Fog centers [9]. Foglets is executed through holder-based insight. The Foglets API considers QoS and weight changing while at the same time migrating continuing on (stateful) data between fog centers. As a future course, one could consider other QoS qualities and consolidate cost data to get a more excessive request of qualified fog plans [10].

By examining, a unique QoS provisioning system (QoPF) is proposed to improve administration situated IoT. As of late, empowering the ideal help structure has been worried about two significant issues: I) Real-time versatile detecting and ii) Performance corruption inside quite a while. The QoPF can be utilized to tackle these issues dependent on QoS. The qualities of the QoPF are shown in separating the intricacy of QoS estimation for the IoT composite administrations [11]. In this manner, the composite help execution is improved and the ongoing versatile detecting issue is additionally decreased.

From the unwavering quality and accessibility point of view, mist administrations and mist networks carry new difficulties to the current organization and administration provisioning techniques [12]. To ensure the accessibility and unwavering quality of the haze benefits, an organized help provisioning instrument that considers both haze and distributed computing are required [13]. For instance, if a mist administration needs a few capacities to handle a surge of information, giving additional copies of those capacities can work on the accessibility of the assistance. Then again, because of the restricted figuring assets of the mist hubs to give accessibility and unwavering quality is certifiably not a direct choice. As a future heading, accessibility might be considered notwithstanding limitations, like idleness, throughput, and security when planning provisioning strategies for mist administrations [2].

3 Safekeeping Encounters of Fog Computing

As Fog Computing might be a promising advancement that extends the advertisement ministration from cloud to the edge of the organization. Like Cloud processing, Fog Computing additionally offers administrations to the top clients with preparing and storerooms. Regarding the issues of Fog processing referenced in Table 1, the considered writing distinguished diverse security challenges like confirmation, protection, information security, and different pernicious assaults. In an alternate work, the creators of [14] proposed three distinct systems for ideal asset usage. First and foremost, an arbitrary approach is utilized to choose the haze hubs to execute undertakings upon appearance. Furthermore, the emphasis will be on the most minimal idleness

Table 1 Outline of summarized work

Year	Problem	Methodology	Pros	Cons
2019	QoE in the progressive circulated and heterogeneous Fog-IoT Environment	Focused on application situation to the reasonable Fog workers utilizing Fuzzy rationale models	QoE-mindful application position	Energy utilization is not assessed
2017	Scheduling in the fog computing	Simultaneous, FCFS and Delay need booking approaches	Computer processor Execution time and postponement	Non-agreeable recreation climate
2016	Computational and correspondence delays in high burden condition	5G Fog-based foundation is reproduced	Correspondence dormancy, figuring inertness	Recreated on incomplete Fog Computing framework
2015	Resource allocation	Reliably base undertaking planning model, an assistance arranged asset management	Resource the executives, execution measure	The enormous scope demand is overlooked

haze gadgets. At long last, the mist assets having the most extreme accessible limit ought to be fundamentally considered [15].

Fog computing is a classical wherein the framework attempts to thrust information preparing from cloud workers to "close" IoT gadgets to diminish idleness time [16]. The execution orderings and the sent spots of administrations make critical impact on the general reaction season of an application. The administration organization issue is one of the appealing exploration arenas of Fog computing [17]. The administration arrangement is a multi-objective improvement issue, there are countless proposed answers for different targets, for example, reaction time, correspondence cost, and energy consumption [18]. In Fog Computing, submission administrations run on both edge hubs (with low idleness access however exceptionally restricted asset capacities) and in the cloud (with higher access dormancy yet basically limitless resources) as well as on conceivable mediator nodes [19]. The issue of information handling in the IoT climate is addressed by an arising worldview of Fog processing. In Fog enrolling, edge devices are used to manage the customer request which has truly taken care of force and are closer to the data sources than cloud resources. Thus, QoS-careful resource the leaders system is proposed for useful organization of resources which considers execution time, network use, and energy usage as QoS limits. The presentation of the proposed technique has been surveyed in Fog enrolling environment using iFogSim tool compartment and the experimental outcomes show that the proposed procedure performs better similar to QoS limits.

4 Proposed Structure for QoS-Aware Fog Service Provisioning

An arrangement for Fog association provisioning, which desires to commit IoT application's undertakings to proper Fog Nodes in a great deal of Fog Nodes. The treatment of data streams at the Fog decreases network traffic and works on the dormancy of time-touchy applications. The significantly considered setup is the Fog Broker, which is obligated for booking the execution of tasks on fitting Fog Nodes utilizing particular orchestrating game plans.

The recommendation of a four-layer-based design for QoS-careful Fog promotion ministration good for visioning is presented in Fig. 1. The fundamental fragments of the framework are IoT Gateways (IoTG), the Fog Broker, the gathering of Fog center points, and more critical level applications.

IoT Gateway: An IoT entryway is responsible for conglomerating the data got from IoT Devices and Sensors, deciphering the shows of the sensors to guarantee their interoperability, and pre-taking care of the data preceding moving it to various levels for additional created planning.

Fog Broker: The Fog Broker is a coordinator part that widens the capacities of the IoT entrance to deal with a gathering of dimness centers.

Fog Node: The customary pieces of a Fog Node might fuse an orchestrate ment structure, an appropriated illuminating system, a data planning engine, data storing, and a pile really taking a look at portion.

The novel resource provisioning (i.e., auto-scaling) is a capable method for managing deal with the IoT obligation fluctuations, and it is the ability to scale out and in Fog Nodes as shown by the moving toward IoT contraptions obligation. Most certainly, the resource provisioning approaches self-administering change when and

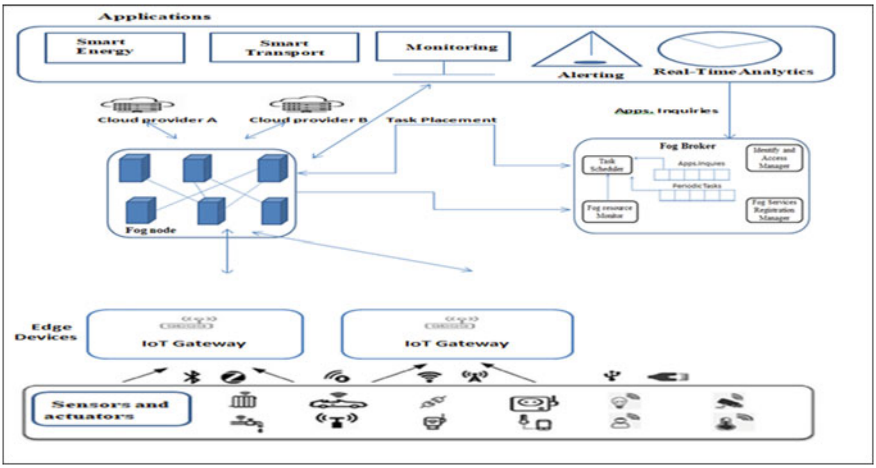


Fig. 1 Design for QoS based fog-administrations provisioning

how to get and pass on Fog Nodes to delight the IoT contraptions resource interest and avoid resource wastage. Resource provisioning approaches should be deftly accustomed to orchestrate with the commitment changes of IoT associations at run time and ruin Fog Node surplus through two scaling works out: the scaling-out movement when then burden developments to cover up QoS necessities, and the surmounting, all things considered, when the store decreases to construct the utilization of Fog Nodes.

A design for haze organizations provisioning, which expects to assign IoT application's errands to suitable haze hubs in a bunch of mist hubs. The handling of information streams at the mist decreases network traffic and works on the dormancy of time-touchy applications. The fundamental idea of the design is the haze agent, which is responsible for planning the execution of errands on fitting mist hubs utilizing different booking arrangements. With regards to mist conditions, planning choices should be made as fast as could be expected. The reasoning is that haze processing is expected to further develop application idleness, and numerous applications might go after the proper assets, and time allotments wanted by an application assignment can be involved by one more application without warning. The goal here is to limit the idleness of undertakings and guarantee that time-delicate applications comply with their time constraints. A technique called "Measure" is to progressively assign assets in mist registering climate, which can keep away from both too high and too low loads.

For the heap balance issue in mist figuring, a custom haze bunching algorithm is proposed to tackle the issue. In this issue, a few clients need to offload calculations and every one of their requests should be taken care of by neighborhood registering group. There is an emphasis on issues about guaranteeing QoS and lessening SLA (Service Level Agreements) infringement and asset the board. On account of QoS confirmation and SLA infringement decrease, the fundamental arrangement of QoS affirmation is productive VM the board. This arrangement can meet client's prerequisites through sensible booking and reconciliation of VMs.

5 Conclusion

Fog computing as another worldview is a generally undiscovered subject with a huge number of unsettled examination questions. An arising Fog figuring worldview takes care of the issue of information preparing in the IoT setting. Late headways in the Internet of Things (IoT) have brought about the ascent of an assortment of helpful IoT applications with shifting Quality of Service (QoS) needs. Fog-Cloud frameworks give an expected climate to asset provisioning for IoT application administrations. In this paper, a QoS mindful asset provisioning system is proposed for proficient resource the executives, which takes execution time, network usage, and energy utilization into account as QoS boundaries.

The combination of IoT with processing advances, for example, Cloud, Edge, and Fog Computing have gotten significant, where the IoT can utilize Cloud Computing as a virtual assets use framework by empowering on request utility estimating models to defeat IoT asset imperatives. Then again, intermingling innovation has arisen to urge IoT to foster a bunch of composite administrations dependent on a nature of administration (QoS) way to deal with address client issues. Albeit this combination considers consistent canny communication among physical and virtual things, it faces difficulties in not just gathering the imperative degree of Quality of Service (QoS), yet in addition fulfilling the client's convoluted requests. Thus, the longing to fabricate a powerful assistance arranged climate has arisen as a key plan thought. The significant objective of this exploration is to foster a powerful Quality of Service Provisioning Framework (QoPF) for Service-Oriented IoT applications.

This construction was planned to increase composite help quality in the IoT application layer by finding a concordance between organization dependability and a reasonable computational time cost.

To evaluate the exhibition of use errands' situation in a mist environment, Cloud Analyst device has been utilized. It is a recreation device that permits mimicking cloud conditions by giving the capacity to characterize server farms and the hosts on every server farm, and indicate three booking policies(closest server farm, improve reaction time, reconfigure progressively with load). Each host can uphold a few virtual machines. The tests were reproduced on MacBook Pro, center i5 processor, 3.40 GHz with 8 GB of memory. As in this unique circumstance, haze hubs are situated in the equivalent environment near IoT gadgets.

6 Results and Discussions

Working with cloudsim is used to be performed by use of the some software and applications. As the software needed is java 6 or newer. Cloudsim will not work on the old versions of Java. And cloudsim 3.0 tool kit which consists of the Cloudsim API, cloudsim version 3.0 is the most recent rendition with every one of the most recent refreshes bug fixes. The updates in this structure are new VM scheduler, new datacenter network model, new VM apportioning and assurance arrangements new power models, new obligation tracks, support for outside positions, and sponsorship for customer portrayed completion of re-enactment.

Cloud Analyst offers assistance for reproducing three plans for load changing across VMs in a clear fog center. These plans are Round-Robin, Equally Spread Current Execution Load, and Throttled. Likewise, it gives three representative planning courses of action: Optimize Response Time (ORT), Closest Fog Node (CFN), and Reconfigure Dynamically with Load (RDL). With this plan and setting reproduce the execution of the application requests for 24 h using these three booking policies, independently. The amount of simultaneous sales a fog center maintains is set to 10, and the store changing technique of the VMs in a fog center point is the Equally Spread Current Execution Load procedure, Round robin, and Throttled independently.

References

1. Guerrero C, Lera I, Juiz C (2018) On the influence of fog colonies partitioning in fog application makespan. In: 2018 IEEE 6th international conference on future internet of things and cloud (Fi Cloud), IEEE, pp 377–384; Minh QT, Nguyen DT, Van Le A, Nguyen HD, Truong A (2017) Toward service placement on Fog computing landscape. In: 2017 4th NAFOSTED conference on information and computer science, IEEE, pp 291–296
2. Skarlat O, Nardelli M, Schulte S, Borkowski M, Leitner P (2017) Optimized IoT service placement in the fog. SOCA 11(4):427–443
3. Yousefpour A et al. (2018) Qos-aware dynamic fog service provisioning. arXiv preprint arXiv: 1802.00800
4. Tran M-Q, Nguyen DT, Le VA, Nguyen DH, Pham TV (2019) Task placement on fog computing made efficient for IoT application provision. Wireless Commun Mobile Comput 2019
5. Singh S, Chana I (2015) QoS-aware autonomic resource management in cloud computing: a system review. ACM Comput Surv 48(3):1–46
6. Vahid Dastjerdi A, Gupta H, Calheiros RN, Ghosh SK, Buyya R (2016) Fog computing: principals, architectures, and applications. arXiv:1601.02752
7. Amir Vahid D, Buyya R (2016) Fog Computing: Helping the Internet of Things realize its Potential." Computer 49, no. 8 (2016); 112–116.
8. Lee W, Nam K, Roh HG, Kim SH (2016) A gateway based fog computing architecture for wireless sensors and actuator network. In: 2016 18th international conference on advanced communication technology (ICACT), IEEE, pp 210–213
9. Agarwal S, Yadav S, Yadav AK (2016) An efficient architecture and algorithm for resource provisioning in fog computing. Int J Inform Eng Electron Business 8(1):48
10. Gupta H, Vahid Dasjerdi A, Ghosh SK, Buyya R (2016) iFogSim: a toolkit for modeling and simulation of resource management techniques in internet of things, edge and fog computing environments. arXiv preprint arXiv: 1606.02007
11. Mung C, Tao Z (2020) Fog and IoT: an overview of research opportunities. IEEE Internet Things J 3(6):854–864
12. Chiang M, Zhang T (2016) Fog and IoT: an overview of research opportunities. IEEE Internet Things J 3(6):854–864
13. Ansari N, Sun X (2018) Mobile edge computing empowers internet of things. (Invited Paper) IEICE Trans Commun E101-B(3):604
14. Vahid Dastjerdi A, Gupta H, Calheiros RN, Ghosh SK, Buyya R (2016) Fog computing: principles, architectures, and applications. In: Internet of Things, Elsevier, pp 61–75
15. Manis Haruna A, Oladipo Francisca O, Ezendu A (2017) User mobility and resource scheduling and management in fog computing to support IoT devices. In The seventh international conference on innovative computing technology (INTECH), IEEE, pp 191–196
16. Mung C, Tao Z (2016) Fog and IoT: an overview of research opportunities. IEEE Internet of Things J 3(6):854–862
17. Amir Vahid D, Buyya R (2016) Fog computing: helping the internet of things realize its potential. Computer 49(8):112–116
18. Anil S, Nitin A (2019) Load balancing aware scheduling algorithms for fog networks. Softw: Practice Exp 3(1):1–19
19. Anif A, Guillaume P (2018) Docker container in fog computing infrastructures. Int J Cloud computing, pp 1–30

Effective Use of Visualization Techniques for Educational Institutional Data Analysis

Ambika Patidar and Sharmishta Desai

Abstract Data can be turned into easily understandable format and relevant business information with data visualization. Nowadays, there is more use of data visualization in business intelligence. Datasets can be visualized in a variety of visual insights using a different dynamic or interactive approach. This paper focuses on Microsoft Power BI operations and interactive display of educational institution databases with multiple modules. It uses live dashboards, interactive reports, charts, graphs, etc. to help users build significant business insight quickly and effectively. Selecting categories, values and fields make visualizations more accurate and gives desired results. Features like slicer and Q&A help to analyze, find out results quickly on larger datasets. Applying filters on visuals based on data features and set range to get insights from large datasets makes it easier. To determine relationship between features, OLAP visuals are created and compared to Power BI based on field values.

Keywords Power BI · Business intelligence · OLAP · Data analytics · Database · Visualizations · Reports · Dashboard

1 Introduction

The demand for a database of educational institutions is growing. An effective solution is required to analyze and visualize the data. Power BI helps in analyzing visualizations and creating effective reports about the data. In this model, Effective use of Visualization Techniques for Educational Institutional Data Analysis has been done. With the diversity of visualizations accessible in Power BI, data visualization aids in better visual insights from the data. Selecting categories, values, fields, and other parameters improves the accuracy of the visualizations and ensures that the user gets the required results. Slicer, Q&A, and other features assist analyze and finding results

A. Patidar (✉) · S. Desai
Maharashtra Institute of Technology, World Peace University, Pune, India
e-mail: ambika.patidar.5232@gmail.com

S. Desai
e-mail: sharmishta.desai@mitwpu.edu.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_45

rapidly on huge datasets. It is more user pleasant and simpler to obtain insights from huge datasets by applying filters to visualizations based on data characteristics and setting the range. Later the results are compared with OLAP Cube. Dataset has been taken from Kaggle. Data Mining is done by use of OLAP tool to extract and discover patterns in large data. Transformation of data is done by using Power Query Editor and Business Intelligence reports, visualizations, and dashboards are generated in Power BI.

2 Literature Review

Created Interactive dashboards, visualizations, and Power BI's operations, the many types of data sources available in the tool, and the various sorts of visual insights or context. The visualizations can be improved by Data Analytics on Power Query editor [1]. BI that can track the development of a student's education, OLAP to analyze to analyze and show data on dashboard. Each school can address business questions about scoring by employing a data warehouse [2]. Infrastructure is built on research, educational data analyses academic processes, and education, data mining uses learning analytics techniques and methodologies, OLAP, and Enterprise Architecture analysis. It is used to perform various experiments to generate explicit knowledge. BI and analytics performance can be improved in OLAP by using Multidimensional Data Representation [3]. Implemented BI and learning analytics, route to address present challenges in the interplay between higher education institutions and learning performance is provided by building a revolutionary closed-loop feedback mechanism [4]. Data warehousing, data mining, and Microsoft cluster techniques were implemented. Discovered a group of kids who shared same interests. This behavior's subjects allow users to investigate the causes and potential consequences for a student's performance. By successfully utilizing huge datasets and data mining technologies, this model can be enhanced [5]. Implemented Social Data Analysis using Decision trees C4.5, Hoeffding trees, naïve bayes, etc. to analyze data and it is useful for institutes for getting feedback about any student or services. The method is traditional and data analytics tools can be used to get insights into social media data [6]. Implemented Very Fast Decision Tree (VFDT) on Hadoop for handling Big Data Applications. Unstructured data can be handled better [7]. Used Random Decision Tree (RDT) algorithms to analyze Social Media data for business analysts and researchers. Compared to other algorithms, it produces accurate results [8]. Created Excel, Power BI, Tableau, and IBM Watson dashboards and compared based on academic execution. Effective detailed data analysis can be done [9]. Study paper demonstrates ideas of Power BI, including analyzing BI and self-service BI, as well as their interaction. Also shown, how to utilize Power BI to create optimal BI solutions for specific businesses [10].

3 Significance of the Topic and Research Gap

Significance is analysis and design of database ensures that all information on the University database remains same for all network users. One strategy to improve data access and quality at a learning institution are to implement an effective data management system. It helps to organize large data and managing data becomes easier. The visualizations, dashboards, and reports can be improved by data analytics. Data transformation should be done on complex datasets. Data Mining tools and techniques can be used to extract and discover patterns in large data. Business Intelligence and analytics performance can be improved in OLAP by using multi-dimensional data representation.

4 Details of Design/Technology

4.1 Data Source

Source: CSV (Comma Separated Values) Dataset from Kaggle.

CSV Files: Student_Performance_Data.csv, Student_Counceling_Information.csv, Department_Information.csv, Employee_Information.csv.

4.2 Brief of Technology Used

Power Query Editor. It is an engine for data transformation and preparation. It comes with a graphical interface for getting data from Power Query sources as well as a Power Query editor for performing transforms. Because engines are used in so many different products and services, the data's eventual destination is determined by where Power Query is utilized. It can be used to run data extraction, transformation, and loading (ETL) processes. Its purpose is to make it easy to make necessary modifications using user-friendly buttons, menus, ribbons & other interactive components. Users can connect to different data sources, preview data, choose to convert from user interface & hundreds of data transformations can be carried out. These data conversion features are universal from all data sources, regardless of basic restrictions on data sources.

Power BI. Power Query is an engine for data transformation and preparation. It comes with a graphical interface It is a set of software services, connections, and applications that work together to turn various data sources into logical, visually engaging interactive insights. Users can connect to data sources with ease, view and uncover essential information, and share the results with selected people. Power BI consists of several components that work together, starting with the following three basics: Power BI Desktop, Power BI service—Online software as a service

(SaaS) service, and Power BI Mobile application—For Windows, iOS, and Android devices. The Power BI Desktop unifies, simplifies, and accelerates business intelligence storage and report design and authoring. Its detailed data can be easily understood by data visualization, which can be transformed into visually convincing and useful business information. Dashboard is a single-page canvas that tells a story using different visual elements. If the user is new to Power BI, understanding the basic concepts of Power BI can help the user to get started.

OLAP Cube. OLAP (Online Analytical Processing) Cube enables quick data analysis based on some dimensions that describe a business demand. It is a type of software that enables users to compare data from many databases simultaneously. It is a tool that can receive and analyze corporate data from many different views. For analysts, aggregating, grouping, and joining data is a typical operation. Relational databases use a lot of resources to perform these processes. OLAP allows data to be pre-calculated and pre-aggregated, which speeds up analysis. OLAP databases are separated by one or more cubes.

5 Design Methodology

5.1 Power BI

The steps to configure a complex data model in Power BI are simple. When data comes in from many transactional systems, it can have dozens of tables to handle. It is all about creating a good data model and simplifying the confusion. Star schema is one of the technologies that make the data model easier to understand. Optimal data granularity depends on report performance and ease of use. A powerful data model makes exploring data faster, easier to aggregate, more accurate reports, less time to write reports, and easier maintenance of reports in the future. Using this dataset, Power Query Editor to extract, convert, and load the dataset's key features is utilized. Then, using four datasets (Student Performance Data.csv, Student Counseling Information.csv, Department Information.csv, and Employee Information.csv), create reports and visualizations. Following the creation of reports, the University database will be published to Power BI Service in order to create interactive dashboards.

5.2 OLAP Cube

The process of designing and building a data cube based on OLAP queries. Business intelligence systems are a good way to perform multi-dimensional online computing and data analysis on vast amounts of data. The analyzed data is often stored in groups of heterogeneous cubes during the decision-making process. OLAP cubes are databases specially developed for reporting. Student Performance Data.csv, Student

Counseling Information.csv, Department Information.csv, and Employee Information.csv graphics are built based on the measure and dimensions for the application of these four datasets. Graphs are made using dimensions and measures.

6 Analytical and/or Experimental Work

6.1 Power BI Implementation

All four datasets are imported and transformed using Power Query Editor. Relation between features of dataset like one-to-many relation of Department ID in Department Information table to Employee Information table, one-to-many relation of Department Name in Department Information table to Student Counseling Information table as Department Admission and Department Choices and many-to-many relation of Student ID in Student Counseling Information table to Student Performance Data table, shown using Model Flow (Figs. 1, 2 and 3).

Two reports of Student Performance have been created. These reports show the performance of students during all semesters in all papers. Student ID is the primary key and Marks is the target column. In report 1, line chart representing the variation of student marks by Student ID. Heat Map of representing student marks by Paper ID. Single row card representation of total number of marks of all students. Two matrix representations for different features of Student Performance Dataset. Report 2 has another heat map showing the number of marks by Student ID and Semester Name (Figs. 4 and 5).

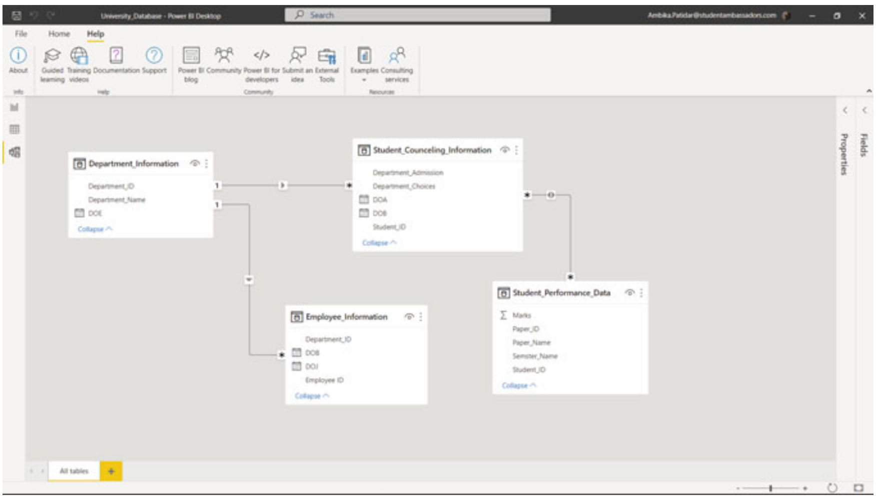


Fig. 1 Schema representation in power BI (Model Flow)

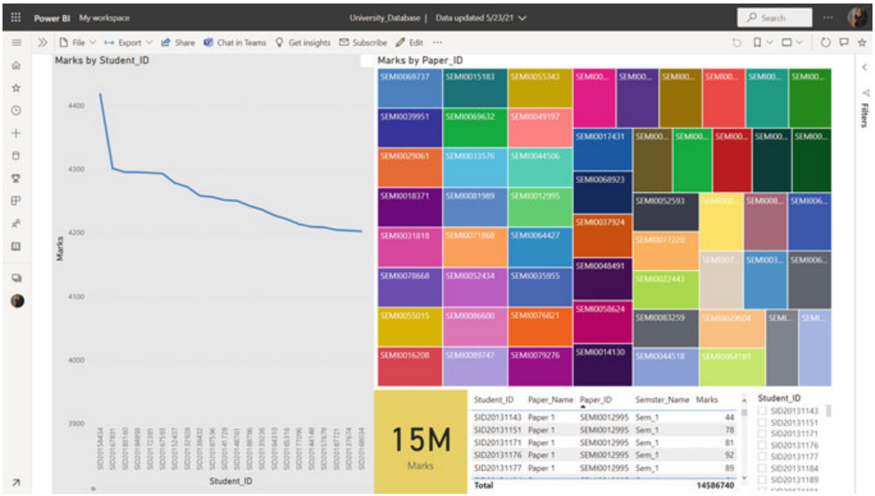


Fig. 2 Student performance report 1

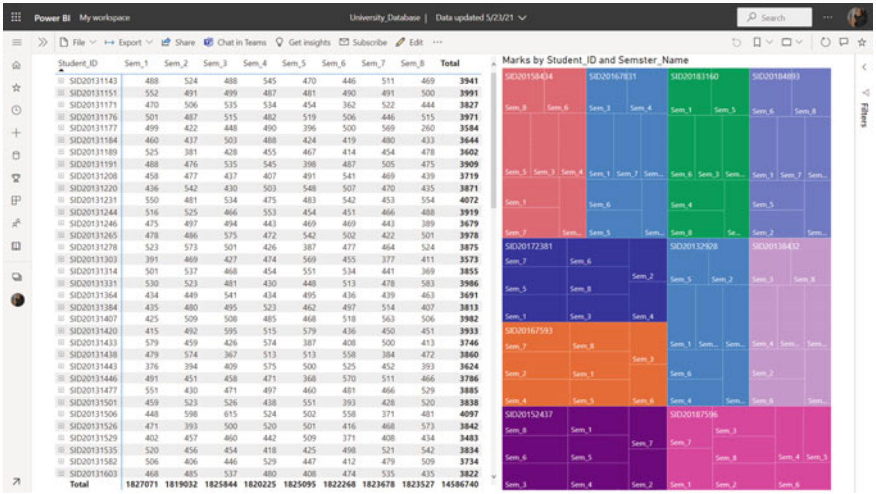


Fig. 3 Student performance report 2

In the above two figures, two reports of Student Counseling have been created. These reports represent the data of students during the counseling and admission process. Student ID is the primary key in above reports. In report 1, heat map showing the count of Student ID by Department Admission and Department Choices, stacked column chart displays the count of Student ID by Department Admission and ribbon chart representing count of Student ID by Department Admission and Department Choices. In report 2, line chart representing count of DOA (Date of Admission) by

Fig. 4 Student counseling report 1

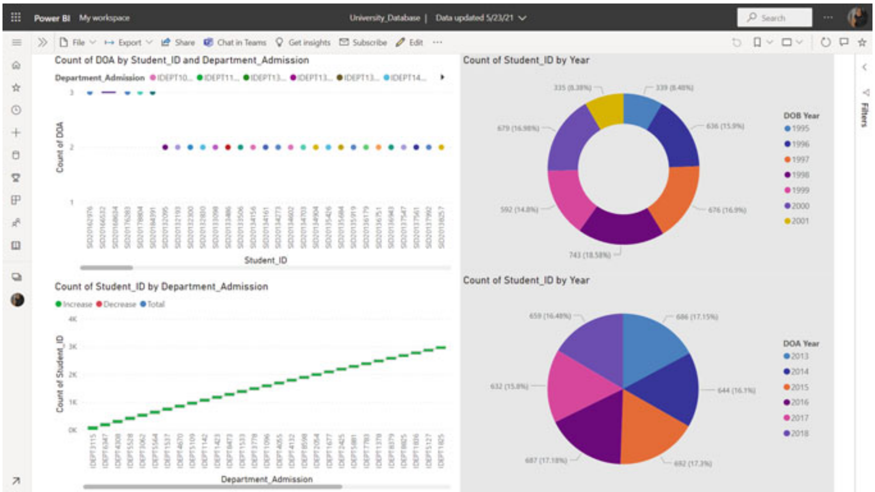


Fig. 5 Student counseling report 2

Student ID and Department Admission, waterfall chart showing count of Student ID by Department Admission, donut chart and pie chart showing count of Student ID by Year (Figs. 6 and 7).

In the above two figures, two reports of Employee Information have been created. These reports represent the Employee data from the time of joining and their departments' information. Employee ID is the primary key in above reports. In report 1, line chart representing count of Employee ID by Department ID, pie chart depicting

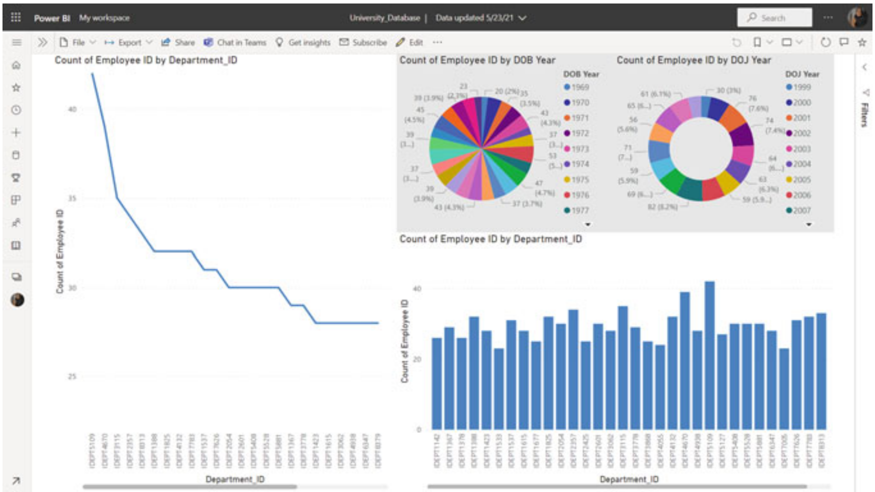


Fig. 6 Employee information report 1

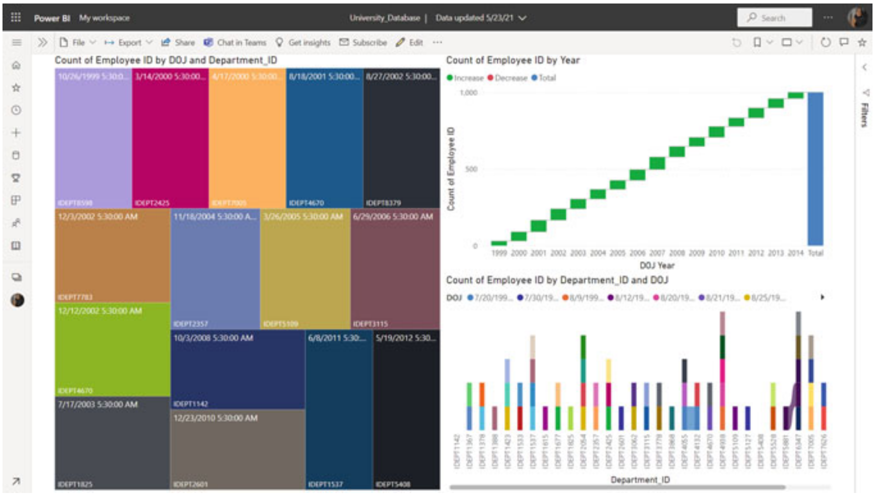


Fig. 7 Employee information report 2

count of Employee ID by DOB (Date of Birth) Year, donut chart depicting count of Employee ID by DOJ (Date of Joining) Year, stacked column chart representing count of Employee ID by Department ID. In report 2, heat map showing count of Employee ID by DOJ and Department ID, waterfall chart showing count of Employee ID by Year and ribbon chart representing count of Employee ID by Department ID and DOJ (Fig. 8).

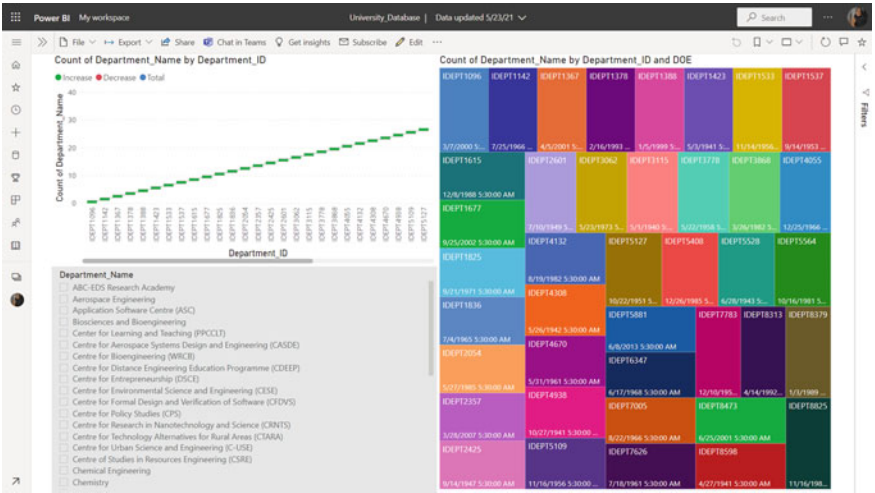


Fig. 8 Department information report

In the above figure, a report of Department information has been created. This report shows the data of all the departments with their IDs. In this report, waterfall chart showing the count of Department Name by Department ID and heat map showing the count of Department Name by Department ID and DOE (Date of Establishment). Slicer showing all the department names so that the user can select one or more departments together and according to that information on visuals will be displayed (Fig. 9).

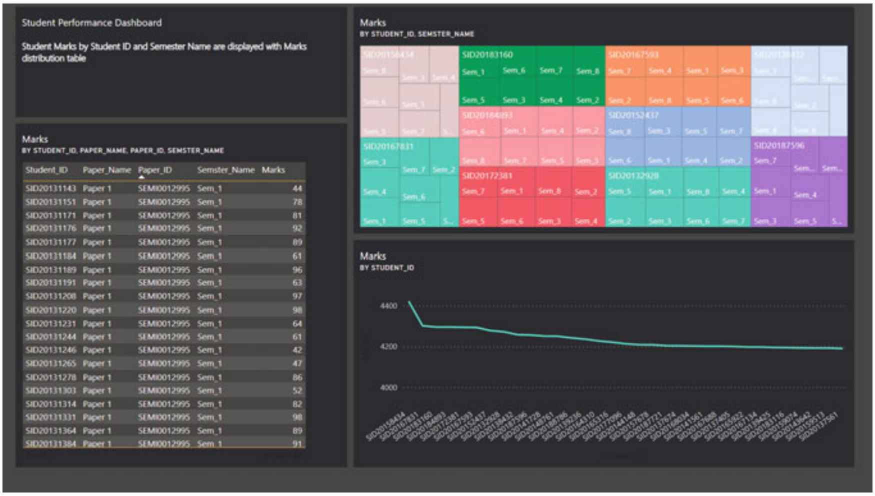


Fig. 9 Student performance dashboard

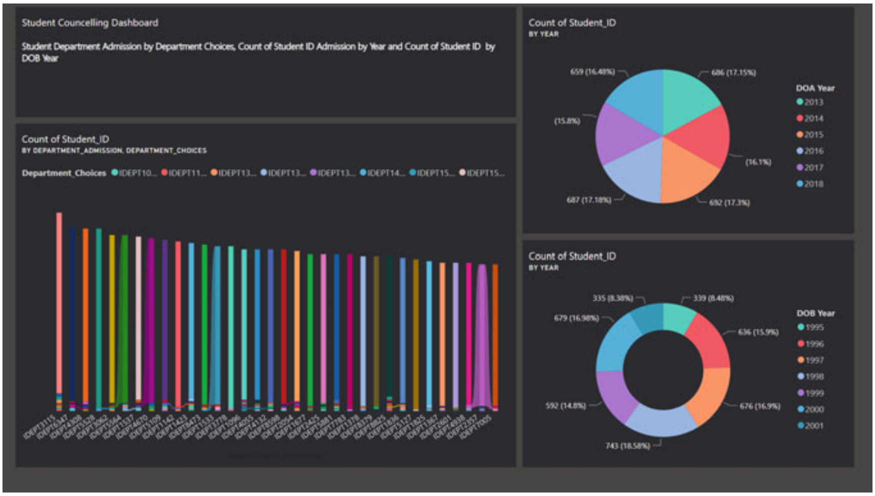


Fig. 10 Student counseling dashboard

Student Performance Dashboard is created by using important visualizations like matrix representations for different features of Student Performance Dataset, heat map showing the number of marks by Student ID and Semester Name and line chart representing variation of student marks by Student ID from Student Performance Reports (Fig. 10).

Student Counseling Dashboard is created by using important visualizations like ribbon chart representing the count of Student ID by Department Admission and Department Choices, pie chart and donut chart showing count of Student ID by Year from Student Counseling Reports (Fig. 11).

Employee Information Dashboard is created by using important visualizations like donut chart depicting count of Employee ID by DOJ (Date of Joining) Year, line chart representing count of Employee ID by Department ID, heat map showing the count of Employee ID by DOJ and Department ID and ribbon chart representing the count of Employee ID by Department ID and DOJ from Employee Information Reports (Fig. 12).

The Department Information Dashboard is created by using important visualizations like waterfall chart showing the count of Department Name by Department ID and heat map showing the count of Department Name by Department ID and DOE (Date of Establishment) from Department Information Reports.

6.2 *OLAP Cube Implementation*

In the above figure, Employee Information Stacked Bar Chart is created in OLAP by using dimensions as features and measure as Count of Employee. Dimensions here

Fig. 11 Employee information dashboard

Fig. 12 Department information dashboard

represent the Employee ID. Hovering on a particular bar shows count of employees (Figs. 13 and 14).

The Student Performance Ribbon Chart is created by OLAP using Dimensions as features and measure as Count of Marks. Dimensions here represent Student ID. Hovering on the chart shows the count of marks for a particular Student ID.

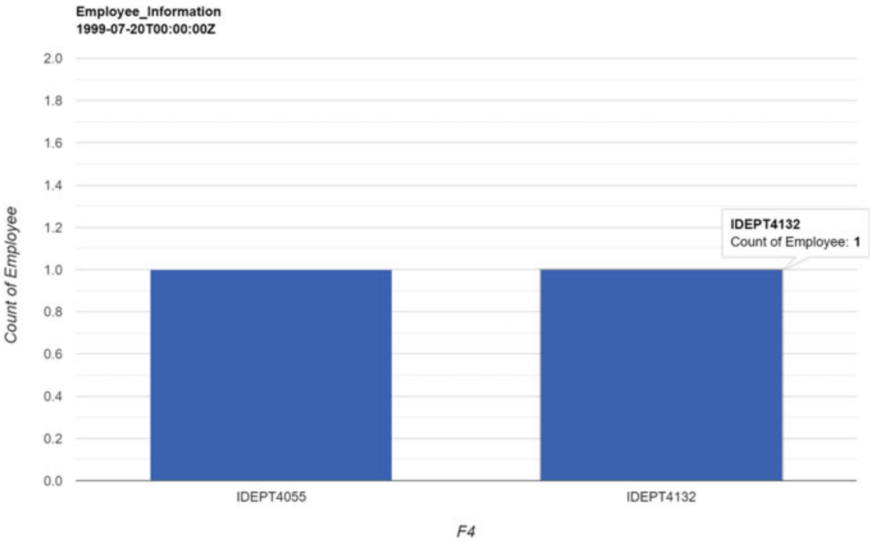


Fig. 13 Employee information chart in OLAP

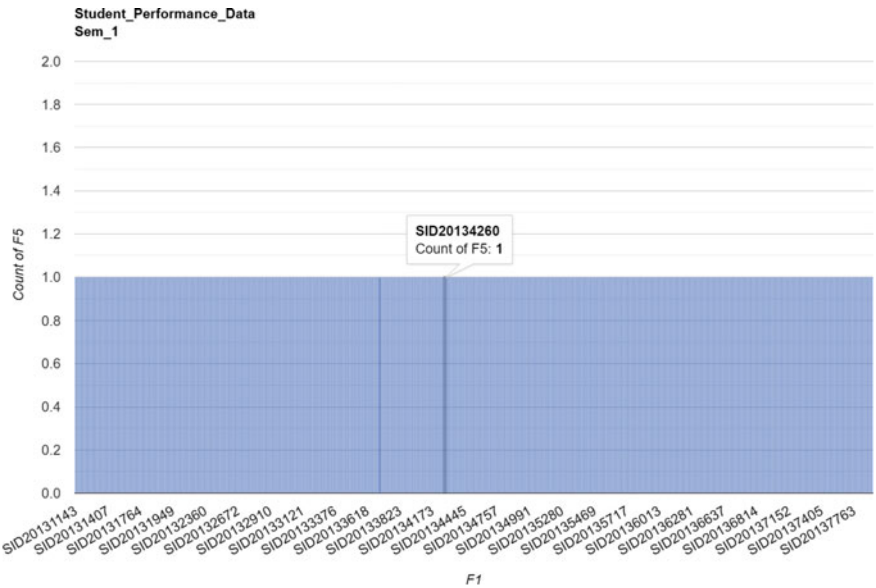


Fig. 14 Student performance chart in OLAP

Table 1 Comparison between the results

Basis	Power BI	OLAP Cube
Visualizations	Different Types of Visualizations can be performed	Limited Visualizations
Graphs	More Interactive	Less Interactive
Data Transformation	Easily done using Power Query Editor	Done by creating dimensions and measures
Insights	Greater insights	Lesser insights
User Interface	User friendly	Less user friendly
Schema	Relationship between dimensions can be easily displayed	One or more cubes in a single database that each are defined by one or more dimensions and measures
Sharing	Database can be shared within individuals or an organization	Database sharing is not easy
Security	Data is more secure	Less security

7 Results

According to the comparison above, Power BI offers more data visualization options and provides better data insights than OLAP Cube. Power BI's user interface is more engaging, allowing users to work more efficiently. In Power BI, a broader selection of type visualizations is accessible, allowing users to analyses data in a variety of ways. ETL (Extraction, Transformation, Loading) processes can also benefit from data transformation using Power Query Editor. Some features, such as slicers and Q&A, make it far superior to OLAP Cube. Power BI is a data analytics solution that can be used for business intelligence in a variety of enterprises. Sharing of data between organization members is easier and confidential data is secure within the organization by providing user access roles (Table 1).

8 Limitations and Future Studies

Future Work can be done on Machine Learning. A predictive machine learning model can be made to predict the results and statistical graphs like histograms, etc. can be created using machine learning libraries. Data Preprocessing is done on the complex dataset to get better visualizations and graphs. By using Machine Learning Algorithms to find out the best accuracy and results.

9 Conclusion

Studied and implemented a university database on Power BI and OLAP Cube to create visualizations, reports, and graphs. Gained insights on data and found schemas and relations between features using Data Analytics tools. Later the results are compared. By comparison, it can be concluded that Power BI is a better Data Visualization tool. Due to the variety of visualizations that can be accessed in Power BI, data visualization helps users gain better visual insights from the data. Choosing categories, values, fields, and other parameters can improve display accuracy and ensure that the user gets the desired results. Slicers, Q&A, and other functions help users quickly analyze and find results in large data sets. By applying filters to the visualization and scoping based on the characteristics of the data, obtaining insights from a large data set will be more enjoyable and simpler for users.

References

1. Geetha Bhargava M, Surya Kiran KTP, Koneru Lakshmaiah DRR Educational Foundation (2018) Analysis and design of visualization of educational institution database using power BI tool. Global J Power BI 1(4) Version 1.0
2. Suganda Girsang A, Alfa Sunarna D, Syaikhoni A, Ariyadi A (2019) Business intelligence for education management system. In: IEEE Xplore
3. Moscoso-Zea O, Castro J, Paredes-Gualtor J, Luján-Mora S (2019) A hybrid infrastructure of enterprise architecture and business intelligence and analytics for knowledge management in education. In: IEEE Xplore
4. Elhassan Ali FA, Klett F (2018) Learning analytics in higher education—the path toward educational management intelligent systems. In: 11th International conference on developments in eSystems engineering (DeSE), IEEE Xplore
5. Villegas-Ch W, Luján-Mora S, Buenaño-Fernandez D (2018) Towards the integration of business intelligence tools applied to educational data mining. In: IEEE Xplore
6. Desai S (2019) Dynamic induction model for student's behavior analysis. Int J Syst Softw Eng 7(1):13–19
7. Desai S, Roy S, Patel B, Purandare S, Kucheria M (2016) Very fast decision tree (VFDT) algorithm on Hadoop. In: Proceedings—2nd international conference on computing, communication, control and automation, ICCUBEA 2016, pp 7860037
8. Desai S, Patil ST (2015) Efficient regression algorithms for classification of social media data. In: 2015 International conference on pervasive computing: advance communication technology and application for society, ICPC 2015, pp 7087040
9. Diamond M, Mattia A (2017) Data visualization: an exploratory study into the software tools used by businesses. J Instruct Pedagogies 18
10. Bansal A, Upadhyay AK (2017) Microsoft power BI. Int J Soft Comput Eng (IJSCE) 7(3) ISSN: 2231–2307

A Comparative Analysis of Automatic Extractive and Abstractive Text Summarization

Madhuri Yadav and Rahul Katarya

Abstract With the advancement of the web, a large amount of data is being generated by people on the web. A shorter yet logical and consistent version of such a large amount of text is very beneficial to easily draw significant conclusions. Summarization is a way to derive a coherent and fluent summary. Summarization can be categorized as Extractive summarization and Abstractive summarization. In this paper, we compare a few Extractive and Abstractive methods of summarization on WikiHow-Dataset. The experiment results show superiority of abstractive summarization methods was observed over the extractive summarization methods when the average precision score was compared. Though pre-trained transformer technology was applied for abstractive summaries, even then, basic extractive summarization techniques performed well. This Analysis helps the researchers to understand the summarization process and performance comparison of the summarization techniques.

Keywords Abstractive summarization · Extractive summarization · Text summarization · WikiHow-dataset

1 Introduction

Before the twenty-first century, users of the web were only a consumer of information, but now every single user is a producer of data. In this century, it has become very common to share detailed information over social media by users. This information includes their thoughts, wishes, anger, emotions, opinions over the news, movies, people, and events happening in their lives. This amount of data from every single user is a lot to handle. With the advancement in sharing capability over social media comes the issue of handling such gigantic data. For a normal human brain, it is difficult to process such large data. Instead of there is a way to produce a synthesized

M. Yadav (✉) · R. Katarya
Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India
e-mail: madhuriyadav.me@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_46

version of data, then it will be easier to read, understand and draw conclusions. With text summarization, this large data can be handled and concluded.

Text summarization is a technique where we intend to generate a smaller synthesized version of the text covering all the important points. Text summarization is an activity of refining the important and significant. Some previous studies suggest the basic process of summarizing a source text involves the following steps: (i) analysis of the source text, (ii) spotting the significant points, (iii) bring down to a smaller version (iv) summary generation [1]. Summarization types are based on many factors type of input, type of output generated, etc. The most popular classification is Extractive and Abstractive summarization which is done on the basis of type of output summary.

An extractive summarization is a technique where sentences that holds some significance, phrases, etc. from the source document and concatenates the selected sentences to generate a distilled version. The importance of sentences is decided based on statistical and linguistic features of sentences [2]. Summaries generated using extractive methods [3] preserve the key text segments from the source text and by condensing the content to generate a shorter version. The deciding factor to point out significant key segments are done by using a number of features. This summary generation is comparatively easier than abstractive summarization.

An Abstractive summarization [4] goes deeper into grasping the notion of the text. It involves the thoughts and understanding of the generator. Automatic abstractive summarization uses linguistic methods to explore the deep notions and generate a shorter version by describing the concept using new phrases and segments.

As shown in Fig. 1 the extractive summary takes just important lines from the source document and generates the summary. While in Abstractive summary the whole new sentences are taken to generate the summary document. In this work, we have taken both extractive and abstractive summarization methods to compare and study. The highlights of the work are given below.

This paper majorly helps in the following ways:

(i) Understanding of Extractive and Abstractive methods of summarization.
(ii) Gives a basic understanding of the steps involved in summarization.
(iii) Give performance comparison of various methods for both the extractive and abstractive approaches.

2 Related Work

Here, we have discussed previous research that has been conducted in this area. We present the summary of related work in two parts: Summarization work in Extractive and Abstractive areas. Firstly the extractive summarization works are discussed and the Abstractive summarization works are discussed.

Extractive summarization is a way to extract a summary from a large text by picking up text segments and phrases that seem important. The authors did single document summarization based on proposed features was done [5, 6]. In [5], authors have done a multi-document text summarization in Marathi text. They used the

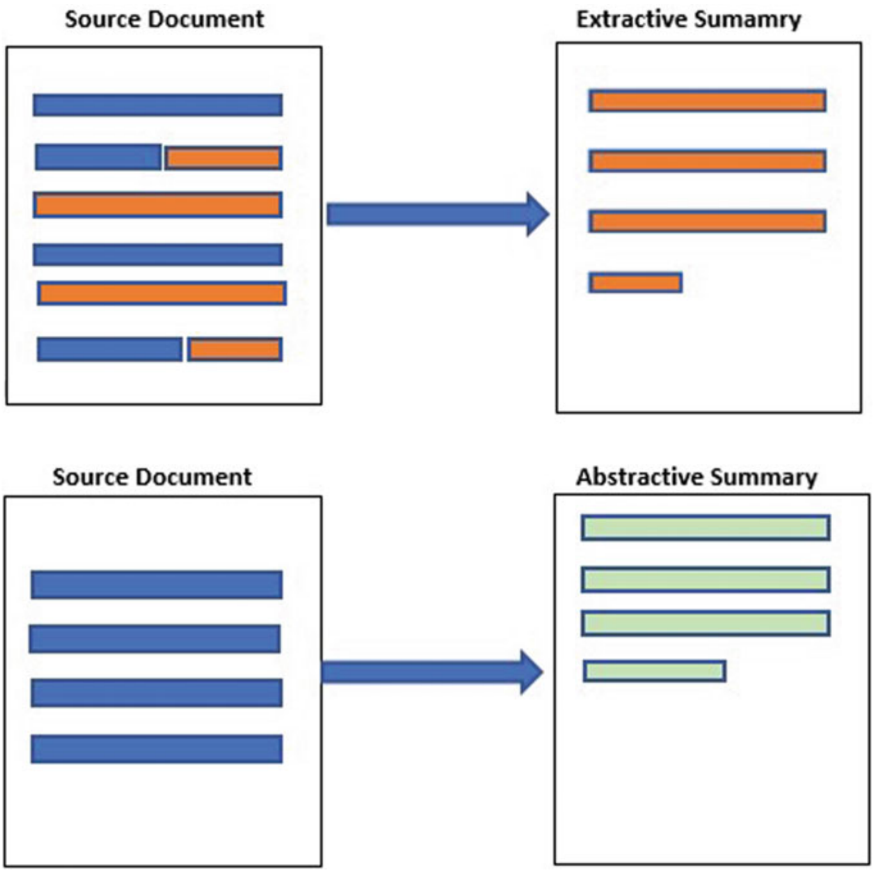


Fig. 1 Extractive versus abstractive summary source document

TextRank algorithm for extracting important features; TextRank is based on the PageRank algorithm. In the TextRank algorithm, the sentences are vertices of the graph, and the edge weights are given based on sentence similarity. In [6] authors used a semantic graph for sentence selection in summary. The method used a Logical Form(LF) parser to represent the sentence in the semantic graph. Gulati [7] proposed a novel technique for multi-document, extractive text summarization on Hindi documents. The input to the system was news articles on sports and politics. For each document, eleven feature values are calculated, and the final score was calculated. Using these scores, sentence selection is made for the summary generation. They combined these summaries, and finally, for summary generation, feature extraction is done. For evaluation of the approach in comparison with human-generated summary precision, recall and F-score were calculated. It was observed that the system achieved an average precision of 73% over multiple Hindi documents. Ganesan et al. [8] did abstractive summarization; they presented a novel framework of abstractive summarization called opinosis based on the graph-based approach. A graph is generated with each node being the unique word from the input sentences, and each directed edges contribute to the structure of sentences. [9] tried to overcome the limitation of [8] that being able to fuse the redundant sentences that should be collapsed together. [10] did topic-based abstractive summarization; they proposed a speech act guided approach for topic summarization for Twitter. They focused on Speech Act Recognition (SAR) to handle the Twitter data properly. They assumed the SAR as a multi-class classification problem, for they used SVM with a linear kernel. Yang et al. [11] the authors developed a model with Topic modeling and Reinforcement deep learning for Abstractive review Summarization. They proposed an end-to-end model with two subtasks where the first task is to capture learn domain-specific aspect and sentiment lexicon representations using a weakly supervised model called Latent Dirichlet Allocation(LDA). The second task is to generate the abstractive review summary. The learned representations from the previous task act as an input to a pointer-generator network which generates a summary using reinforcement learning. Rudra et al. [12] worked on the disaster-specific situational tweets. They divided the process into two-step in. First, they did extractive summarization using an existing summarization technique [13]. From the collected information, a bigram-based word graph is generated. To generate the abstractive summary, an optimization technique based on integer linear programming (ILP) is used. The authors also took noise and redundancy into consideration.

3 Summarization Process

In this section, we have discussed the methodology that we went along with to perform automatic text summarization. The objective was to analyze a number of Extractive and Abstractive Text Summarization techniques and compare them against each other to identify their advantages and shortcomings. In order to do that, the following steps were involved that are given in Fig. 2:

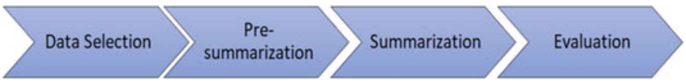


Fig. 2 Basic process followed for summarization task

3.1 Data Selection

For text summarization, there are many datasets that are available like TAC,[1] DUC,[2] Opinosis Dataset[3], WikiHow-Dataset[4], etc. The Dataset used was WikiHow-Dataset [19]. The WikiHow-Dataset consists of a collection of 230,000 articles. For each article human-generated summary is included in the dataset from an online knowledge base written by different human authors. For summary generation and comparison for this work, we have taken twenty articles to evaluate the result and compare it.

3.2 Pre-summarization

It is a very important step in natural language processing. The dataset taken is already well-structured. So only parser and tokenizer were used. For parsing, PlainText Parser was used, and for tokenization, Sumy Tokenizer was used.

3.3 Summarization

For summarization, Two approaches were selected: the Extractive and Abstractive approaches of summarization. The selected methods are discussed briefly.

3.3.1 Extractive Summarization

It is a method of summary generation; here, the sentences are chosen from source data based on their importance. The Sentences selected are directly added to the summary without any change. The final summary is a collection of sentences that entirely belongs to original input documents.

1. *Latent Semantic Analysis (LSA)* is based on a mathematical technique called Singular Value Decomposition (SVD) to point out the patterns of relationships between the terms and concepts. The basic concept is that the occurrence of the words in the same contexts will have similar meanings. LSA gives the weightage of documents belonging to different topics. Summarization is done by selecting the top N sentences from each topic depending on the weightage. This gives us

[1] https://tac.nist.gov/data/index.html.

[2] https://duc.nist.gov/data.html.

[3] http://kavita-ganesan.com/opinosis-opinion-dataset/#.XnCMuKgzZPY.

[4] https://github.com/mahnazkoupaee/WikiHow-Dataset.

a summarization where we get the highest weighted sentences from each of the topics for the summary [20].

2. *LexRank* [4] is a graph-based approach used for calculating the significance of the sentence. In this method, a connectivity matrix is created that works as an adjacency matrix that contains intra-sentence cosine similarity of sentences. Sentence selection is based on centroid sentence selection. A central sentence is selected, which is the mean of all the sentences. Based on this sentence other sentences are ranked.
3. *Luhn (TF-IDF)* [14] This is one of the earliest approaches to automated text summarization. The approach is based on TF-IDF, where the "significance" of a word is based on its frequency in the document. A significant word vector is made, and the sentences with more significant words have higher "scores." The final summary is a collection of the N highest-scoring sentences.
4. *Kullback–Leibler Sum (K-L Sum)* [15] In this approach, the main focus is on sentence selection for a summary generation. Here, the target summary is fixed to L words. The sentences are selected greedily based on the remaining size of the target summary (Less than L word) and the extent of unigram similarity to the original text. This method is based on the concept of Divergence, where we check to what extent a particular probability distribution differs from another distribution. If this divergence measure is low, similarity in summary and source document will be high.

3.3.2 Abstractive Summarization

Abstractive summarization a method where the summary is a distilled version of the source document. This generated summary should be as near as possible to a human-generated summary. Abstractive summarization creates a summary by rephrasing the input document by maintaining the original meaning of the document. The following are the model used:

1. *Generative Pre-Trained Transformer (GPT-2)* [16] It is a language-based method that predicts the next token when given an input token. This transformer is pre-trained on a very large corpus of English data. As it is an unsupervised model, it was pre-trained on the raw texts only without any pre-labeling or without any human supervision. Which makes it possible to work on publicly available data. The model a mask-mechanism to find out the next possible token on the basis of the previous input tokens. The inner representation of the English language is learned by the model to use further for feature extraction. The model works very well for Reading Comprehension, translation, summarization, etc.
2. *Masked Language Modeling (XLM—MLM)* is a prediction task where a blank token is called a mask token, given the token around the blank token. need. Masked language modeling is a language modeling technique to train the model without any human annotation. Various NLP tasks can be completed when using this pre-trained model. A BERT model is used, pre-trained on raw texts (no

Table 1 Performance table for extractive versus abstractive summarization methods

	Extractive summarization				Abstractive summarization	
	LSA	LexRank	Luhn	KL-Sum	GPT-2	XLM
Precision	0.5265	0.5605	0.5673	0.4463	0.6647	0.6642
Recall	0.2130	0.2121	0.1909	0.2229	0.1687	0.1689
F1- Score	0.2839	0.2906	0.2702	0.2725	0.2490	0.2490

human labeling), to make predictions and generate a summary. The transformer open-source library is used for this purpose [17].

3.4 Evaluation

Evaluation of the generated summary is the main task if we want to measure and compare the performance of the applied techniques. There are various evaluation metrics available one of them is ROUGE [18]. which is been used in this work. Evaluation metrics used ROUGE (Recall-Oriented Understudy for Gisting Evaluation) introduced in the Document Understanding Conference (DUC) Summarization task[1]. Rouge compares the quality of machine-generated summary that is produced by the summarization method against a set of summaries that are human-generated. ROUGE-1 is used for generating evaluation scores and comparisons. ROUGE-1 scores on the basis of overlap of unigrams of machine-generated summary and reference summary. Table 1 shows the mean values of the scores for twenty sample summaries.

4 Experimental Results and Evaluation

In this section, we have experimented with different models of Extractive and abstractive summarization and analyzed their results.

4.1 Analysis

For comparison, we have taken three parameters Precision, Recall, and F1-score. In Table 1, we have taken these for Extractive Summarization techniques: LSA, LexRank, Luhn, and KL-Sum. From the experiments, it has been observed that Summaries, KL-Sum visibly falls short in comparison to the other methods in this category in terms of precision and F1-score. There is very little difference between the scores of LexRank and Luhn; LexRank outperforms LSA. While comparing abstractive methods, it was observed that both the methods gave similar scores. Both the methods outperformed extractive methods considering the precision scores. Table 1 shows the mean values of the scores for twenty sample summaries for all the methods mentioned above for experimentation.

5 Conclusion and Future Work

Summarization is a facility that is extremely helpful for reaching a conclusion. We noticed that a tremendously large amount of data is available over the internet to process. Automatic summarization makes it possible to process such a large amount of data by saving time and energy. We analyze two approaches of summarization extractive and abstractive methods with six different algorithms. Experimental results show that even if pre-trained transformers were used for abstractive summarization, basic extractive summarization methods performed quite well. While observing Extractive Summaries, KL-Sum visibly falls short when compared to the other methods in this category. There is very little difference between the scores of LexRank outperforms LSA. For abstractive summaries, both methods performed quite equally. In the future, we plan to work more on the unstructured data, and also, we plan to improve the summary generation for abstractive and extractive summarization algorithms.

References

1. Maybury M (1995) Generating summaries from event data. Inf Process Manage 31:735–751
2. Gupta V, Lehal G (2010) A survey of text summarization extractive techniques. J Emerg Technol Web Intell 2
3. Khosravi H, Eslami E, Kyoomarsi F, Dehkordy P (2008) Optimizing text summarization based on fuzzy logic. In: Proceedings of the seventh IEEE/ACIS international conference on computer and information science (Icis 2008). pp 121–130
4. Radev D, Hovy E, McKeown K (2002) Introduction to the special issue on summarization. Comput Linguistics 28:399–408
5. Sarwadnya V, Sonawane S (2018) Marathi extractive text summarizer using graph based model. In: 2018 fourth international conference on computing communication control and automation (ICCUBEA)
6. Banu M, Karthika C, Sudarmani P, Geetha T (2007) Tamil document summarization using semantic graph method. In: International conference on computational intelligence and multimedia applications (ICCIMA 2007)
7. Gulati A, Sawarkar S (2017) A novel technique for multidocument Hindi text summarization. In: 2017 international conference on nascent technologies in engineering (ICNTE)
8. Ganesan K, Han J, Zhai C (2010) Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions. In: Proceedings of the 23rd international conference on computational linguistics. Beijing, pp 240–248
9. Bhargava R, Sharma Y, Sharma G (2016) ATSSI: abstractive text summarization using sentiment infusion. Proc Comput Sci 89:404–411

10. Zhang R, Li W, Gao D, Ouyang Y (2013) Automatic twitter topic summarization with speech acts. IEEE Trans Audio Speech Lang Process 21:649–658
11. Yang M, Qu Q, Shen Y, Lei K, Zhu J (2018) Cross-domain aspect/sentiment-aware abstractive review summarization by combining topic modeling and deep reinforcement learning. Neural Comput Appl 32:6421–6433
12. Rudra K, Goyal P, Ganguly N, Imran M, Mitra P (2019) Summarizing situational tweets in crisis scenarios: an extractive-abstractive approach. IEEE Trans Comput Soc Syst 6:981–993
13. Rudra K, Ghosh S, Ganguly N, Goyal P, Ghosh S (2015) Extracting situational information from microblogs during disaster events. In: Proceedings of the 24th ACM international on conference on information and knowledge management
14. Luhn H (1958) The automatic creation of literature abstracts. IBM J Res Dev 2:159–165
15. Haghighi A, Vanderwende L (2009) Exploring content models for multi-document summarization. In: NAACL '09: proceedings of human language technologies: the 2009 annual conference of the North American chapter of the association for computational linguistics. pp 362–370
16. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language models are unsupervised multitask learners
17. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Davison J, Shleifer S, von Platen P, Ma C, Jernit, Y, Plu J, Xu C, Le Scao T, Gugger S, Drame M, Lhoest Q, Rush A (2020) Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations
18. Lin C (2004) ROUGE: a package for automatic evaluation of summaries. ACL-2004
19. Koupaee M, Yang Wang W (2018) WikiHow: a large scale text summarization dataset. CoRR
20. Dumais S (2005) Latent semantic analysis. Ann Rev Inf Sci Technol 38:188–230

Secure Publisher-Subscriber Model on IoT Nodes Using Lightweight Authentication Algorithm

Gudapati Ramyasri and Shaik Jakeer Hussain

Abstract The Internet of things is getting developed in many areas like smart homes, healthcare, smart building, and industrial-related applications. For transmission of these data, IoT requires a lightweight protocol like MQTT for these purposes. MQTT will not be the best security protocol. Secure data transmission is essential. Confidentiality, Integrity, and Accessibility are the security functions that ought to be maintained. MQTT mainly presents the Transport layer security on behalf of the Transmission control protocol suitable for IoT applications with lightweight requirements. This paper mainly contains the secure transmission of data using payload encryption and data encryption before transmission.

Keywords Internet of things · Lightweight · MQTT · Security · Authentication

1 Introduction

Recent advances in the development of the modern ecosystem of the Internet and communications technology have led to the development of smart data-driven applications of the Internet of things (IoT) [1]. As IoT cannot be implemented using regular internet architecture and communications protocols, the design of a lightweight protocol with a secure authentication-based encryption algorithm is mandatory [2, 3]. In message queue, telemetry transport protocol communication between devices is done through an intermediary device called a broker [4, 5]. Figure 1 shows the Publisher to subscriber communication model for data transmission between publisher and broker through a broker.

MQTT is a lightweight protocol with a publish-subscribe model containing three main entities of publisher, subscriber, and broker. Publisher and subscriber do not communicate with each other and cannot transmit data directly. They exchange data with each other through a broker based on hosted topics. The broker acts as a central authority for the exchange of data between other entities [6, 7]. Data transmitted

G. Ramyasri (✉) · S. J. Hussain
School of Electronics, VFSTR, Guntur, India
e-mail: ramyasri.gudapati@gmail.com

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_47

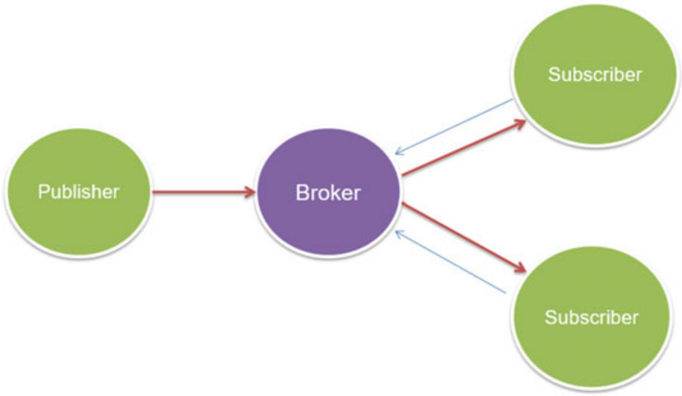


Fig. 1 Publisher-subscriber model

from the publisher is publication and request of particular topic data from the broker is a subscription. In MQTT, secure transmission of data and privacy is a big question mark. Publisher-subscriber model with secure protocols will provide more beneficial effects on IoT [8, 9]. As traditional security algorithms take more period, MQTT should be implemented with a lightweight encryption algorithm and authentication model for the secure transmission of data and authentication of devices.

In general IoT applications like healthcare, home automation, and industrial applications data from the sensors are transmitted to the broker as publications and brokers transmit the data to subscribers. Here a group of sensors transmits the data to the gateway and the gateway acts as a publisher [10, 11]. Due to security threats, MQTT is used in limited private security networks. Usage of MQTT in public Third-party networks is not secure. Some authors had implemented MQTT protocol with SSL and TLS features which lags in the confidentiality of communication paths. Due to this implementation of MQTT with payload encryption is required. Still to provide trustworthiness from broker to subscriber or from publisher to broker Authentication of users is required by reducing malicious users and middleware attacks [12, 13].

Many authors have done their research work on providing security for IoT data using basic TLS/SSL features which are not suitable for high amounts of data transmission. Some others have done the research work on security protocols missing the authentication of user and transmitting less secured data [14, 15]. This paper [12] has done research work on authentication and access control of data and data during transmission is not encrypted. Due to this reason, there are channels of hacking data during transmission. These will result in the several confluences.

Basic Traditional Encryption Algorithm mainly requires Authentication and Encryption algorithm which is very complex and those algorithms cannot implement end-to-end encryption with this decoupled publisher and subscriber model [16, 17]. This paper mainly presents lightweight encryption and authentication model with key management protocol. Here data is getting transmitted in an encrypted

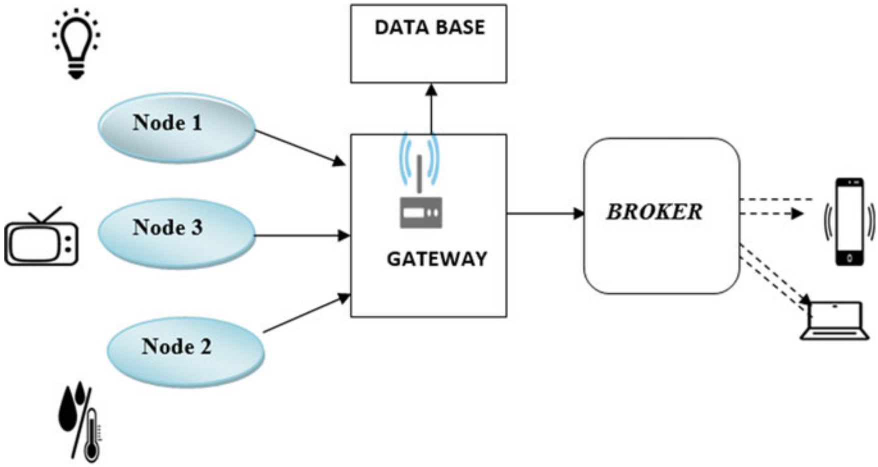


Fig. 2 MQTT protocol using gateway

form between peer-to-peer nodes. This results in the secure transmission of data which increases the confidentiality, integrity, and accessibility of data.

All the sensors are considered unique publishers it is consuming more energy. So one category-related data from sensors can be sent from one gateway which results in the reduction of time consumption and energy consumption for transmission of data. Figure 2 shows the implementation of the MQTT protocol using the gateway where data from all sensors are collected by the gateway and from the gateway they are sent to the broker as a publisher [18, 19].

This paper mainly works on the secure transmission and authentication which is suitable for the end-to-end communication between nodes as follows.

- Implementation of advanced protocols is done instead of basic TCL/SSL features which results in the decrease of the bandwidth.
- Authentication algorithm results in the secure access to the stored data which results in the reduction of many confluences and increase in the transmission rate.
- Implementation of this algorithm will result in the reduction of storage size and transmission rate.

2 IoT Security Threads

IoT is facing many issues related to data transmission. A limited amount of security protocols in IoT will produce more data manipulation and transaction attacks compared to traditional internet usage [20]. IoT threats are the main reason for data transmission attacks. Here attackers will try to gain unauthorized access and will do the actions without access rights.

2.1 Authentication

Due to insecure authentication for devices, users and lack of credential data fulfilled for authentication more unauthorized users can access data. Even authorized users can access more prevailed data other than the prescribed limit. Unprotected data of devices, password recovery, and lack of strong passwords will result in ineffective authentication.

2.2 Interfacing

Insecure interfacing to the IoT devices is mainly done due to lack of data credentials, weak passwords, ineffective parameters, fake accounts, insecure storage of account data.

2.3 Services

IoT is providing an insecure interface to the network services and providing denial of service, DoS through the network interface. These all will provide authorized access to the data and particular accounts usage.

2.4 Encryption

Due to the plain text transmission from the IoT device to the user, intruders will gain free access to the data that is getting transmitted, IoT devices, and also the data storage location.

2.5 Privacy

Insecure storage of personal or sensitive information. This results in the lack of data protection either during data collection or data transmission. Privacy issues also include insecure interfaces with the cloud data storage and also IoT devices. Due to unauthorized access and unencrypted data transmission, intruders can get access to IoT devices and data [21].

2.6 Security

The lack of a strong password and lack of permissions for interfacing with IoT devices will result in the IoT security threads of unauthorized access. The main reasons for the insecure configuration of IoT devices are the lack of monitoring, lack of strong passwords, and lack of authorized permission levels [4].

2.7 Firmware

Lack of insecurity for the firmware that is getting updated frequently. Attackers will wait for the updated files where sensitive information is getting stored. As Firmware is not concerned with security, the update of firmware will not give any change. Firmware thread is mainly due to the insecure update, lack of encryption for data files that are getting updated, and verification of the stored files during retrieval.

3 Related Work

Many researchers have developed different security protocols in MQTT with the reduction of confidentiality and Authentication. Let's compare the drawbacks of those protocols.

SSL/TLS Based Security

SSL/TLS based security in MQTT will provide Authentication of the user based on the authorization certificate key along with the signature scheme. Here the drawback is the reduction in the confidentiality of Topics data that is getting transmitted. The data is not encrypted before transmission [5].

Access control

Other users have implemented different access control mechanisms of attribute-based access control (ABAC), role-based access control (RBAC), and policy-based access control (PBAC). Here access control of subscribers to the particular topic is authorized but data that is getting transmitted is not encrypted [9].

Basic Encryption Algorithm

Many Authors have implemented peer-to-peer encryption algorithms where the sender encrypts the data and sends the encrypted data to the receiver [3]. Here the publisher and broker require the direct communication path without any intermediary broker. This is not satisfactory for the MQTT protocol.

Attribute-based encryption Algorithm

The attribute-based encryption algorithm is implemented as an asymmetric AES algorithm on payload using a third party. Here it is very complex.

4 Publisher-Subscriber Security Model

Publisher-Subscriber should have authorized security protocol for broker, publisher, and subscriber. This paper mainly presents the below-mentioned security protocols of data encryption using Asymmetric cryptography with the private key, public key, and Payload encryption with symmetric block cipher security protocol [9].

4.1 Publisher-Subscriber Cryptographic Organization

This paper mainly publishers and subscribers of the MQTT protocol will have a unique cryptographic ID that is used for authentication purposes [13]. This Id cannot be misused or can be hacked or can be guessed as it is random no which is in a hexadecimal form. This is used for authentication and communication between the publisher and broker as well as between the subscriber and broker.

4.2 Broker Cryptographic Organization

The broker must have a cryptographic Authentication ID manager for the transmission of data between the broker and other clients. The broker must face problems during interaction with the clients. Due to frequent and continuous requests of data-heavy flow and overlaps might occur. The broker must manage these simultaneous operations with heavy speed.

4.3 Key Management

The broker must maintain a security protocol for interaction. For this data encryption is done using the asymmetric cryptographic algorithm and authentication is done on payload encryption using symmetric block cipher cryptographic encryption algorithm [19].

The key pairs of the public key, private key, and random number that are generated for each client will act as an identity for authentication and are present at the devices internally. Here the private key is present at the device itself and used during decryption or encryption of data. Random Keys are encrypted and sent during transmission for identity and authentication. This reduces unauthorized access and man-in-the-middle attacks [10]. The public key will be present at the repositories and will be linked up with the unique Id of the Device. Cryptographic ID will link publisher and subscriber pairs with the broker public key. This will provide secure communication between the clients and also represents the unique one-time transmission of data from publisher to broker and also broker to the subscriber [16].

4.4 Authentication

In this Authentication method, the connection between the client and broker is formed by the connection request from the client to the broker. For this connection, the client sends the publisher a unique node ID and encrypted form of data along with random key Rk_a. Using a Unique node ID broker will get the public key of the publisher from public key repositories. Using the public key of the publisher, the broker will decrypt the message. By comparing decrypted data, unique node Id and first message unique node ID broker will confirm the publisher has encrypted data with its private key.

The broker will send the two parts in a message. One part will encrypt data with the private key of the broker and the second with the public key of the client. First encrypted data will contain the unique node ID of the broker and random key sent for connection from the client. The second part will contain the unique node ID of the client plus the random key received from the client and another two random keys of Rk_p and Rk_s. Clients will receive the data decrypts the first part with the public key of the broker and the second part with the private key of the publisher. It confirms that it had received it from the broker. For verification of the broker, the publisher resends the second part of the message encrypted with the public key of the broker. The broker will decrypt it with its private key and confirm the data sent to the publisher is correct. Now it confirms the mutual connection between the publisher and broker (Figs. 3 and 4).

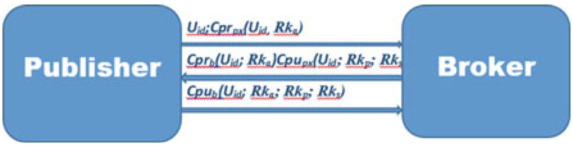


Fig. 3 Publisher-broker mutual authentication

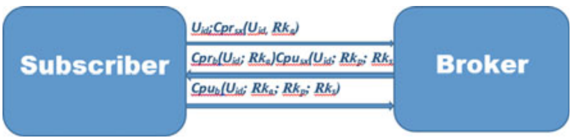


Fig. 4 Subscriber-broker mutual authentication

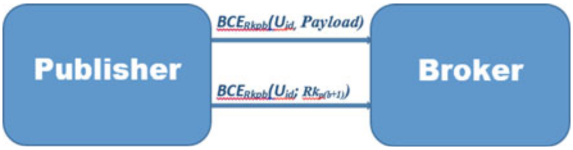


Fig. 5 Publisher-broker block cipher encryption

In the same way as authentication between the publisher and broker. Mutual authentication between the subscriber and broker is also done using the unique node Id of subscriber and broker. The private key of the subscriber is used for encrypted transmission of the Random key and unique node ID. Confirmation of authentication from broker to the subscriber is sent using the private of broker and public key of subscriber. The broker will also encrypt and send random keys used for data transmission. After verification of unique node ID and Random keys Authorized Authentication is formed between subscriber and broker.

4.5 Payload Encryption

Payload encryption is mainly done using the symmetric block cipher encryption security protocol which consumes less amount of time compared to other encryption technologies like asymmetric encryption (Fig. 5).

Data from the publisher to the broker is sent through the encrypted form payload which contains the unique node ID of the Publisher along with the data. This encryption is done using symmetric block cipher technique and Random key which is known only to the publisher and Broker (Fig. 6).

In Authentication three steps are followed for connection formation. In the first step Random key along with a unique node, Id is sent from the publisher to the broker for authentication using the private key of the publisher. In the second step, the Broker will send the publisher unique node ID with the random key of authentication Rk_a in an encrypted form with the private key of broker and along sends random keys of Publisher and subscriber, i.e., Rk_p and Rk_s with the unique node ID of broker encrypted with the private key of the broker. The broker will verify it based on the encrypted message sent from the publisher containing Unique Id, $Rk_{a,}Rk_{p,}$ and Rk_s. so these Keys are known only to the publisher and broker.

Data from the publisher to the broker is sent in the form of a payload along with the unique node ID of the publisher encrypted with random key $Rk_{pb.}$In the

Fig. 6 Subscriber-broker block cipher encryption

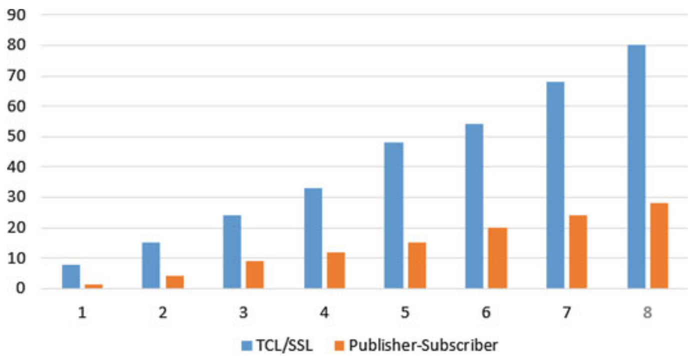


Fig. 7 Message size consumption for communication

acknowledgment, the broker sends the Unique node Id of the publisher along with another random key $Rk_{p(b+1)}$ generated by the cryptographic system. These two are encrypted using the block cipher technology with a random key of $Rk_{pb.}$ This $Rk_{p(b+1)}$ is used as a random key for the next transaction between publisher and broker.

Block cipher encryption between subscriber and broker is done on payload using the unique Id of subscriber and broker. Random key *Rksb is* used as an encryption key for transmission of encrypted payload and a unique ID. Updated random key $Rk_{s(b+1)}$ of is used for encryption during the next transaction of payload.

The transmission of the data is done in an encrypted form in MQTT form which results in the reduction of message size. Figure 7 shows the analysis of message size for each transmission on monthly basis. The size of data that is getting transmitted with proposed publish-subscribe model is consuming less amount of size compared to the TCL/SSL-based transmission. Here the data is considered based on the total unique data transmitted for each month.

5 Conclusion

This paper mainly presents the encryption and Authentication model which will control the access of data from IoT devices using a secured lightweight MQTT protocol. Without any modifications of MQTT protocols, additional features are added which result in providing confidentiality, integrity, and access control of IoT data from sensors and users for smart devices for analysis. Asymmetric Elliptic curve cryptography is used for Authentication and symmetric block cipher encryption algorithm is used for payload i.e. data encryption.

References

1. Luca R et al. (2016) A semantic publish-subscribe architecture for the Internet of Things. IEEE Internet of Things J 3.6:1274–1296
2. Ravi Kishore K, Ramya Soratkal S (2016) MQTT based home automation system using ESP8266. In: 2016 IEEE region 10 humanitarian technology conference (R10-HTC). IEEE
3. Ismail B, Österberg P, Song H (2019) Security of the internet of things: vulnerabilities, attacks, and countermeasures. IEEE Commun Surv Tutorials 22.1
4. Nasi T et al. (2017) MQTT-topics management system for sharing of open data. In: 2017 international conference on digital arts, media and technology (ICDAMT). IEEE
5. Gudapati R, Sharma S (2017) Optimization of energy exhaustion for Smart Camera Network in vision task. In: 2017 third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE
6. Chen F et al. (2020) A study on MQTT node selection. In: 2020 16th international conference on mobility, sensing, and networking (MSN). IEEE
7. Chen F et al. (2020) A review on the study on MQTT security challenge. In: 2020 IEEE international conference on smart cloud (SmartCloud). IEEE
8. Meena S et al. (2015) Secure MQTT for the internet of things (IoT). In: 2015 fifth international conference on communication systems and network technologies. IEEE
9. Prabaharan J et al. (2017) Wireless home automation and security system using MQTT protocol. In: 2017 2nd IEEE international conference on recent trends in electronics, information and communication technology (RTEICT). IEEE
10. Al-Fuqaha A et al. (2015) Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun Surv Tutor 17.4:2347–2376
11. Chintan P, Prafulchandra Doshi N (2020) Secure lightweight key exchange using ecc for user-gateway paradigm. IEEE Trans Comput
12. Urs H, Truong HL, Stanford-Clark A (2008) MQTT-S—A publish/subscribe protocol for wireless sensor networks. In: 2008 3rd international conference on communication systems software and middleware and workshops (COMSWARE'08). IEEE
13. Wei-Tsung S, Chen W-C, Chen C-C (2019) An extensible and transparent thing-to-thing security enhancement for mqtt protocol in IoT environment. In: 2019 global IoT summit (GIoTS). IEEE
14. Tetsuya Y, Sasaki Y (2016) Comparison with HTTP and MQTT on required network resources for IoT. In: 2016 international conference on control, electronics, renewable energy, and communications (ICCEREC). IEEE
15. Dikii Dmitrii I (2020) Remote access control model for MQTT protocol. In: 2020 IEEE conference of russian young researchers in electrical and electronic engineering (EIConRus). IEEE
16. Ramyasri G (2021) Smart camera network supervision for competent exploitation of energy recourses in vision task. APTIKOM J Comput Sci Inform Technol 6(1):67–74
17. Ousmane S, Ngom I, Lishou C (2019) Lightweight security scheme for MQTT/MQTT-sn protocol. In: 2019 Sixth international conference on internet of things: systems, management, and security (IOTSMS). IEEE
18. Yanqi Z et al. (2018) Secure pub-sub: blockchain-based fair payment with reputation for reliable cyber physical systems. IEEE Access 6:12295–12303
19. SeongHan S et al. (2016) A security framework for MQTT. In: 2016 IEEE conference on communications and network security (CNS). IEEE
20. Syed TS, Safdar GA (2019) Energy-efficient GCSA medium access protocol for infrastructure-based cognitive radio networks. IEEE Syst J 14(1):288–297
21. Ozansoy CR, Zayegh A, Kalam A (2007) The real-time publisher/subscriber communication model for distributed substation systems. IEEE Trans Power Delivery 22(3):1411–1423
22. Aimaschana N et al. (2016) Authorization mechanism for MQTT-based Internet of Things. In: 2016 IEEE international conference on communications workshops (ICC). IEEE

Survey of the Data Vortex Switch Next-Generation Architectures

Amrita Soni and Neha Sharma

Abstract Today high-performance computing systems (HPCS) and large Data Center Networks (DCNs) are experiencing exponential growth in the high network traffic due to several emerging applications such as mobile data consumption, multimedia information, and cloud computing. To deal with this high data rate traffic, DCNs and HPCS require thousands of servers that are interconnected with high bandwidth capacity switches. All-optical interconnection networks could provide a viable solution to these applications by offering high throughput, low latency, and reduced power consumption.

Keywords Interconnection network · Original Data Vortex · Augmented Data Vortex · Cluster DV · High-Performance Computing

1 Introduction

Interconnection Networks are of high importance in all large-scale high-performance computing (HPC) workloads, core routers, supercomputers, and high-performance computing systems (HPCS) to get high bandwidth, high throughput, and low latency [1]. Mostly, the core routers are used in telecommunication whereas HPCS are in parallel and distributed systems. Buffering, scalability, and latency are very important design parameters to achieve high bandwidth in OPS networks. It is suggested that to take maximum advantage of fiber optic technology, the use of optical buffering, switching algorithms and complex routing should be avoided. To provide above-stated requirements the interconnection network namely the Data Vortex network presents a good example that fulfills the requirements of HPCS applications [2–6].

A. Soni (✉) · N. Sharma
Department of Electronics and Communication, Ujjain Engineering College, Ujjain, M.P 456010, India
e-mail: amuujn@uecu.ac.in

N. Sharma
e-mail: nehatripathi@uecu.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_48

Currently, workloads like data analytics or working on irregular data structures seem to be very important. The reason behind this fact is, these applications require pointer-based or link list-based data structures to save data and present irregular data structures with unpredictable access patterns. Hence, a new interconnection network (IN), DV is proposed which fulfills both traditional as well as irregular system requirements [7].

Objective. The main objective of this survey paper is as follows:

- A detailed study on traditional and next-generation architectures and to explore the key role of DV in the implementation of HPC applications and large Data Center Networks.

This paper presents a survey of some important next-generation architectures of DV switch that have been recently presented in the research literature, as suitable all-optical interconnection networks. Section 2 presents Background of the Data Vortex Switch architecture. Section 3 presents the next-generation architectures, and finally, Sect. 4 concludes the findings of the survey and outlines for future work.

2 Background of Data Vortex Switch Architecture

In paper [3, 8], the numerical simulations of the DV for various performance parameters are discussed. Further analyses of results are obtained from an experimental setup by multi-hop WDM packets routing in a re-circulating test-bed [3]. DV architecture employs a hierarchal switching system design where the WDM technique is used with header encoding to reduce the header processing and routing latency at each hop [4]. Data Vortex architecture has been studied by applying non-uniform and bursty traffic conditions [6]. In DV, optical crosstalk was a big hurdle to use OPS technique. To reduce this problem, in reference [9], it has been demonstrated that WDM 10 Gigabits/sec packet routing can be possible through two fully functional nodes of Data Vortex [9]. A generalized algorithm is proposed to find all possible routes and pathways between every source–destination pair in the data vortex switch. This algorithm can be used for various applications like survivability and quality of service implementations [10]. Low BER was a great issue. To resolve this problem in reference [11], 72 SOA-based cascaded switching nodes have been studied by an experimental setup [11]. In another experiment [12], on 8 node sub-network, realistic traffic is emulated by using the SPLASH-2 parallel computing benchmark [12]. A new WDM all-optical packet switching node is presented for distributed routing deflection networks [13]. The correct routing behavior for 14 channel WDM packets has been verified by implementing a 12 port DV OPS with 36 interconnected nodes and latencies below 60 ns have been obtained [14]. An OPS interconnection network is introduced, which is comprised of 36 silicon optical amplifier (SOA) based switching nodes. For 12 input ports to 12 output ports, it displays complete packet routing functionality [15].

A new experiment is performed on slot timing margins and latency characteristics in a fully implemented 12 port optical packet switching network. It is investigated that due to switching transients, the packets are routed shortly by 400 ps for each node [16]. A novel FDL-based packet injection control module (ICM) is designed which mediates injection of packets into OPS routers. It is demonstrated that ICM can resolve contention in optical networks [17]. A new experiment has been performed to measure receiver power penalty, optical signal to noise degradation, and dynamic optical power range for multiple wavelength packets in 12 $\times$ 12 DV-based OPS architecture [18]. It is investigated that for 58 hops with 8 channels, 10 Gigabits/s per channel payload covering 24.2 nm of the C-band, BER of 10^{-9} can be maintained [19]. In this experiment, the calculation of clock to data skew tolerance window of one hundred and 50 picoseconds is measured using an embedded clock signal. Thus in large-scale computing systems by the use of the high degree of parallelism provided by Wavelength Division Multiplexing, the time of processor/memory access can be reduced [20]. Two experiments are performed in order to support the flexibility of optical packet format in the OPS network. In these experiments, payloads of varying wavelength, timing offset, payloads of varying size, varying duration, and format are generated [21]. A large network size cannot always improve the performance, but increasing the total count of angles and also a few amounts of them used for input and output, improves the performance of system. This also reduces the requirement of total number of nodes. This type of system can achieve low average latency and 99.9% acceptance under realistic traffic loads [22]. In this test-bed, researcher found that in a single node a very small generated PDG accumulates ASE noise which can affect the optical OSNR deterioration [23].

For the first time, the equivalent chained planar (MIN) model which is a planar diagram of the 3-D model network is presented, which simplified the performance analysis of the entire network and also its topological comparisons with other chained MIN also [24]. In DV switch, fault tolerance and reliability analysis have lacked attention. The fault tolerance and reliability of ADV were evaluated and presented in [25]. In reference to [26], by using the simulation method, the angle value of Data Vortex is studied. It is investigated that angle values show tremendous effects on network performance. A novel buffer architecture is discovered in [27]. This modern architecture is supported with individualistic read and write processes without affecting the rest of the network [27]. The parameters like acceptance rate, median latency, 99.9% latency are compared by using simulation software. The system is simulated with and without ICM. For 64 ports DV IN (interconnection network), injection strategies and input port populace policies are also studied. These two parameters are controlling the acceptance rate and latency of 64 ports DV [28].

In [29], the result of the primary DV switch is compared with the result of the new ADV switch. After comparing both the results it is found that the augmented data vortex switch with increased fault tolerance also improves the latency while throughput is almost similar. In [30] a new node design is proposed in which a combined time and space domain conflict resolution method is used. In DV switch, the results of various parameters are studied under uniform and non-uniform traffic, are also presented by simulation method [31]. To achieve a high level of performance for end-to-end communication, Avg. delay must be minimized. If the network is designed with high message acceptance and low Avg. delay penalty then the complete potential of photonic system can be used. For a photonic network like DV, clustering is the most suitable method [32]. A review on OPS interconnection network which is based on the various performance parameters of DV architecture is done in reference [33]. In reference to [34], the latency performance of the ADV switch under worst faulty condition is evaluated. In [35], earlier fault tolerance evaluation performed for 3×3 ADV is now evaluated for 4×4 DV and compared. For DV, ADV, and 4×4 DV the cost-effectiveness is also calculated [35]. In reference to [36], a new K-ary scheme is introduced. In this paper, the 4-ary decoding system and regular binary decoding systems are compared.

The Data Vortex architecture shows a significant effect of angle parameters on accepted traffic and average latency. A comparative study of DV architecture has been done with angle parameter $A = 1$ and angle parameter $A > 1$ [37]. A modified K-ary decoding scheme has been proposed by Q. Yang for DV architecture. Here, the main focus is on the study of difference between the network behavior of binary Data Vortex network and 4-ary (where $k = 4$) network [38]. In [39], K-ary DV network with node buffering and without node buffering has been studied. In [40], for the first time proposed and evaluated the terminal reliability of DV switch and compared it with ADV switch. In [41], DV architecture is proposed in a backward (reverse) direction named as reverse data vortex (RDV) switch. In RDV architecture, control logic, hardware model, and modified priority scheme are discussed. To enhance the network performance, in [42] it is shown that three input node design in Data Vortex is advantageous as compared to two input nodes. In reference [43] Q. Yang has proposed a new routing node in Data Vortex architecture, to provide speedup and parallel processing capability. In reference to [44], original three-dimensional design structure is converted into multiple parallel planes. Here author explored an alternate layout which permits the routing path arrangement to be implemented as multiple parallel planes. In [45], the Q. Yang et al. have proposed modular DV buffered and non-buffered designs. A new method of switching named as hybrid switching [47] explored to ensure good quality of service (QoS) in OPS interconnection networks. In the reference [46], bidirectional data vortex is proposed. A new path synchronization strategy in DV interconnection networks that contain reduced physical delay.

3 The Next-Generation Architectures

3.1 The Cluster Architecture of Data Vortex Network

Most of the current large Data Center networks (DCNs) are considered to be hierarchical architectures, in a way that they have group of server racks, forming clusters. In a common DCN architecture, the arrangements of servers are done into racks and interconnected to top of rack (ToR) switches. The ToR switches manage the inter-server communication within the rack. ToR switches are interconnected by an intra-cluster optical switch, which forms a small size cluster network. High-capacity inter-cluster optical switches are capable to provide connections to several small size cluster networks. A hierarchical DCN architecture was required with the capability of managing two types of traffics, intra-cluster communication traffic among racks and varying inter-cluster traffic. From previous research, we know that a single DV architecture treats all the large number of I/O ports with uniformity and manages all traffic with similar process. It can also handle bursty and spatial non-uniform traffic very smoothly. In DCN architecture to accommodate all above-stated characteristics, a new cluster architecture of DV is proposed [48], which can support faster communication for heavier traffic with adaptability to traffic locality patterns.

This section presents an overview of the cluster architecture of DV network. It is well explained with 16 clusters/groups of servers in ref [48]. Here, each cluster is of similar size. It has 64 servers or I/O points that communicate heavily within the group. As in each cluster, servers have 80% of the traffic load for other servers, the locality parameter is observed to be as 0.8, and also to reach other clusters 20% inter-cluster traffic has routed across a separate sub-network.

This architecture has two types of sub-networks.

DV-L sub-network: It handles traffic within each cluster. This architecture has 16 DV-L sub-networks. Each DV-L has height of $\mathrm{H_L} = 64$ and angle $\mathrm{A_L} = 2$. To match the cost of the networks, $\mathrm{A_L}$ in the cluster network and A in the single Data Vortex network can be varied. To make network simple Ain $= 1$ is taken. The total number of cylinders used in the DV-L sub-network is given by $C = \log_2 H + 1$ because before the exit of packet from optical domain an extra cylinder is provided.

DV-G sub-network: It handles global traffic outside of the cluster. Here $H_G = 16$ and $A_G = 64$ are taken to sort the traffic to the correct cluster. While $A_G = 64$ is taken for multiple entry points based on the total number of links in and out of the DV-G sub-network. The total number of cylinders is given by $C = \log_2 H$ used in DV-G network because before entering the DV-L network the packets remain in the optical domain.

As shown in Fig. 1, global or local (G/L) decision block separates the local cluster traffic and global traffic. The traffic flow and interconnections between the DV-L and DV-G sub-networks can be understood by labeling the packet target address with (cluster, height) indexes, respectively. When packet enters the DV-L sub-network, the local traffic index is given by (cluster, height, local) with the required header bits. These header bits are encoded in the physical layer. When packet enters the DV-G sub-network, they follow different input interface with the indexes of global traffic (cluster, height, global). In this sub-network, they identify the global traffic by decoding the corresponding header bits attached to the cluster index. After the separation of traffics an E/O interface can be set up, in which very small number of header bits is needed to distinguish its local height address for local traffic.

When traffic goes from global traffic, it needs more header bits. Some header bits indicate the destined cluster and remaining header bits indicate the height within the designated cluster. This architecture needs almost double E/O interfaces, which

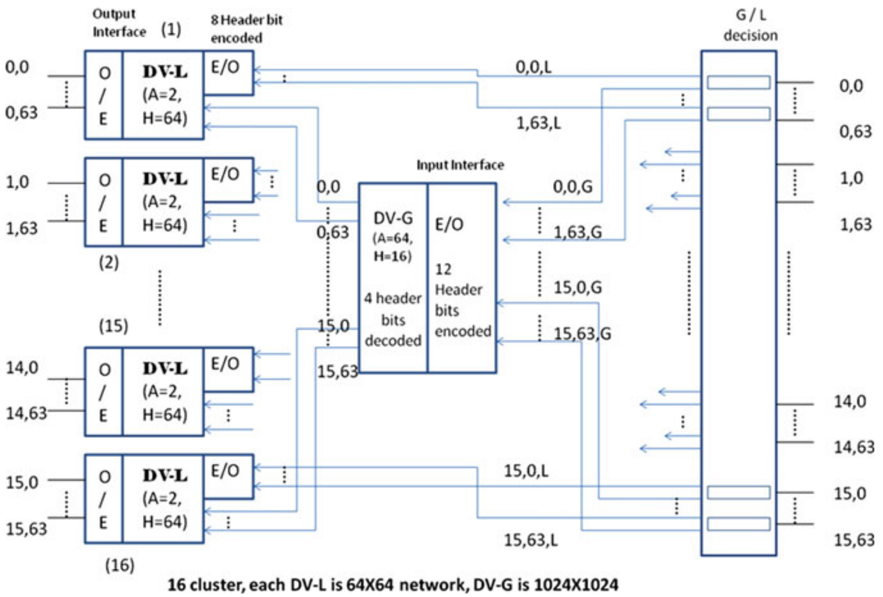


Fig. 1 Cluster architecture of data vortex for large DCN [48]

results in increased cost. Thus the separation of traffic in the optical domain is preferred by using very fast switches which reduce cost of the network.

Here, the routing performance is evaluated by simulation method and it is observed that it has improved throughput and reduced latency with comparable cost. Thus cluster DV architecture is most suitable design for large DCN applications.

3.2 The Data Vortex Network Architecture

This architecture of data vortex supports both the traditional and the irregular emerging workloads. The packets are of fixed lengths, 64 header bits, and 64-bit payload. The new architecture is designed by two blocks: the Data Vortex switch and the Data Vortex interface controller (VIC). The Data Vortex switch controls the switching in the distributed manner with the use of control signals. Hence the complexity of the logic structure has been minimized. When the traffic flows in the network if desired output path is unavailable then the packet is routed through an additional switching node in the same cylinder and creates a deflection design. In this paper, the authors have implemented an electronic version that is based upon the previous optical switch designs. For detailed study of the DV network papers [7, 49] can be referred.

As shown in Fig. 2, the cluster node is interfaced to the DV switch by using VIC, which is a custom Network Interface Controller. Thus, for each cluster node,

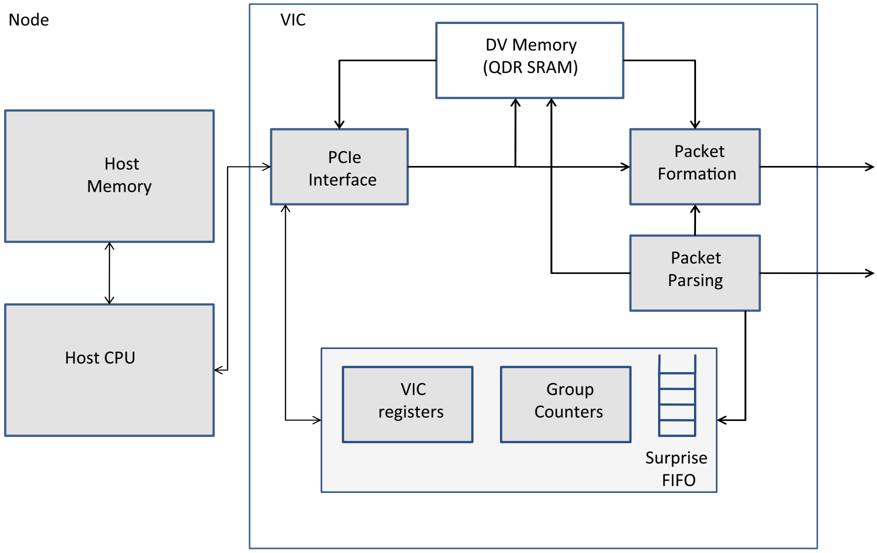


Fig. 2 Block diagram of VIC [7, 49]

it requires at least 1 VIC. The PCI express (3.0 $\times$ 8 boards) integrates a FPGA with 32 MB DV memory. The Quad Data Rate Static Random Access Memory (QDR SRAM) is addressable for reading and writing from the cluster node and from the network. QDR SRAM can also be utilized as a shared memory that is globally addressable.

VIC also facilitates hardware support for group counters, and a network addressable FIFO input queue. The current VIC design can provide up to 64 group counters. The "surprise packet" queue or FIFO queue permits each switching node to receive and buffer 1000 of 8-byte messages within the cluster. To transfer data in between the host memory, VIC memory, and the network, the VIC facilitates two Direct Memory Access (DMA) engines.

The performance evaluation and scalability of 32 VICs per node have been observed and it is found that DV network is more efficient than traditional High-Performance Computing (HPC) networks [49].

4 Conclusions and Future Work

This paper presents a survey on traditional Data Vortex switch architectures and next-generation architectures. It also tries to explore their key role in HPC applications and large Data Center Networks. The advanced cluster architecture of DV and a new Data Vortex network architecture designed for data-driven irregular applications with an overview of the non-blocking congestion-free interconnection network switch and

VIC are also explored in detail. After study, it is found that all-optical packet-switched interconnection network called Data Vortex (DV) can serve as a suitable candidate for interconnection among servers within a large DCN or HPCS applications.

References

1. Nezhadi A, Koohi S (2019) OMUX: optical multicast and unicast capable interconnection network for data centers. J Opt Switch Netw 33:01–12. https://doi.org/10.1016/j.osn.2019.01.002
2. Yang Q et al (2000) Opt Comput 4089:555. SPIE
3. Yang Q, Bergman K, Hughes GD, Johnson FG (2001) WDM packet routing for high-capacity data networks. J Light Technol 19:1420–1426. https://doi.org/10.1109/50.956129
4. Reed C (1999) Multiple level minimum logic network. U.S. Patent 5 996 020, November 30
5. Yang Q, Bergman K (2002) Traffic control and WDM routing in the data vortex packet switch. IEEE Photon Technol Lett 14:236–238
6. Yang Q, Bergman K (2002) Performances of the Data Vortex switch architecture under nonuniform and bursty traffic. J Light Technol 20:1242–1247. https://doi.org/10.1109/JLT.2002.800330
7. Gioiosa R, Warfel T, Yin J, Tumeo A, Haglin D (2016) Exploring data vortex network architectures. In: 2016 IEEE 24th annual symposium on high-performance interconnects
8. Yang Q, Arend MF, Hughes GD, Johnson FG, Bergman K (2000) WDM/TDM optical-packet-switched network for supercomputing. In: Optics in computing, international society for optics and photonics, pp 555–562. https://doi.org/10.1117/12.386875
9. Lu W, Bergman K, Yang Q (2003) WDM routing with low cross-talk in the Data Vortex packet switching fabric. In: Optical fiber communication conference
10. Sharma N, Chadha D, Chandra V (2004) Algorithm for finding all routes in Data Vortex switch. In: Proceedings of international conference, Cochin, India
11. Lu W, Liboiron-Ladouceur O, Small BA, Bergman K (2004) Cascading switching nodes in data vortex optical packet interconnection network. Electron Lett 40:895–897. https://doi.org/10.1049/el:20040586
12. Small BA, Shacham A, Bergman K, Athikulwongse K, Hawkins C, Wills D (2004) Emulation of realistic network traffic patterns on an eight-node data vortex interconnection network subsystem. J Opt Netw 3:802–809. https://doi.org/10.1364/JON.3.000802
13. Small BA, Shacham A, Bergman K (2005) Ultra-low latency optical packet switching node. IEEE Photonics Technol Lett 17:1564–1566. https://doi.org/10.1109/LPT.2005.848276
14. Small BA, Liboiron-Ladouceur O, Shacham A, Mack JP, Bergman K (2005) Demonstration of a complete 12-port Terabit capacity optical packet switching fabric. In: OFC/NFOEC technical digest. Optical fiber communication conference, vol 3, p 3. https://doi.org/10.1109/OFC.2005.192870
15. Shacham A, Small BA, Liboiron-Ladouceur O, Bergman K (2005) A fully implemented 12 × 12 Data Vortex optical packet switching interconnection network. J Light Technol 23:3066
16. Small BA (2005) Dynamic power considerations in a complete 12×12 optical packet switching network. In: 31st European conference on optical communications, ECOC 2005. IEE, Glasgow, UK, pp v4-873–v4-873. https://doi.org/10.1049/cp:20050762
17. Shacham A, Member S, Small BA, Member S, Bergman K A wideband photonic packet injection control module for optical packet switching routers. IEEE Photon Technol Lett 17:2778–2780
18. Small BA, Bergman K (2005) Slot timing considerations in optical packet switching networks. IEEE Photonics Technol Lett 17:2478–2480. https://doi.org/10.1109/LPT.2005.857595

19. Liboiron-Ladouceur O, Small BA, Bergman K (2006) Physical layer scalability of WDM optical packet interconnection networks. J Light Technol 24:262–270. https://doi.org/10.1109/JLT.2005.859852
20. Liboiron-Ladouceur O, Gray C, Keezer DC, Bergman K (2006) Bit-parallel message exchange and data recovery in optical packet switched interconnection networks. IEEE Photonics Technol Lett 18:779–781.https://doi.org/10.1109/LPT.2006.871651
21. Small BA, Lee BG, Bergman K (2006) Flexibility of optical packet format in a complete 12× 12 data vortex network. IEEE Photonics Technol Lett 18:1693–1695. https://doi.org/10.1109/LPT.2006.879934
22. Hawkins C, Wills DS (2006) Impact of number of angles on the performance of the data vortex optical interconnection network. J Light Technol 24:3288–3294. https://doi.org/10.1109/JLT.2006.879508
23. Liboiron-Ladouceur O, Bergman K, Boroditsky M, Brodsky M (2006) Polarization-dependent gain in SOA-based optical multistage interconnection networks. J Light Technol 24:3959–3967. https://doi.org/10.1109/JLT.2006.883122
24. Sharma N, Chadha D, Chandra V (2006) Fault tolerance in multiplexed Data Vortex all optical interconnection network. In: National conference on communications, IIT Delhi
25. Sharma N, Chadha D, Chandra V (2007) The augmented data vortex switch fabric: an all-optical packet switched interconnection network with enhanced fault tolerance. Opt Switch Netw 4:92–105. https://doi.org/10.1016/j.osn.2007.02.001
26. Hawkins C, Small BA, Wills DS, Bergman K (2007) The Data Vortex, an all optical path multicomputer interconnection network. IEEE Trans Parallel Distrib Syst 18:409–420. https://doi.org/10.1109/TPDS.2007.48
27. Small BA, Shacham A, Bergman K (2007) A modular, scalable, extensible, and transparent optical packet buffer. J Light Technol 25:978–985. https://doi.org/10.1109/JLT.2007.891176
28. Shacham A, Bergman K (2007) Optimizing the performance of a data vortex interconnection network. J Opt Netw 6:369–374. https://doi.org/10.1364/JON.6.000369
29. Sharma N, Chadha D, Chandra V (2007) Performance evaluation of the augmented data vortex switch fabric: an all-optical packet switched interconnection network. Opt Switch Netw 4:213–224. https://doi.org/10.1016/j.osn.2007.04.001
30. Shacham A, Bergman K (2007) On contention resolution in the data vortex optical interconnection network. J Opt Netw 6:777. https://doi.org/10.1364/JON.6.000777
31. Iliadis I, Chrysos N, Minkenberg C (2007) Performance evaluation of the Data Vortex photonic switch. IEEE J Sel Areas Commun 25:20–35. https://doi.org/10.1109/JSAC-OCN.2007.021505
32. Hawkins C, Wills DS, Liboiron-Ladouceur O, Bergman K (2007) Hierarchical clustering of the data vortex optical interconnection network. J Opt Netw 6:1179–1190. https://doi.org/10.1364/JON.6.001179
33. Liboiron-Ladouceur O, Shacham A, Small BA, Lee BG, Wang H, Lai CP, Biberman A, Bergman K (2008) The Data Vortex optical packet switched interconnection network. J Light Technol 26
34. Sharma N, Chadha D, Chandra V (2008) Determination of fault tolerance and latency of all optical switch—the augmented data vortex. In: WOC 2008, Qubec, Canada
35. Sangeetha RG, Sharma N, Chadha D 4×4 optical data vortex switch fabric: a fault tolerance study. In: International conference, New Delhi, India
36. Yang Q (2008) Optimum routing and forwarding arrangement in bufferless data vortex switch network. In: Seventh international conference on networks, ICN 2008, pp 633–638. https://doi.org/10.1109/ICN.2008.94
37. Dong L, Dou Q, Feng Q, Dou W (2008) A comparison study of the data vortex topologies with different angle parameter under asymmetric I/O mode. In: 2008 International conference on computer science and information technology, pp 453–457. https://doi.org/10.1109/ICCSIT.2008.139
38. Yang Q Performance evaluation for different arrangements of routing and forwarding paths within bufferless data vortex networks

39. Yang Q (2009) Performance evaluation of k-ary Data Vortex networks with bufferless and buffered routing nodes. In: 2009 Asia communications and photonics conference and exhibition, ACP, pp 1–2. https://doi.org/10.1364/ACP.2009.FL4
40. Sharma N, Chadha D, Chandra V (2010) Terminal reliability evaluation of the augmented data vortex all optical interconnection network. In: 2010 second international conference on communication systems and networks, COMSNETS 2010, pp 1–2. https://doi.org/10.1109/COMSNETS.2010.5431966
41. Sangeetha RG, Chandra V, Chadha D (2010) Optical interconnection reverse Data Vortex network. In: 2010 international conference on signal processing and communications, SPCOM, pp 1–5. https://doi.org/10.1109/SPCOM.2010.5560564
42. Yang Q (2011) High throughput exploration of Data Vortex network. In: 16th opto-electronics and optical communications, pp 535–536
43. Yang Q (2011) Speedup of data vortex network architecture. 5
44. Yang Q (2012) Performance evaluation of a planar layout of data vortex optical interconnection network. In: 2012 IARIA, the international conference on fiber optic and photonics, Chennai, India
45. Burley B, Yang Q (2012) Modular design of Data Vortex switch network. In: 2012 ACMIEEE symposium on architectures for networking and communication systems, ANCS, pp 145–146
46. Sangeetha RG, Chandra V, Chadha D (2013) Optical interconnection bidirectional data vortex network: architecture and performance analysis. J Light Technol 31:1283–1294. https://doi.org/10.1109/JLT.2013.2246542
47. Yang Q (2014) Hybrid switching in modular data vortex optical interconnection network. In: 2014 16th international conference on transparent optical networks, ICTON, pp 1–4. https://doi.org/10.1109/ICTON.2014.6876471
48. Yang Q (2016) Design of optical data vortex cluster network for large data centres. In: 10th international symposium on communication systems, CSNDSP
49. Gioiosa R, Tumeo A, Yin J, Warfel T, Haglin D, Betelu S (2017) Exploring data vortex systems for irregular applications. In: 2017 IEEE international parallel and distributed processing symposium, IPDPS, pp 409–418. https://doi.org/10.1109/IPDPS.2017.121

Technology, ICT, Digital Transformation in Vietnam in the Integration Era—And FDI Effects for Digital Technology and ICT Sector

Thi Hang Nguyen

Abstract The fields of digital technology and ICT sectors in Vietnam require lots of investment, internal and from external sources such as FDI investment. FDI capital from developed countries also helps developing countries take advantage of modern technology, learn advanced management experience to develop economic fields, create motivation to develop industries. Smart production. Developed countries have used technologies such as IoT, machine learning, and artificial intelligence (AI) to optimize supply chains, improve factory efficiency, accelerate product innovation, and enhance value service. In the context of global integration, there are many foreign businesses choosing Vietnam in general and Thai Nguyen in particular as an investment location. This paper uses methods of meta-analysis, comparative statistical methods, survey methods, qualitative and quantitative analysis methods. The results of the article will objectively assess the impact of FDI on the growth of digital tech, ICT sectors and the whole economy. At the same time, propose solutions to increase the efficiency of using capital and technology to develop the digital economy on the basis of taking advantage of digital technologies from countries and large economic groups.

Keywords Digital transformation · ICT sector · Digital economy · Digital education · Industrial revolution 4.0 · FDI · Economic growth · Smart manufacturing

JEL Classification F21 · F23 · P42 · P45

1 Introduction

First, we mention about technology, ICT, digital transformation in Vietnam in the Industrial revolution 4.0: We all know, digital transformation is the integration of technology and digital into production and business activities. Digital transformation

T. H. Nguyen (✉)
Thai Nguyen University, University of Information and Communication Technology, Thai Nguyen, Vietnam
e-mail: nthang@ictu.edu.vn

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_49

in the industrial revolution 4.0 will help countries have more resources, especially technology resources [1–3].

Second, FDI roles for digital technology and ICT sector: According to the Ministry of Information and Communications, Vietnam has now attracted more than 26,000 foreign investment projects (foreign investment) with a total registered capital of nearly 350 billion USD [4–7]. The FDI sector has had a spillover effect on the development of all sectors of the economy, including the important contribution of multinational information technology corporations. However, attracting and transferring technology from the foreign investment sector has not been effective, and foreign investment from multinational corporations in research and development and digital technology is still limited. Meanwhile, the world economy is entering a growth phase mainly based on technology and innovation. Therefore, Vietnam will have a policy to attract investment in the field of information and communication technology (ICT) in the period of digital transformation [8–13]. In the context of digital transformation, Vietnam is increasingly focusing on attracting multinational corporations, especially corporations operating in the technology field, to invest in the production and export of technology products.

International integration in the digital era opens up opportunities in economic development, attracting capital resources as well as technology. Many recent studies show that the value created by digital transformation from four sectors, including automotive, consumer goods, electricity generation, and logistics (logic) amounted to 100 trillion USD by year 2025. In the context that the economy is undergoing great changes, the need to attract investment capital is increasing, meeting the requirements of sustainable development in a developing country like Vietnam. In Vietnam, foreign direct investment (FDI) includes the following basic forms: (A) *Venture business* (This is a form of investment of enterprises established by two or more cooperative enterprises in the host country on the basis of a joint venture contract. New businesses are established in the form of limited liability companies, having legal status under the laws of the host country); (B) *Enterprises with 100% capital of foreign investors—this is* 100% foreign-invested enterprises are owned by foreign investors but are subject to the control of the law of the receiving country. Enterprises with 100% foreign investment capital must be invested, established, and subject to the state management of the host country; (C) *Business Cooperation Contract* (a form of investment signed between investors in order to cooperate in business for profit sharing and product distribution without establishing a legal entity). And when forms of 100% FDI or joint ventures develop, this form will tend to decrease sharply; (D) *Investment contract BOT, BTO, BT* (the form in which one party to sign the contract must be the state. Investment fields are infrastructural works such as roads, bridges, ports, airports, hospitals, manufacturing plants, electricity, water …; forced by the deadline to transfer without compensation to the State); (E) *Investing in buying shares or merging and acquiring businesses.*

The context of international integration in the digital technology era has opened up great opportunities for countries to fulfill their digital transformation goals. At the same time, it also creates opportunities to attract resources for infrastructure development, digital technology, and human resources with digital skills, especially capital and technology resources to improve production capacity. It is also the emergence of the trend of globalization and international economic integration that has increased foreign direct investment capital to developing countries like Vietnam. Nguyen Phu Tu and Huynh Cong Minh (2010) tested the two-way relationship between FDI and economic growth in 64 provinces and cities of Vietnam. The results of the study show that the relationship between FDI and economic growth manifests itself in many aspects, including positive contributions to economic growth, job creation, technology transfer, and negative effects related to social problems. The author Dao Thi Bich Thuy (2012) stated the impact of FDI is also evident in the mobilization and transfer of technology to modernize and improve production capacity and technological level for the investee country itself [14–16].

2 Methodology

To assess the impact of FDI on economic growth in the current digital transformation context, the study used a combination of methods. Including synthetic analysis methods, comparative statistical methods, survey methods, qualitative and quantitative analysis methods. The method of meta-analysis aims to synthesize collected and searched documents to systematize theories related to the impact of capital on economic development.

According to the Cobb–Douglas model:

$$Y(t) = A_f(K, L) \tag{1}$$

FDI has an impact on economic growth and economic restructuring. The regression function showing the correlation between growth and economic restructuring is written in the form:

$$Y(t) = A * K_t\beta_1 + L_t\beta_2 \tag{2}$$

Therefore, taking the logarithm on both sides, we will have the following estimation model:

$$L_nY(t) = L_nA + K_t\beta_1 + L_t\beta_2 + U_t \tag{3}$$

where: A is an unobserved variable in the model or a constant, representing productivity. K is physical capital, L is labor participating in the economy. Capital K includes FDI.

Then, (1) reflects the growth of the economy, as measured by GDP or GDP per capita. Or (2) can be considered as a variable reflecting economic restructuring.

U_t is the error of the model, showing other factors besides capital K and labor L, which affect the dependent variable $Y(t)$ (2). U_t can be the process of international integration, globalization, or the context of digital transformation. β_1, β_2 are the parameters of the regression model; $L_nA = \beta_0$ is the parameter of the regression model (3).

3 Main Findings

3.1 Technology, ICT, Digital Transformation in Vietnam in the Integration Era

Digital economy is a term that has been mentioned for a long time before the concept of the Fourth Industrial Revolution appeared. Along with the Fourth Industrial.

Revolution is the trend of digitization or digital transformation taking place quite synchronously across the entire economy. It can be affirmed that the core element of the Industrial Revolution 4.0 is digital transformation, with the integration of digitization, connectivity/hyperconnectivity, and intelligent data processing as the foundation. People are also gradually getting used to digital technology being applied in all fields and economic sectors. Therefore, the digital economy is gradually becoming a key activity that changes the way of life and operation of the economy-society.

Digital economy is economic activity that use digital information, digital knowledge, digital technology, and digital data as the main factors of production. The digital economy takes the use of the internet, the information technology network as an operating space, and uses digital technology and digital platforms to increase labor productivity. According to Oxford research, the digital economy is an economy that operates mainly on digital technology, especially electronic transactions conducted through the internet. Therefore, the digital economy is also sometimes referred to as the internet economy, the new economy or the web economy.

The context of digital transformation helps to change both production methods (resources, infrastructure, business operations) and (ii) economic structure. Vietnam is ranked 48th out of 60 countries with the fastest digital economy transformation in the world and 22nd in digital development. This is a good sign, showing Vietnam's initial efforts in creating favorable conditions to promote the development of a digital economy based on digital technologies. Digital technology also helps to create new business models, challenge new or replace old business models. Therefore, the digitization of the economy is a technological revolution tied to specific goals. To integrate into the common trend of mankind, we must accept new business models, accept new technologies that fundamentally change industries, collectively known as X-tech. Fintech, EdTech, AgriTech all show creativity that destroys the old, using the new instead. If Vietnam accepts the new, the world's new technology will approach Vietnam, along with a new industry.

To implement digital transformation, Vietnam can choose many paths. But basically, it is necessary to follow the steps. Initially, Vietnam needs to accelerate the digitization of fields and industries, accelerate the digital transformation of the

Government, the digital transformation of enterprises, and the digital transformation of society. All of these are aimed at increasing operational efficiency, increasing labor productivity to create a remarkable development of the economy. The next step, Vietnam needs to use digitalization as a competitive advantage to create breakthrough development. Next, Vietnam needs to form a comprehensive digital economy, all fields are digitized, new generation digital industries and new industries are formed. This is the growth engine for the economy in the current context.

3.2 FDI Investment in ICT, Digital Technology Sectors

Currently, Vietnam is considered an attractive destination for foreign investors. Many global clothing manufacturers such as Nike and Adidas moved their production bases to Vietnam in 2009 and 2012. Vietnam's National Digital Transformation Strategy focuses on smart manufacturing, respectively. Internet of Things (IoT), Artificial Intelligence, Big Data, Blockchain Technology, Online Data, Cybersecurity, Financial Technology, 3D Printing Technology, 5G Ecosystem, Smart Cities. That has opened the way for other foreign businesses to choose Vietnam as a production center in the Asian market with big names like Intel, Samsung, Panasonic, Nokia, Microsoft, and LG. The revolutionary change of the digital economy compared to the traditional economy along with the great impacts of the digital economy has brought many favorable opportunities for domestic enterprises. In that context, information and data have become important and valuable resources in the overall economy. Along with traditional resources, the main development resources are increasingly associated with high technology, information, and human intelligence from developed countries. It is the development of digital technology that allows people to collect data from many different sources, form big data and analyze big data to create high value for the economy. Along with the emergence of emerging technologies such as Artificial Intelligence, Internet of Things, Robotic, Machine Learning, Blockchain Technology has been creating great potential for the economy. Therefore, developing a digital platform for economic development is the current trend.

Vietnam's national digital transformation strategy has also clearly outlined the goal that by 2025, the digital economy will account for 20% of GDP and about 30% of GDP by 2030. The strategy clearly analyzes how Vietnam is focusing on developing Digital Government, Digital Economy, and Digital Society. On the other hand, Vietnam has also established Digital Enterprises. To achieve this goal, Vietnam has taken measures to attract FDI in the development of high-tech parks, concentrated information technology parks, develop 5G mobile network infrastructure, and promote online public services. Through domestic platforms, develop cloud computing technology, ensure network safety and security, improve capacity in artificial intelligence (AI), Internet of things (IoT), data science, Big Data, Cloud Computing, Blockchain.

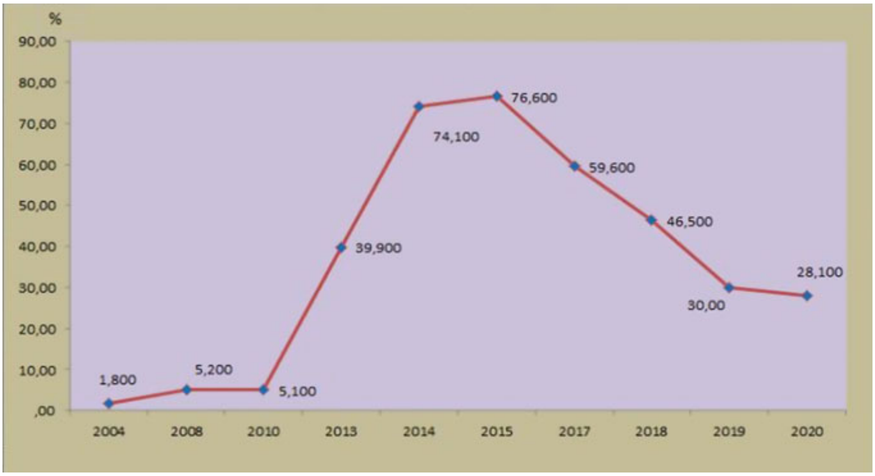


Fig. 1 FDI increases the average annual economic growth rate. *Source* Author's summary from Thai Nguyen Provincial Statistics Office

4 Discussion on Roles of FDI for Economic Growth

4.1 *FDI Increases the Average Annual Economic Growth Rate*

FDI inflows also promote economic restructuring, contributing to an increase in the average annual economic growth rate.

The average growth rate of Thai Nguyen province is quite high (Fig. 1). The period 2004–2020 will reach 24%/year; the period 2004–2005 reached 25%; period 2006–2010 reached 31%; period 2011–2015 reached 23%; in the 2016–2020 period, reaching 17%. Compared with the average rate of Vietnam, the average annual growth rate of budget revenue in Thai Nguyen province in the period 2011–2020 increased by 19.1%/year, nearly two times higher than the growth rate of 9.9%/year of country. Compared with other provinces in the Northern Midlands and Mountains region of Vietnam, the province's budget revenue has reached the highest and has always ranked 1st from 2010 to present.

4.2 *Industry Growth Rate*

Industrial production value in 2004 (at current prices) reached 7.7 trillion VND, by 2020 it will reach 879 trillion VND, 114 times higher than the industrial scale in 2004. In which, the processing and manufacturing industry always accounts for a high proportion and is the driving force for growth; in 2004 accounted for 88.9%

of the total value of industrial production, by 2020, 99.1%; in which the industrial production value of the FDI sector accounts for a large proportion, accounting for only 8% in 2004 and 91.5% of the province's total industry by 2020. The proportion of industry in general in the province's GRDP increased rapidly, in 2004 the proportion of industry accounted for 28%, by 2020 it will account for 51.6% of the total GRDP.

Vietnam has actively participated in integration into the global and regional economy, signed many bilateral and multilateral agreements, and cooperated with partners around the world to take advantage of international opportunities. Trade and investment. Multilateral and bilateral trade agreements have been and will have a positive impact on FDI attraction and activities of this sector in Vietnam.

The ADF unit root test results show that the first difference of the time series in this study is stationary series. This is because the series Loga FDI, Loga industrial GDP, Loga economic restructuring, Loga GDP per capita are all stationary series of 1st order (Table 1).

The AIC information indexes all stop at difference, so the data continues to be fed and tested for cointegration.

The results of the cointegration test in the variables measuring the impact of FDI compared to economic development indicators are shown quite clearly.

Loga FDI and Loga Industrial GDP achieved statistical results of 17,991 and are co-linked.

Loga FDI and Loga GDP per capita achieved statistical results of 17.93 and are co-linked.

Loga of FDI and Loga of economic restructuring achieved statistical results of 22.91 and co-linked. This shows that there is a linear relationship between FDI capital and economic development indicators.

Vietnam is in the process of accelerating industrialization, modernization, and international economic integration, so the capital demand of the national economy is very large. FDI has significantly supplemented and compensated for the need and shortage of that capital. FDI capital plays an important role, reflected in the

Table 1 ADF unit root test results between FDI and economic variables

Unit root test ADF			General conclusions
Criteria	Value level achieved	First-order difference	
Loga FDI	−1.589	−5.392	Stop at rank 1
Loga GDP of industry	−3.197	−3.789	Stop at rank 1
Loga of economic restructuring	−1.979	−4.683	Stop at rank 1
Loga of GDP per capita	−2.803	−3.592	Stop at rank 1

Source Authors calculation

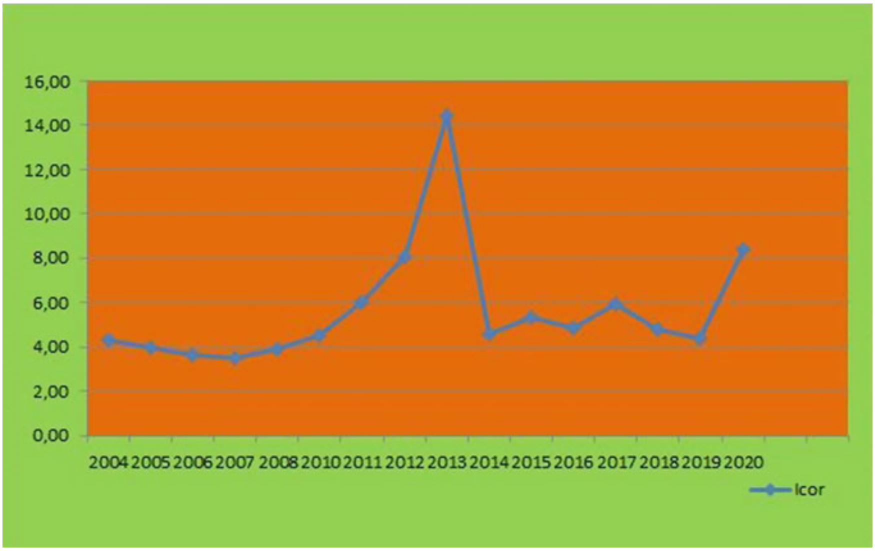


Fig. 2 ICOR evaluate the effectiveness of Using FDI. *Source* Authors calculation

increasing total development investment capital of this sector in the period 2011–2015, the FDI sector contributed over 22% of the total investment capital of the whole society. However, the efficiency of capital use of Vietnam in general and Thai Nguyen province, in particular, is still very low.

Although Vietnam and Thai Nguyen both have some strengths in attracting FDI, which foreign investors are very interested in. But in general, the efficiency of using FDI capital of Thai Nguyen province is very low (Fig. 2). If compared with other countries such as Korea, which are investing a huge amount of FDI in Vietnam and Thai Nguyen today, there is a clear difference. At the same time, Korea has the same level of economic development as Vietnam today, i.e., the period 1970–1990, the efficiency of using Korean investment capital was very high. Specifically: Korea's ICOR in the period 1970–1974 reached 1.97; in the period 1985–1989 reached 2.79.

5 Conclusion

The 4.0 technology revolution will create many challenges for the Vietnamese economy as well as attract FDI into Vietnam. Because the current structural status of the economy is still inadequate, the transition is slow, and the growth is not sustainable. The efficiency of investment capital (ICOR coefficient) of state-owned enterprises is still far behind that of the private sector. The imbalance and lack of mechanism in policy have led to a large difference in the ICOR coefficient of the FDI sector compared to the non-state sector.

The 4.0 technology revolution and the digital transformation context create many advantages for attracting FDI into Vietnam. The international context has many advantages for Vietnam to participate more and more deeply in international playgrounds. The 4.0 technology revolution will have the strongest impact on the manufacturing industry, which is one of the industries attracting the most FDI in Vietnam in general and Thai Nguyen province in particular. The influence of technological change in the global economy, the mechanism of technology diffusion through import and export activities has changed the flow of international trade.

Next, Attention should be paid to FDI projects of high quality, with high added value, using modern and environmentally friendly technologies, especially in the field of information technology and friendly technology. It is necessary to develop infrastructure, train high-quality human resources, research and development, and modern services to help domestic enterprises participate in the global value chain of transnational corporations. In addition, there should be solutions to attract and encourage industrial projects to gradually shift from outsourcing to production in order to select large and reputable investors to invest in market development. With the efforts of the government and ministries, Vietnam has achieved success and made a breakthrough from being an importer of technological products to an exporting country ranked in the top group in the world. Not only that, but Vietnam has also started piloting a project in conjunction with five Australian universities to provide foreign online courses to domestic students. This has helped Vietnam rise to the top position in the online education system in Asia, toward developing digital human resources to participate in the digital economy. Vietnam has committed to improving the production value chain through FDI in parallel with efforts to improve the level of human resources. This is a direction that is appropriate to the context and represents a smart vision, creating a breakthrough for a country that has gone from agriculture to become a technology exporter, creating a competitive advantage.

References

1. Anwar S, Phi Lan N (2010) Foreign direct investment and economic growth in Vietnam. Asia Pacific Bus Rev 16(1–2)
2. Huy DTN, Linh TTN, Dung NT, Thuy PT, Van Thanh T, Hoang NT (2021) Investment attraction for digital economy, digital technology sector in digital transformation era from ODA investment-and comparison to FDI investment in Vietnam. Laplage em revista, 7(3A):427–439
3. Huy DTN et al (2021) Impacts of internal and external macroeconomic factors on firm stock price in an expansion econometric model—a case in Vietnam real estate industry. In: Data science for financial econometrics. Springer, pp 189–205
4. Hang NT, Huan NV (2017) Evaluation of the electronics industry in VietNam in Thai Nguyen by the analysis of financial situation. Int J Rec Sci Res 8(4)
5. Ghirmay T, Grabowski R, Sharma S (2001) Exports, investment, efficiency and economic growth in LDCs: an empirical investigation. Appl Econ 33(6):689–700
6. Marc A (2011) Is foreign direct investment a cure for economic growth in developing countries? Structural model estimation applied to the case of the south shore Mediterranean countries. J Int Bus Econ 898

7. Hang NT (2021) Research on green economic development toward a sustainable economic growth strategy: case study in Thai Nguyen province. Dev Econ 286
8. Hang NT (2018) International integration and development of industrial clusters in Thai Nguyen province. Int J Res Bus Econ Manag 2(2)
9. Hang NT Research on the influence of FDI on economic development, application in Thai Nguyen province. In: Proceedings of the National Conference. Social Publishing House, Labor
10. Hang NT (2017) Evaluation of the growth quality of Thai Nguyen industry in the period of global integration. Ind Trade J 2:82–89
11. Hang NT (2017) Testing the influence of factors on the industrial development of Thai Nguyen province. J Ind Trade 7:82–89
12. Hang NT (2016) Impact of foreign direct investment on the development of industry in Thai Nguyen province. J Econ Manag-Ministry Plann Invest 7+8(77):70–80
13. Nguyen TH (2021) Developing a marketing strategy to attract FDI in industrial development in Thai Nguyen province. J Econ Forecast-Ministry Plann Invest 10(26)
14. https://baotintuc.vn/khoa-hoc-cong-nghe/. Access date 11/10/2021
15. Tu NP, Minh HC (2010) Econ Dev Rev 28(11):54–72
16. Thuy DTB (2012) The impact of foreign direct investment on economic growth in the developing economy model. VNU J Sci Econ Bus 28(1):193–199

FinTech: Application of Artificial Intelligence in Indian Banking

Khusboo Srivastava and Somesh Dhamija

Abstract Indian Banking system have a lot on their plate including the dynamism of emerging competitors, shifting demographics, augmenting expectations of customers, changing regulations and its compliance. Technology offers solutions in the form of FinTech applications. Using data from Report of working group on fintech and digital banking (2017) by Reserve bank of India, this study captures the role of financial technologies or fintech in the financial industry, banking sector in particular. The paper is believed to add value by extensively highlighting the recent trends and cutting edge innovations in fintech for Indian banking. Fintech is now an integral part of Indian banking system, it has transformed challenge into an open door for more adaptability, better serviceability in some areas of the bank. This study suggests a pathway for the evolution of a fintech in changing the business and customers. It will probe deeper into the structured evolution of fintech-based innovations and also relationship between fintech and financial inclusion. The analysis recommends that there can be frontier in creation of value in further exploration of the new insights on upcoming financial technology and its consequences on banking industry.

Keywords Financial inclusion · P2P · Crowd funding · Blockchain · Big data

1 Introduction

Fintech has been a great disruptor and has digital power to change the entire industry. It is an omnibus term now which uses technology to deliver all kinds of financial services. Fintech embrace traditional financial institutions, start-ups, venture capitalists and regulator [1, 2]. FinTech stands for financial technologies, and broadly

K. Srivastava (✉) · S. Dhamija
GLA University, Mathura, UP, India
e-mail: khusboo.srivastava@gla.ac.in

S. Dhamija
e-mail: somesh.dhamija@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_50

Table 1 Domains of financial market activities

Payments clearing and settlement	Deposits lending and capital raising	Market provisioning	Investment management	Data analytics and risk management
Mobile and web based payments digital currencies distributed ledger	Crowd funding peer-to-peer lending digital currencies distributed ledger	Smart contracts cloud computing e-aggregators	Robo advice smart contracts e-Trading	Big data artificial intelligence and robotics

Source WEF 2017

used and have applications in the financial services sector, this technology are principally utilized by financial institutions to support the backbone of their businesses [3]. As per Financial Stability Board (FSB), it is as "FinTech is technologically enabled financial innovation that could result in new business models, applications, processes, or products with an associated material effect on financial markets and institutions and the provision of financial services". In simple words, various industry involving insurance sector, asset management company and payment company and many others employ fintech to offer financial services. The usage of smart phones for advanced access to banking services, investing administrations and digital currency are instance of development intending to increase financial services open to the overall population. Companies based on both start-ups and established financial technology is integral part of fintech which are attempting to supplant or upgrade use of financial services provided by existing financial organization. The methodical study on the definition of fintech affirms "fintech is a new financial industry that applies technology to improve financial activities" [4]. Products and services under fintech currently used in practice comprises of Peer to Peer (P2P) lending platforms, crowd funding, blockchain technology, distributed ledgers technology, Big Data, smart contracts, Robo advisors, E-aggregators, etc. [5, 6]. These products are having their existence in global finance, which with or without intervention of a nodal intermediation agency, create association of the lenders and borrowers, share information between seekers and providers. It has been classified into various categories [7]. According to the domains of financial market activities, Table 1 shows recent classification and it is as follows.

2 Literature Review

FinTech is now the most buzz word with its latest applications, processes, products, or innovative business models in the financial circles, collection of innovative products and financial services by providing solutions via the Internet. Fintech is classified into various phases based on period of years [8]. The first phase, FinTech 1.0 (1866–1967) was the beginning of technology linked financial services but was very costly while FinTech 2.0 (1967–2008) is the phase where financial institutions accepted the presence of technology and its importance in improving the quality of financial services and experience of customers. As per Consumers International, 2017, we are currently in third phase Fintech 3.0 (2008 to present). Phase has embarked upon the wave of start-ups and technology firms. Now the preference of customers has also shifted from banking to other financial companies. The term “fintech” is a summation of two words “finance” and “technology” [8]. Traditional banking and financial institutions are now challenged by fintech which embrace series of innovative services which includes crowd funding, payment solutions based on mobile technology which has grown the mobile payment market to more than $ 1 trillion globally, online portfolio management and international funds transmission [9, 10].

2.1 Objective and Research Gap

The objective of this research study is to provide a thematic review of emerging fintech ecosystem India and to bring its significance to fintech users, entrepreneurs, investors, policy makers and regulators.

As best of my knowledge, various aspects to unfold fintech in Indian banking system has been attempted for the very first time in this research paper. It concisely analyze the various reports on fintech in India and also present a review of literature on financial technology by stating its dynamism in Indian financial system. It helps in the identification of significant inroad in financial inclusion so that unbanked /underserved customers may also access the better options for financial services.

3 Research Methodology

Research is based on secondary data. This paper has been prepared by reviewing various literatures available along with various other reports on fintech in India. Report of working group on fintech and digital banking (2017) by Reserve bank of India is primarily used. Based on literature review of the selected papers and exploratory nature of this study, the paper encompasses the different aspects of fintech in association with Indian banking which include how fintech is transforming banking, enriching the customer’s experience, enhancing the quality of financial services, impacting the business models and empowering the financial inclusion in India. This study helps in extending the knowledge by highlighting the new dimensions of fintech innovations along with their leading practices in India.

3.1 Fintech is Transforming Banking and Bringing It Firmly into the Digital Arena

The last decade has been incredible for fintech and its advancement in various fields. In 2019, about 64% of worldwide customers on boarded one or more fintech platform. It stated that fintech or digital innovations have played pivotal role in transforming financial markets [11]. Fintech is quickly evolving developing fragment of the financial services sector where technology enabled business model start-ups and other new market entrants are challenging operations of the existing financial ecosystem [12, 13]. The unprecedented advancement in India's fintech sector is supported is by the presence of untapped vast market base, focused innovative start-ups and benevolent government strategies and guidelines. Fintech adoptions are taken up by all traditional banking institutions and non-banking financial companies (NBFCs) to speed up along with various dynamic start-ups. This new wave of revolution is rapidly flourishing in financial services sector and for banking it is like paradigm shift. Fintech in India can possibly give serviceable way out for the issues like low market penetration, unavailable history of credit and cash based transactions in economy faced by the traditional financial institutions. Indian banking and financial services sector can be substantially transformed only when there will be participation of regulators, market players and investors. Fintech players are redesigning the pathways of business and customers by driving down the cost of daily transactions. Now, India have progressed significantly as digital economy. India has presently more than 1,500 fintech start-ups, both big and small, out of which half were established in the course of last two years. In India, fintech industry reached 282% between 2013 and 2014, and in 2015, it had touched USD 450 million. Currently in India, more than 400 fintech based organization are functional and their investments are estimated to grow 170% by 2020. According to NASSCOM, the Indian market for fintech software is determined to reach USD 2.4 billion by 2020. The Indian fintech sector's transaction is expected to touch USD 73 billion in 2020 from USD 33 billion in 2016. The various dimensions of fintech in Indian financial markets are given in Table 2.

Fintech is new driver of all business models. One is crowd funding platforms which allow businessman to source funds from the crowd of online investors therefore is source of entrepreneurial finance [14]. Equity crowd funding as "an open

Table 2 Indian fintech industry division basis

Payment	Online financial products
(a) Payment banks (b) Mobile wallets (c) Payment gateway (d) Payment infrastructures i. ATM ii. mPOS	(a) Lending (b) Insurance (c) e-NPS (d) MF/Broking

Source Statista website

marketplace for entrepreneurial finance that takes place on a two-sided online platform and operates within a social media environment" [15]. P2P is a "platforms that facilitate financial services via direct, one-to-one contracts between a single recipient and one or multiple providers" [16]. Peer to Peer (P2P) lending is like borrowing from friends and relatives, allows anyone to raise debt from an online community at a mutually agreed term and conditions of profit sharing mechanism. P2P provide platform similar to micro financing. Digital wallet providers have already enabled people to rural India without bank accounts, to zip money across borders and even overseas using mobile phones. Digital advancement in the form of fintech is redefining the financial services. India has become now more apparent of numerous fintech start-ups, accelerators and incubators. India has everything making it work to build up itself as a worldwide fintech hub. Since India is a huge platform of underserved/undeserved customers, having high mobile penetration, rapidly changing demographics, an innovation driven start-up, consumer behavior and pool of advance technology therefore India has a high chance to establish fintech. It is now coming with new technologies to overcome old financial services, it includes mobile based payments, funds transmissions, fund raising, lending and management of portfolio. The Government of India's push toward a digitized economy with demonetization, the implementation of the Goods and Services Tax (GST), introduction of initiatives such as the Unified Payments Interface (UPI) and the connection of financial services such as bank accounts to Aadhaar have led to a spurt in the growth of fintech. Union Budget 2018 has given further impetus to the fintech community through announcements on the National Programme on Artificial Intelligence, cash flow-based lending TReDs, unique business identifiers and a focus on block chain, etc. While these initiatives will benefit fintech, one of the key requirements for ensuring a sustainable fintech ecosystem is the creation of a friendly, unambiguous and standardized regulatory regime. Artificial intelligence is defined as "AI is best understood as the overarching field that seeks to create complex machines that can exhibit all characteristics of real human intelligence" [17]. It alludes to the substitution of human knowledge by innovation. Blockchain is pointed out as "the fastest-growing area of FinTech innovation" and "held large future potential in financial services" [18]. It is the electronic payment system where system output itself ensures security of the payment. A fact was represented that Robo-advisors can trim down the other extra costs in delivering financial services and also charge lesser fees [19]. Using Robotics and AI in the banks address the issues of risk and cost by its ability of cognition, manipulation and interaction.

3.2 Risks Involved in Fintech

Functioning of fintech attracts lots of risk [15, 20, 21]. Risks are expected to manifest from two key sources as follows:

Table 3 Risks and opportunities emerge from financial technologies and innovation on banking system

Risk	Opportunity
Strategic and profitability risks	Improvement in banking processes
Increased interconnectedness between financial parties	Innovation in data utilization for the purpose of marketing and management of risk
Greater operational risk is systemic	Increased competition has favorable impact on financial stability
Greater operational risk is idiosyncratic	Regtech
High risk due to third-party/vendor management	
Compliance risk which includes consumers and data protection regulation	
Money laundering which can be treated as financing terrorism	
Volatility in funding sources of bank and risk related liquidity	

Source BCBS 2017

1. Micro financial risks—risk related to credit, liquidity, operational specifically cyber and legal risk, leverage and maturity mismatch.
2. Macro financial risks—growth of credit which is not sustainable, expanded interconnectedness or relationship, advantage given for high risk bearing institutions, procyclicality, contagion and systemic significant.

Various risk and opportunities associated with financial technology are discussed in Table 3.

3.3 Fintech Changing the Business Model

The advent of technology in offering financial service has perpetually transformed the working style of the organizations. The business based on traditional model or conventional investors is now not in existence as availability from crowd sourcing to mobile payments has provided many options to entrepreneurs. It has been now convenient to not only start a business, but also has improved its scope of expansion. For instance, in crowd sourcing, entrepreneur may get investors online without meeting them. Therefore long period of investor discussions can be avoided; entrepreneurs can see the funds move in inside only weeks as opposed to months. Digital technology has transformed the conventional banking solution by transferring funds across borders in very less time which has empowered tiny firms and individuals fund transfer far cheaper. A paper revealed that the fintech revolution, as a result, will become the backbone of transactional economics in India [22].

Thus fintech has substantially driven down costs and also fintech adapted firms can acquire immense investments as they are definitely proficient than customary banks, not having similar overheads and responsibility that banks have. Their overall size likewise improves and adapt in a manner greater companies can possibly aspire. Dynamic and innovative companies team up with a shared aim of entering new markets and expanding their customer base. Fintech like crowd funding, mobile based payments and international funds transfer are metamorphosing small businesses startups and go worldwide, and they are making it simpler than at any other time to begin and maintain a business. Crowd financing implies that you can fundraise rapidly and efficiently from individuals everywhere [23]. It has coordinated the way toward discovering start-up capital and reduced the timings of meetings from months to few weeks.

3.4 Fintech Enhancing Customer Experience

FinTech is a disruption which is influencing customer behavior and their expectations incredibly [24–26] by proving the cutting-edge tools to the customers. Innovative technology solutions provide seamless customer experience by prioritizing their needs in designing the tools and efficient service delivery. The mobile wallets have been very popular in last few years. Year 2019 alone accounts for 2.1 billion mobile wallet users. Though the youngest group of shoppers is driving this payment innovation forward, but these new technologies are no longer just for Generation Z. Entire banking platform is available in the mobile itself from payments to online managing portfolio. Specifically millennial prefer to handle financial affairs in simple and convenient way. Blockchain and artificial intelligence integration with the system have made this platform safer for financial transactions than traditions banks. Lower fee, faster turnaround time and accessibility are the reasons behind the inclination of customer. With proliferation of innovative tools and services, it is high time for businesses to embrace the latest technologies as now economy is connected economy where customers have 24/7 access to the information and those who refuse will not succeed and flourish in this market. Even bank should strategize for analytics of open data solution to provide all data from all sources to offer customer centric products and services to deliver delighted experience. Nevertheless, fintech is in its nascent phase in India. It is empowering the customers by offering services through unconventional channels like cloud computing and social media which will become primary medium by 2020 for customer interaction and platform for customer to research and compare the bank offerings. For establishing fintech ecosystem, partnership is required between banks and new entrant which includes designing and development of product by start-ups and distribution and infrastructure capabilities by the banks. Thus it is an attempt to diversify the risk and embrace the faster profits. Factors driving this win–win partnerships are scalability, vision, product life cycles, buy vs. build and trust. Fintech firms are now catering to the huge unbanked markets of small business for profits and growth. Even small fintech firms have attracted massive investments globally due to their technology-based business models [27]. Thus disruptive technology will transform the Indian financial landscape in few coming years.

3.5 Digital Banking Acting as Big Innovation in Financial Services

As per report of financial services technology 2020 and beyond: Embracing disruption, one of the biggest products of fintech is "digital payments" which accounts 25% of fintech market. Technological developments and increased interests from new players are leading to a fundamental reimagining of the process and business model of the financial services industry. There are many examples of digitization of behavioral pattern of consumers into financial products on technological platform. "Paytm-ICICI Bank Postpaid" is short-term credit product launched by collaboration of Paytm and ICICI Bank. It proposes instant offers to Paytm customers for movie tickets, payments of bills, booking of flights and goods purchase etc. It is reported that there will be no documentation or physical on boarding involved. The product relies on the credit assessment of customers by bank on real time basis that uses a various data from the financial background and digital behavior of the customer including the frequency and buying behavior and CIBIL score in understanding the creditworthiness of a customer. The bank has harnessed Big Data to design a new algorithm that measures potential of customers and also limiting risk of fraudulent and dubious payments. Analytics of big data leads to growth in the number of electronic records which are used by financial services to store data, to derive business insights thus improving scalability. It aids intelligent combination between bank and customer sharing symbiotic relationship by providing customized banking solutions [28]. It has capability to anticipate the risk and the demand of customers with great precision than ever before.

3.6 Fintech Empowering Financial Inclusion

Fintech 3.0 has picked up its pace and is now focusing on covering the urban including migrant, domestic help, micro establishment and small time entrepreneur and rural segment comprises of agri-laborers, women, MSME and tiny entrepreneur. There is requirement of approach adoption which is responsive and regulatory in nature to boost financial inclusion [29]. Fintech is reflection of financial inclusion now. Financial inclusion is connecting under banked people of India to mainstream of financial system. Financial inclusion has mainly three dimensions which comprises of access, usage of financial services, and the quality of the financial products and the way of its delivery. Financial inclusion is process of ensuring universal access of various financial products and service at reasonable price [30]. Fintech gave a new gateway to it which is very well recognized and well accepted fact. Fintech is merger of financial service with technology to facilitate financial inclusion by democratizing access to financial system to those individuals and small businesses that were not the part of formal financial market [31]. Fintech companies are becoming platform for all financial solution for small business and acting as catalyst for growth of MSMEs. Digital innovations are pervasive in all aspects of the financial services and have profound impact in enriching financial inclusion. It renamed financial inclusion as digitally derived financial inclusion where digital medium is used to avail financial services to unbanked and underserved clients at affordable cost. It comprises of digital device, digital transactional platforms, agents and offer of additional financial products and services. The principles of digital financial inclusion are promoting and expanding the digital financial services, accessing the associated risk by establishing the legal and regulatory framework, focusing on consumer data protection, encouraging financial literacy to spread the awareness and also tracking its progress. Fintech is key to unlock all issues associated with financial inclusion. It is revolution in financial arena accelerating the growth of country by bridging the gap between informal sector and formal sector. In comparison to brick and mortar banks, online association wins with the emergence of mobile technology [32]. Mobile device, smartphones and wide network coverage fuelling the fintech revolution by converting non-consumer to consumer that at ease and affordable cost. Large numbers of small business have made collaboration with fintech players for credit and payment solutions with less costly options like peer to-peer lending, direct finance [33]. The Reserve Bank is encouraging financial inclusion with schemes like Jan Dhan accounts, Aadhaar enrolment and provision of license to payment banks and small finance banks. Evolving landscape of financial inclusion is backed by Fintech companies, start-ups, financial ecosystem and government aids across the nation for getting bigger picture of financial inclusion in the following areas mentioned in Table 4 by leveraging the financial technology expertise.

3.7 Future of Scope of FinTech in India

Fintech has already altered the financial landscape forever with its salient features. Book written in 2016 evaluated that as with other fintech based industry versus traditional banking outlook [34]. At present, 40% of the population are still having no access to the banking facilities and almost 87% are cash payments. Therefore India has potential opportunity for new start-ups driven by financial services technology. It is anticipated that with mobile usage, there can be increase from 64% in 2018, and Internet penetration is contributing for the development of fintech in India. According to RBI's latest data, digital transaction has increased 6.05% about 1.06 billion as per data in 2017. Additionally, by certain evaluations, 90% of small scale businesses have no access to any of the recognized financial services. These lacunas in accessibility

Table 4 Financial inclusion in the following areas

S No.	Application of fintech	Financial inclusion areas	Implications
1	Easy cash transfer app	For social transfer of cash by government	To augment the personal income of the poor people as it may direct to medium-term sustainable inclusion path
2	Modified bank fintech app	For long term regular transfer of cash under welfare programme	To open account for females of lower income population under Sukanya Shiksha Scheme
3	Integrated Aadhaar infrastructure	For linking each individual credit account with Aadhaar	To stabilize the credit system and to improve accessibility
4	Enhancement in mobile banking	For seeking greater support with mobile based technology	To boost last mile service delivery system
5	Land records digitization	For credit eligibility certificates (CEC) with the help of Aadhaar-linked mechanism	To develop all agrarian segments by providing formal credit supply
6	Loan/payment app	For supporting corporate social responsibility (CSR) initiative	To nurture self help groups (SHGs)
7	Direct account transfer app	For replacing agricultural input subsidies by the government on fertilizers, irrigation and power	To introduce decoupled income support scheme
8	Crop insurance app	To introduce universal crop insurance scheme by government	To cover all crops and to strengthen small and marginal farmers by providing impetus to crop sector
9	Multiple guarantee app for agencies	For exploring the possibilities for counter guarantee and reinsurance	To provide collateral free credit to micro and small enterprises
10	UID for MSME app	For introducing UID for all MSME borrowers	To share the information with credit bureaus

Source Developed by author based on Report of the Working Group on FinTech and Digital Banking (2017)

of innovative technology have proved that there is scope of fintech development in India. It is the now very important aspect to adopt the technology to get more intelligent about the customer's needs.

4 Conclusion

Main insight drawn from this research study is a paradigm shift of banking industry to relatively new industry in India. A major contribution of this study has been to increase our understanding of fintech innovations. It has reflected that the ascent of fintech has opened up doors for potential outcomes. By not only ensuring the efficient financial services delivery and to stay competitive at market, new innovative financial technology has also left behind traditional financial methods by improving the financial stability [35]. It shows that now businesses can offer more services in fewer prices. It will help financial institutions and entrepreneurs to make them more aware of the latest opportunities and developments in financial technology as it may improve the business and will provide new vision to survive in competitive market. There are a plethora of fintech start-ups emerging in India, across all segments in financial services. This represents a significant opportunity to all FIs looking to maximize on the potential outcomes. The pace at which fintechs are emerging; there is no denying the fact that our country has enormous entrepreneurial potential. The fintech transformation is as a rule additionally empowered by the activity of the Government and regulatory bodies which are all set the additional mile to empower financial inclusion and advancement in banking sector in India. Study recommends banks and other financial institutions to effectively work with new companies for their shared advantage because now there will be sharing economy in every part of financial ecosystem. Financial institutions thus have a great set of challenges ahead of them, primarily beginning with the overall user experience. As the ecosystem continues to evolve and redefine itself, FIs have to ensure that their core offerings of products and services have to change drastically. Beginning with on boarding the customer, the experience of processing documentation and creating new accounts has to improve in line with customer expectations which are in turn, being redefined by the fintech start-ups. It is the responsibility of Indian banking system and other financial institutions to embrace the disruption to survive in the global competition.

References

1. Lee SH, Lee DW (2015) On Issue and outlook of wearable computer based on technology in convergence. J Korea Convergence Soc 6(3):73–78
2. Chanson M, Risius M, Wortmann F (2018) Initial coin offerings (ICOs): an introduction to the novel funding mechanism based on blockchain technology
3. Jung D, Dorner V, Glaser F, Morana S (2018) Robo-advisory. Bus Inf Syst Eng 60(1):81–86
4. Goldstein I, Jiang W, Karolyi GA (2019) To FinTech and beyond. Rev Fin Stud 32(5):1647–1661
5. Belanche D, Casaló LV, Flavián C. Artificial Intelligence in FinTech: understanding robo-advisors adoption among customers. Ind Manag Data Syst
6. D'Acunto F, Prabhala N, Rossi AG (2019) The promises and pitfalls of robo-advising. Rev Fin Stud 32(5):1983–2020
7. Frame WS, Wall LD, White LJ (2018) Technological change and financial innovation in banking: some implications for fintech

8. Arner DW, Barberis J, Buckley RP (2015) The evolution of fintech: a new post-crisis paradigm. Geo J Int'l L 47:1271
9. Böhme R, Christin N, Edelman B, Moore T (2015) Bitcoin: economics, technology, and governance. J Econ Perspect 29(2):213–238
10. BCBS BC, o. B. S. (2017) Sound practices: implications of fintech developments for banks and bank supervisors. Retrieved from Basel-Switzerland: https://www.bis.org/bcbs/publ/d415.htm
11. Gomber P, Kauffman RJ, Parker C, Weber BW (2018) On the fintech revolution: interpreting the forces of innovation, disruption, and transformation in financial services. J Manag Inf Syst 35(1):220–265
12. Jagtiani J, Lemieux C (2018) Do fintech lenders penetrate areas that are underserved by traditional banks? J Econ Bus 1(100):43–54
13. Jagtiani J, Lemieux C (2019) The roles of alternative data and machine learning in fintech lending: evidence from the lending club consumer platform. Financ Manag 48(4):1009–1029
14. Cumming D, Hornuf L (2018) The economics of crowdfunding. Palgrave Macmillan, London, UK
15. Gefen D, Pavlou PA (2012) The boundaries of trust and risk: The quadratic moderating role of institutional structures. Inf Syst Res 3(2):940–959
16. Odinet CK (2017) Consumer bitcredit and fintech lending. Ala L Rev 69:781
17. Philippon T (2017) The fintech opportunity' New York. NY University, Stern School of Business. mimeo
18. Chen MA, Wu Q, Yang B (2019) How valuable is FinTech innovation? Rev Fin Stud 32(5):2062–2106
19. Abraham F, Schmukler SL, Tessada J (2019) Robo-advisors: investing through machines. World Bank Res Policy Briefs. 2019 Feb 26(134881)
20. Gefen D, Pavlou P (2006) The moderating role of perceived regulatory effectiveness of online marketplaces on the role of trust and risk on transaction intentions. In: ICIS 2006 proceedings, vol 81
21. Fang Y, Qureshi I, Sun H, McCole P, Ramsey E, Lim KH (2014) Trust, satisfaction, and online repurchase intention. MIS Q 38(2):407-A9
22. Shim Y, Shin DH (2016) Analyzing China's fintech industry from the perspective of actor–network theory. Telecommun Policy 40(2–3):168–181
23. Ahern DM (2018) Regulatory arbitrage in a FinTech world: devising an optimal EU regulatory response to crowdlending
24. Sironi P (2016) FinTech innovation: from robo-advisors to goal based investing and gamification. John Wiley & Sons
25. Moenninghoff SC, Wieandt A (2013) The future of peer-to-peer finance. Schmalenbachs Zeitschriftfürbetriebswirtschaftliche Forschung. 65(5):466–487
26. Li Z, Li W, Wen Q, Chen J, Yin W, Liang K (2019) An efficient blind filter: location privacy protection and the access control in FinTech. Futur Gener Comput Syst 1(100):797–810
27. He MD, Leckow MR, Haksar MV, Griffoli MT, Jenkinson N, Kashima MM, Khiaonarong T, Rochon MC, Tourpe H (2017) Fintech and financial services: initial considerations. International Monetary Fund
28. Yu S, Johnson S, Lai C, Cricelli A, Fleming L (2017) Crowdfunding and regional entrepreneurial investment: an application of the Crowd Berkeley database. Res Policy 46(10):1723–1737
29. Guild J (2017) Fintech and the future of finance. Asian J Public Aff 18:17–20
30. Schindler JW (2017) FinTech and financial innovation: Drivers and depth
31. Bofondi M, Gobbi G (2017) The big promise of FinTech. Eur Econ 2:107–119
32. Haridasan AC, Fernando AG (2018) Online or in-store: unravelling consumer's channel choice motives. J Res Interact Mark
33. Mackenzie A (2015) The fintech revolution. London Bus School Rev 26(3):50–53

34. Zetsche DA, Buckley RP, Arner DW, Barberis JN (2017) From FinTech to TechFin: the regulatory challenges of data-driven finance. NYUJL Bus 14:393
35. Sarma M, Pais J (2008) Financial inclusion and development: a cross country analysis. In: Annual conference of the human development and capability association. New Delhi, pp 10–13

Role of Artificial Intelligence in Revolutionizing Cancer Detection and Treatment: A Review

Sunanda Kulshrestha and Anjana Goel

Abstract Artificial Intelligence (AI) is embraced as a rapidly growing area in the field of computer science. It is defined as the ability of computer system to get the tasks done that conventionally requires human intelligence. The technology has been exploited in the field of medical sciences too, where the AI helps in the detection and analysis of any disease that helps the patient to get personalized medicine. This has led to the development of precision medicine and digital health as sector in the field of medicine. Unlike other diseases, cancer is the most under-diagnosed disease in the world affecting many people. The major challenging area in the field of cancer is the late detection and failure to predict the progression and severity which has fatal results for the sufferer. The challenge of detection of cancer is been helped through by AI. AI has played a major role in the last few years in field of cancer detection and is found to be a success in terms of exploring the type of cancer and the extent of its severity. AI has been found successful in detection cancers including Breast cancer, Lung cancer and skin cancer. The technology of AI is a significant and growing field in terms of its medical application in clinical oncology and also has many future perspectives.

Keywords Artificial intelligence · Cancer · Detection · Digital health · Precision medicine

1 Introduction

Cancer is a global burden and leading cause of death in the world. In the latest paper by Sung et al. [1], 10 million deaths have been recorded by cancer in world in year 2020 with 19 million new cases and many premature deaths. This has direct socio-economic effect on any country. Cancer is a very unpredictable and complex disease because of its adaptive process of interacting with the microenvironment. Its prediction of progression is still the biggest challenge in the field of oncology due

S. Kulshrestha · A. Goel (✉)
Department of Biotechnology, GLA University, 17km Stone, NH-2, Mathura-Delhi Road, P.O. Chaumuhan, Mathura, U.P 281 406, India
e-mail: anjana.goel@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_51

to it arising new complexities and aggressive behaviour with advancement of stages [2].

Focusing on the concept of AI, a field of research where the computer system mimics, do tasks that require the human intelligence [3]. The system interprets large data sets, images, etc. and produces results on the basis of it. The introduction of NGS (Next generation sequencing) has application in prediction of risk, early detection of cancer, prediction of biomarkers, analysis of high-resolution images etc., that has enhanced the innovation and the use in field of oncological research [4].

The date interpretations from the images, etc. in AI are done by coded algorithms and this constitutes the machine learning [5]. Machine learning is a subset of AI that are used interchangeably although are distinct. Machine learning allows machine to learn from data automatically by recognizing patterns and programming. The Machine learning algorithms such as neural network are developed to facilitate the learning ability of machine that helps in making decision and solving problems, the main function is to mimic human brain and make decision after interpreting the data for example images [6]. Apart from this, Deep learning which is also an inevitable part of AI which has applications in field of robotics, drug discovery, diagnostics etc. The AI is beneficial in accessing big data sets from genetic screenings that is generated to provide personalized medicines to the patient [7]. Use of algorithms to process the big datasets generated by NGS (Next Generation Sequencing) is the future of healthcare system. The data is generated through image processing, past health record of the patient, data from clinical trials etc. are needed to be summarized and processed with help of AI with decreased errors. The sector of cancer studies has been the one which has been utilizing the AI for early detection, targeted therapy, precision of future and other information from genetic studies of the patient [8].

The directions in which AI has been used in field cancer studies are been described in Fig. 1.

2 Next Generation Sequencing (NSG)

NGS is an application of AI and helps in detection of mutations and aberrations in cancer-causing agents and genes. With the use of NSG, genomic analysis has rapidly evolved in clinical practices and shows promising outcomes in field of personalized or precision treatment [4]. Sanger's Sequencing method has been the classical approach, but with the introduction of second-generation sequencing, the large-scale sequencing of genomic materials (DNA and RNA) has been possible with high throughput and cost effectiveness. The samples that could be used include blood samples, tumour samples from biopsy, cell lines, liquid biopsy, and FFPE (Formalin-fixed paraffin embedded blocks). The famous Human Genome Project is the example of practical implication of NSG [9]. The approach is widely used in cancer research and diagnostics to look and identify the alterations in cellular transcriptome and molecular pathways [10]. RNA sequencing has been seen to be used widely in cancer samples in clinics. But as the technology advances, it was observed that library preparation was needed for the second-generation sequencing that required labour and also increases the chances of errors. To mitigate this, third generation was introduced which is more efficient than last one. DNA and RNA in portable handheld devices could be sequenced the example being The Oxford Nanopore Technology. The device could sequence single molecule in less time and low cost with quality in longer reads [11].

3 Cancer Imaging in Clinics

AI is a powerful tool in diagnostics and it helps in monitoring the disease, assist in surgical procedures and hence, helps in management of disease. Improved qualities of CT (Computerized Tomography), MRI (Magnetic Resonance Imaging), PET scans (Positron Emission Tomography) have been achieved with help of AI. Images are being analysed using Deep learning providing accurate results with speed. The accurate analysis is important to detect cancer as early that ensures a better check on prognosis. The AI has facilitated computer aided image interpretation, improved quality of images, and radiomics with improved speed and cost effectiveness [12].

For examples, mammography, screening of breast cancer is crucial to diagnose breast cancer, but with young women having dense breast, ultrasound is the only preferred technique. The medical imaging is crucial in early detection which could help in decreasing the mortality rate. Apart from this, in developing countries where proper trained radiologist and medical imaging is not always available, AI in form of computer aided systems can help in detection of breast cancer.

Studies by Rodríguez-Ruiz et al. 2018 compared the interpretation of mammography with manual and with AI. As expected, AI assistance worked better than manual with quick image analysis and accuracy. The AI systems were also found to work even in dense breasts. In article by McKinney, 2020 published in Natures, the introduction of AI system that has the capacity better than human experts in breast cancer prediction has been stated about [13]. Alanazi et al. 2021 have described the CNN (Convolutional Neural Network) to boost breast cancer detection which achieves 87% accuracy with improved accuracy of 8% [14]. Similar studies have been done by Jutsi et al. 2020 in detection of skin cancer through images. Zhou et al. 2017 have studied the use of markers in detection and imaging of CNS (Central Nervous System) tumours [15, 16].

Patil et al. 2020 highlighted the use of AI in lung cancer detection with help of ML. Various models were being designed to look for initial stage and categorizing lung nodules in meaningful manner in benign or malignant form. Khangar et al. 2021 lately reviewed the implication of AI in oral cancer detection [17–19].

4 AI and Digital Pathology

Detection and imaging are not the only work done by AI; it is also the future of pathology. The laboratories will soon be transformed into digital one decreasing the intense labour, increasing efficiency of work whilst maintaining the quality. The AI in pathology helps the investigator to view the images for accurate analysis in large scale and high definition. In this way, identification of unique biomarkers associated with the cancer responsible for prognosis will help in getting the correct treatment done [20]. At present, the area is under study and gaps need to be filled to procure proper outcome of the inter-disciplinary studies.

5 AI and Cancer Therapy

Cancer is disease that goes multidrug resistant at any time and that plays an important factor in relapse of cancer even after treatment. This plays a major challenge in the field which could be dealt only with identification of genes, epigenetic factors and molecular pathways that induced the resistance. For instance, estrogen receptor positive breast and ovarian cancer have been seen to become drug resistant [21]. For such analysis, RNA sequencing and mutation analysis data is obtained from NGS and other bioinformatics tools that reveal the factors behind.

6 AI and Drug Discovery

AI has led to development of new cost and time effect platforms to deliver high throughput data in field of drug discovery. The examples include, the laser capture micro-dissected RNAseq (LCM-RNAseq) which is a NSG tool and it identifies the differentially expressed genes in histological samples from tumours. A study by Civita et al. 2019 on human glioblastoma concluded that molecular events are region specific. The upregulation of growth factors signaling pathways was studied in tumour cells and served as therapeutic targets for glioblastoma. Results from the largest genomic program, TCGA led to work in Lung MAP. Clinical trials were conducted by National Cancer Institute (NCI)-USA for lung squamous cell carcinoma with personalized treatment [22].

7 Identification of Relevant Signatures in Drug Resistant Cancer

In the paper by Dorman et al. 2015, the use of AI in signatures derivation responsible for paclitaxel and gemcitabine resistance in Breast Cancer has been reported. The mutation and gene expression has also been studied using machine learning. As the effectiveness and advancement of chemotherapy is increasing, multidrug resistance and recurrency of cancer has also been an associated problem. Another study by Hossain et al. 2019 where expression of mRNA leading to initiation of tumour formation and progression has been studied. 41 genes have been identified that could predict the clinical traits, progression, and predictive mechanistic related to ovarian cancer [23, 24].

8 AI in Onco-Surgery

AI-based application and development in field of oncological surgery has been a very fascinating and looked upon area. Machines have been designed using AI to assist in surgery. 30.6% mastectomy (Breast surgery) has been successful in AI assisted program. The image processing helps in guidance of needle biopsies and predict high risk cancer lesion [25].

9 Conclusion and Discussion

AI has a significant impact on the healthcare sector and its branches. It is an exclusively and intensely researched topic at present time and researches are been carried to get the maximum benefit of AI and machine learning for the medical system. AI has been used widely for the detection of disease and proving itself to be a cost and time-effective approach. The system analyses big data in less time and gets the results which in turn helps in the early detection of the disease and timely treatment. At present, clinical oncology is the largest sector taking advantage of AI. Cancer has been a burden on the healthcare system with the highest mortality, less survivals and increasing new cases. The major challenge in the field is the timely detection of cancer and type of cancer. Getting familiar with cancer helps in getting personalized treatment to the patients which is also called as the precision medicine. Apart from clinical, AI is also playing an important role in research with the introduction of NSG. Risk prediction, identification of biomarkers used for detection, analysis of large dataset, analysis of images etc. in all terms, the AI showcases itself as a promising technology for oncology studies. AI is an Inter-disciplinary area that could be exploited in ways to meet the challenges faced by the medical field in terms of cancer.

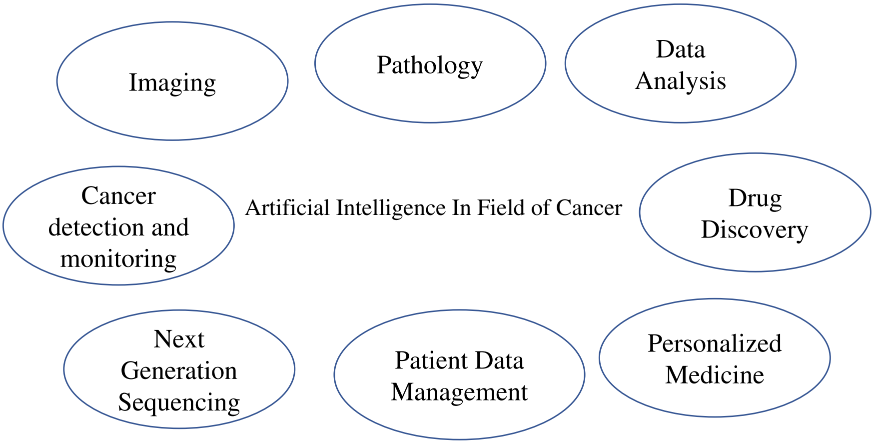


Fig. 1 Use of AI in field of CANCER

References

1. Sung H, Jacques F, Rebecca LS, Mathieu L, Isabelle S, Ahmedin J, Freddie B (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca Cancer J Clin 71:209–249
2. Wenya LB, Ahmed H, Matthew BS, Maryellen LG, Nicolai JB, Alireza M, Tavis A, Omar A, Christopher A, Ian FD, Raymond HM, Rulla MT, Clare MT, Charles S, Udo H (2019) Artificial intelligence in cancer imaging: clinical challenges and applications. Ca Cancer J Clin 69:127–157
3. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. N Engl J Med 380:1347–1358
4. Dlamini Z, Flavia Z, Rodney H, Rahaba M (2020) Review artificial intelligence (AI) and big data in cancer and precision oncology. Comput Struct Biotechnol J 18:2300–2311
5. Joshi AV (2020) Machine learning and artificial intelligence. Springer Nature Switzerland
6. Jiang F et al (2017) Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2(4):230–243
7. Adir O et al (2020) Integrating artificial intelligence and nanotechnology for precision cancer medicine. Adv Mater 32(13):e1901989
8. Davenport T, Kalakota R (2019) The potential for artificial intelligence in healthcare. Future Healthc J 6(2):94–98
9. Kulski JK (2016) Next generation sequencing—advances, applications and challenges (next-generation sequencing—an overview of the history, tools, and "Omic" applications)
10. Wang Y et al (2020) Changing technologies of RNA sequencing and their applications in clinical oncology. Front Oncol 10:447
11. Kchouk M, Gibrat J, Elloumi M (2017) Generations of sequencing technologies: from first to next generation. Bio Med 9(3)
12. Gore JC (2020) Artificial intelligence in medical imaging. Magn Reson Imaging 68:1–4
13. McKinney SM, Marcin S, Shetty S (2020) International evaluation of an AI system for breast cancer screening. Nature 577:89–94
14. Alanazi SA, Kamruzzaman MM, Nazirul IS, Alruwaili M, Alhwaiti Y, Alshammari N, Hameed SM (2021) Boosting breast cancer detection using convolutional neural network. J Healthc Eng 5528622

15. Rodriguez-Ruiz A et al (2019) Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology 290(2):305–314
16. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, Sondermann W, Fröhling S, Hekler A, Schmitt M, Maron RC, Brinker TJ (2020) Artificial intelligence in skin cancer diagnostics: the patients' perspective. Front Med 7:233
17. Patil S, Ibtisam HM, Mona MA, Abrar MA, Mohammed AJ, Lizbeth R, Thirumal AR, Sadiq MS (2020) Reviewing the role of artificial intelligence in cancer. Asian Pac J Cancer Biol 5(4):189–199
18. Zhou H, Vallières M, Bai HX, Su C, Tang H, Oldridge D, Zhang Z, Xiao B, Liao W, Tao Y, Zhou J, Zhang P, Yang L (2017) MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro Oncol 19(6):862–870
19. Khanagar SB, Naik S, Al Kheraif AA, Vishwanathaiah S, Maganur PC, Alhazmi Y, Mushtaq S, Sarode S, Sarode G, Alessio Z, Luca T, Patil S (2021) Application and performance of artificial intelligence technology in oral cancer diagnosis and prediction of prognosis: a systematic review. Diagnostics 11:1004
20. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A (2019) Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 16(11):703–715
21. Nagpal K et al (2019) Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. NPJ Digit Med 2:48
22. Kumar S et al (2018) Tracking plasma DNA mutation dynamics in estrogen receptor positive metastatic breast cancer with dPCR-SEQ. npj Breast Cancer 4:39
23. Hossaina MdA, Islamc SMS, Julian MW, Quinnd FH, Monid MA (2019) Machine learning and bioinformatics models to identify gene expression patterns of ovarian cancer associated with disease progression and mortality. J Biomed Inf 100:103313
24. Dormana SN, Baranovaa K, Knollb JHM, Urquharte BL, Marianif G, Carcangiug ML, Rogana PK (2016) Genomic signatures for paclitaxel and gemcitabine resistance in breast cancer derived by machine learning. Mol Oncol 10:85e
25. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD (2018) High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. Radiology 286(3):810–818

Data Mining and the Associated Social Benefits

Antara Titikhsha Barua, Sharika Tasnim, Zulfikar Alom, Zeyar Aung, and Mohammad Abdul Azim

Abstract Several sectors in our society have employed technical analytics, but now healthcare and the economy are more interested in Artificial Intelligence. A big "running start" for that choice has been data mining. This article focuses on the architecture of data mining in healthcare and the economy. The attention to the study consists of (i) stately changes in the health sector, (ii) the methods and applications of data mining rooted deep in the various industries, and (iii) the approaches and developments of data mining in the aforesaid fields.

Keywords Data mining · Machine learning · Healthcare

1 Introduction

The word "mining" is used for its original meaning of extracting valuables, whereas, in computer science, the valuables are digital information. Until the 1980s, scientists referred to the word "data fishing" to resemble the same ideas of "data mining" for the present day. Regardless of word choice, the meaning encompassed with the words is still similar, extracting, discovering, and learning patterns from an extensive data system. It is surprising to witness the progress of data mining in the last thirty years.

A. T. Barua (✉) · S. Tasnim · Z. Alom · M. A. Azim
Department of Computer Science, Asian University for Women, Chattogram, Bangladesh
e-mail: antara.barua@auw.edu.bd

S. Tasnim
e-mail: sharika.tasnim@auw.edu.bd

Z. Alom
e-mail: zulfikar.alom@auw.edu.bd

M. A. Azim
e-mail: azim@ieee.org

Z. Aung
Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates
e-mail: zeyar.aung@ku.ac.ae

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_52

With the explosion of data in our day-to-day life, we have stepped into a realm of computerization where we want all the data stored in electronic devices safely. The explosion of data in every sector—computer science, medicine, healthcare, education, business, politics, security—has compelled computer scientists, business professionals, medical professionals, and education experts to be concerned about handling data or extracting data from large data pools.

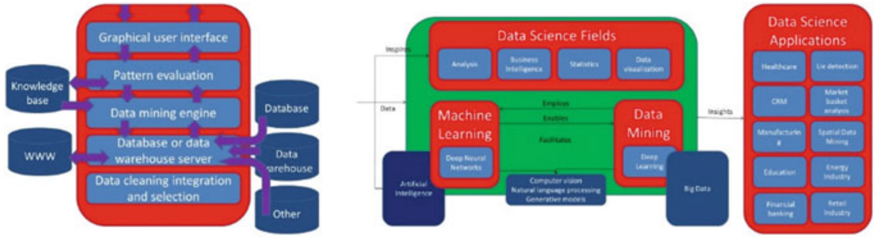
For instance, Wal-Mart runs millions of transactions weekly in its thousands of branches around the world [1]. The healthcare industry in the United States has included Knowledge Discovery, a unique data mining method, in the database to help enhance the skills and knowledge of public health analysts and policymakers [2].

With technology advancements, an urge to handle a tremendous number of data effectively is becoming mandatory. Nowadays, data mining techniques have been applied to control the overwhelming amount of data generated every second. Data mining's popularity has significantly increased after being introduced to the many social sectors, specifically public health and business sectors, which enable the availability of the excessive amount of data and information about the public to the stakeholders to be used in data mining. The critiques will remain the main challenge for the technology. Nevertheless, fast development, challenging opportunities, and collaborative initiatives alongside the social contracts have made data mining inseparable from the various industries.

1.1 Contributions

The main three contributions of the paper are (i) describing the significance of data mining into the public service industries focuses on deriving examples from healthcare decisions and customer care or sales in businesses (ii) narrating the architectures of popular applications in the healthcare and business sector that uses data mining as their primary resource to focus on the deep-rooted utopia of this technology into our society while also devising the different steps of the usage and (iii) addressing the changes made by data mining into these particular sectors to have a comparative view of the benefits that everyone in society is appreciating.

The remainder of this paper is organized as follows. After summarizing data mining and its architecture, in Sect. 2, we review the data mining process regarding a short description of each step. Section 3 concerns data mining applications in mainly healthcare sectors by giving a narrative view on medical decision-making and policy planning. Section 4 takes an approach centered on the economic use of the technology by discussing how data mining works by taking a step-by-step application building process. Section 5 concludes the paper.



(a) Data Mining Architecture (b) Data Mining the big picture.

Fig. 1 Data mining

2 Data Mining

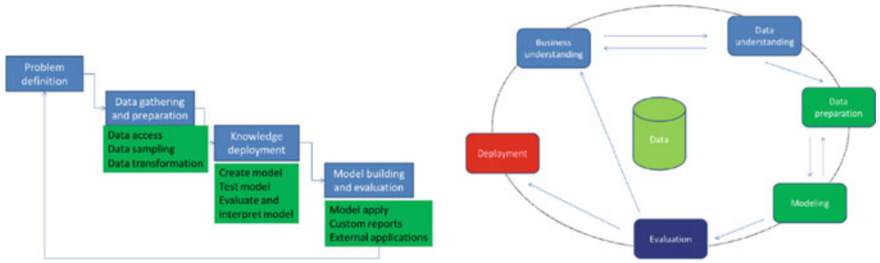
What is Data Mining? Reference [3] defines, "data mining refers to the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis." Data mining results from the evolution of information technology have occurred naturally over the past few years [1]. The main task of data mining is to detect patterns and structures in datasets. Moreover, data mining uses sophisticated mathematical algorithms to segment the data and evaluate future events' probability. Data mining is also known as Knowledge Discovery in Data (KDD) [3].

Data Mining Architecture: The significant components of data mining systems are (i) data source, (ii) data mining engine, (iii) data warehouse server, (iv) pattern evaluation module, (v) graphical user interface, and (vi) knowledge base. Figure 1a presents the components of the data mining architecture and its external interactions with different external servers' databases and repositories. The big picture, where data mining stands, and the interactions of different components and the fields are depicted in Fig. 1b.

3 Data Mining Process

The data mining process starts after the deployment of the solution. Figure 2a and b provides the descriptions for the processes and interaction. The consisting five steps [3] are (a) Data preparation (b) Model building (c) Model testing and computing lift for a model (d) Model scoring and € Model deployment.

Data Preparation: The first stage of the data mining process is data preparation that deals with all the activities that aid in constructing the final sets of data extracted from initial raw data [3]. Data preparation activities comprise the selection of attributes, table and case, data transformation, and cleaning to model the tools [3].



(a) Data Mining Steps. (b) Data Mining Essential Processes.

Fig. 2 Data mining process

Model building: In this phase of the data mining process, various kinds of techniques such as Bayesian classifier, *k*-nearest neighbors, association rule mining, rule-based classifier, artificial neural networks, rough sets, clustering algorithms, and genetic algorithms are applied to build models that solve numerous problems related to large datasets [4]. Same and different data models can be used for the same or different data-related issues. The ultimate goal of model building is to understand the data to provide sophisticated predictions and solutions.

Model testing and computing lift for model: After building a model, the model goes through the *testing phase* so that the quality and the accuracy to provide the outcome can be checked for the target model [3].

Model applying (scoring): As per Ref. [3], "The application of a classification model to new data is called *applying the model*, and the data is called *apply data* or *scoring data*." A classification model plays a vital role in finding patterns in predictor attributes and target attributes. There are a bunch of different classification algorithms that help apply new data. For instance, Oracle Data Mining (ODM) Scoring Engine allows data preparation, building models, importing a model, and applying the model to data [3].

Model deployment: The last stage of the data mining process is model deployment. In this stage, all the data mining models get integrated into a model. Thus, within an environment, it takes the input to provide output. The ODM Scoring Engine reinforces the typical schemes of deployment [3].

4 Applications of Data Mining

Nowadays, data explosion has taken place almost in every sector. Data mining is being applied in business analysis, healthcare systems, retail industry, bioinformatics, education, and politics everywhere in the era of information technology. In this paper, we only focus on how data mining is improving healthcare and business.

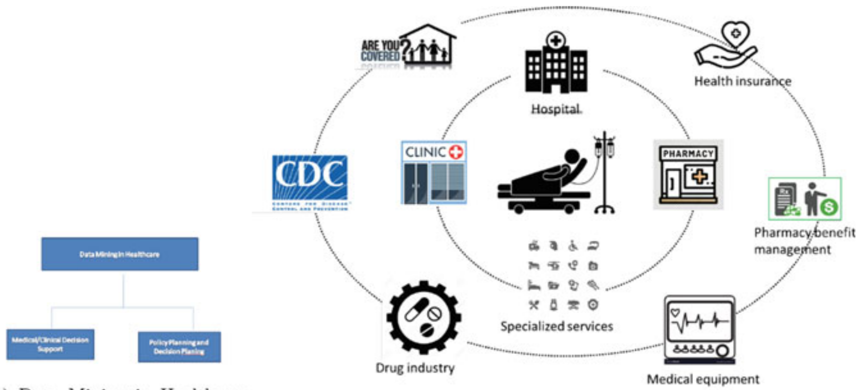
4.1 Data Mining in Healthcare

In the healthcare system, medical data is being expanded every second. For instance, several issues, such as the lack of tools to access healthcare-related data, the necessity of detecting disease, and problems from MRIs, fMRIs, CT scans, PETs, providing treatment at the right time, etc., are aided with data mining. For instance, the patients admitted to hospitals every day produce many data about their gender, blood pressure, symptoms, height, weight, etc., in hospitals. The data is stored in a database of the hospitals.

However, most of the time, it is often observed that a patient whose previous medical history has already been stored in the hospital, asked to submit all the details again. This issue contributes to more problems (i) the same medical history is again asked to store in the database, (ii) the doctor is taking more time to analyze the patient's previous medical history to connect with his/her current health condition.

In this way, much time gets wasted identifying the disease, and it delays having the right number of medical tests. Delay in medical tests leads to the inappropriate analysis of medical reports and medical images as the whole process is time-consuming. Thus, most of the time, doctors struggle to provide the right amount of treatment at the right time, which increases the possibility of a severe medical condition. Hence, doctors, medical, and healthcare professionals are now applying data mining to improve healthcare systems to resolve these issues. That is how the cost of the treatment can be reduced too. A simple demonstration of the data mining in healthcare is presented in Fig. 3.

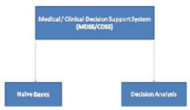
Medical/Clinical Decision Support: Many data mining algorithms in Medical Decision Support System (MDSS)/Clinical Decision Support System (CDSS) are implemented to improve the healthcare system, as shown in Fig. 4 where the taxonomy and the data flow of MDSS and CDSS are shown in the sub-figures. The concept of



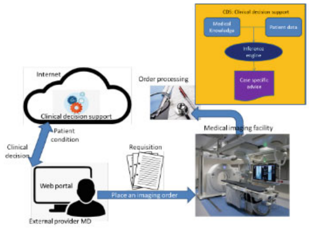
(a) Data Mining in Healthcare Taxonomy.

(b) Healthcare Ecosystem.

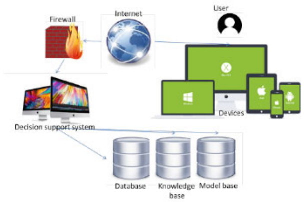
Fig. 3 Data mining in healthcare



(a) Medical/Clinical Decision Support System (MDSS/CDSS) Taxonomy.



(b) CDSS Data-flow.



(c) MDSS Data-flow.

Fig. 4 Medical/clinical decision support system (MDSS/CDSS)

CDSS is not new, as it was first used back in 1980. Since then, it has been undergone perfunctory evolution. CDSS refers to various tools and interventions that may be computerized and non-computerized [5]. Figure 4b and c presents how computational algorithms help the Medical/Clinical Decision Support System (MDSS/CDSS).

Moreover, Ref. [6] argues that Medical Decision Support System (MDSS) is a broad concept. Since MDSS includes patient history, medical test history, pathology of the diseases, family history, etc., diverse data mining approaches have been applied for MDSS. Naive Bayes, rule-based systems, decision analysis, etc., are the most common data mining techniques that have been used in MDSS. Reference [7] mentioned that the Bayes formula could be used to estimate the probability of a diagnosis. The Naive Bayes (NB) algorithm uses the Bayes rule [8]. Reference [9] depicted that "Bayes rule refers to a decomposition of a conditional probability that is frequently used in a family of learning techniques. The learning techniques collectively called Bayesian Learning." Bayes rule can be mathematically defined as:

$$P(z|w) = \frac{P(z)P(w|z)}{P(w)} \tag{1}$$

$$P(\text{pneumonia}) = \frac{P(\text{pneumonia}) \times P(\text{fever}|\text{penumonia})}{P(\text{cough} \cap \text{fever})} \tag{2}$$

where z and w are events; $P(z|w)$ refers to the probability of z given w is true; $P(w|z)$ denotes the probability of w given z is true. And finally, $P(z)$ and $P(w)$ denote the independent probabilities of z and w, respectively.

The Naive Bayes (NB) is widely used to analyze the diagnosis, which helps MDSS as NB is wholly based on assumptions of a particular disease's symptoms. For instance, if a patient has cough and fever symptoms, based on calculation with Naive Bayes, it can be determined whether a person has pneumonia or not.

From the above equation, it can be concluded that there is no dependency between cough and fever. Furthermore, it helps diagnose whether the patient has pneumonia or not. Thus, Naive Bayes applications of data mining are widely helpful for MDSS/CDSS.

Another data mining technique that helps MDSS/CDSS is decision analysis. As per Ref. [10], decision analysis refers to “choice under uncertainty, that is consistent with a set of judgments and preferences for consequences.” Here, the consequences mean that the uncertainty is modeled as probabilities. Reference [11] defines decision analysis as a systematic process. This book also argues decision analysis is prescriptive as it can prescribe what decision should be made upon assigning probabilities of uncertain events. Moreover, the utility model plays a significant role in decision analysis. Since decision analysis helps decide uncertainties, this is widely used in medical/clinical decision support. As per Ref. [12] “In clinical medicine, the analysis consists of determining the best action (diagnostic or therapeutic) for any given patient.”

According to Collins English Dictionary, therapeutics refers to the branch of medicine concerned with treating disease. On the other hand, in Cambridge Dictionary, diagnostic refers to identifying a particular illness using a combination of signs and symptoms. Thus, clinicians can distinguish between diagnostic or therapeutic and reach the best treatment decision with a decision analysis algorithm.

Additionally, this decision analysis algorithm helps physicians decide which decision is best for the patient’s current health condition. For instance, this kind of algorithm helps determine diseases with similar symptoms like Alzheimer’s and dementia. Alzheimer’s and dementia both are neurodegenerative diseases. According to NIH National Institute on Aging (NIA), “Alzheimer’s is a disease which slowly demolishes memory and thinking skills and the capability to perform simple tasks.”

According to World Health Organization (WHO), “Dementia refers to a syndrome which de- grades cognitive functions, i.e., the ability to process thought beyond what might be expected from normal aging.” Just like Alzheimer’s disease, it also causes degradation of memory, incapacitates learning capacity. It is also said that dementia causes Alzheimer’s disease. Since both disorders have similarities and found that dementia leads to Alzheimer’s disease, physicians and doctors get confused to determine whether the patient has Alzheimer’s or dementia (more precisely, moderate or severe stage of dementia).

The gravity of the problem concerns an arduous task to detect the particular disease to provide the proper treatment to the patient. Thus, decision analysis—a data mining technique is a workable tool in decision-making support for this kind of neurodegenerative disease where the differences are very subtle.

Policy Planning and Decision Planning: Policy Planning and decision planning in healthcare systems play a vital role in dealing with healthcare infrastructure and fraud detection issues to understand data. According to the World Health Organization (WHO), “Health policy refers to decisions, plans, and actions undertaken to achieve specific healthcare goals within a society.” Policy planning and decision planning in healthcare play a significant role in maintaining the country’s infrastructure and healthcare systems.

Among the many issues that are urged to be figured out, fraud detection is an emergent healthcare system component to maintain a smooth healthcare infrastructure. Political, administrative, and personal issues together aid fraudulent activities

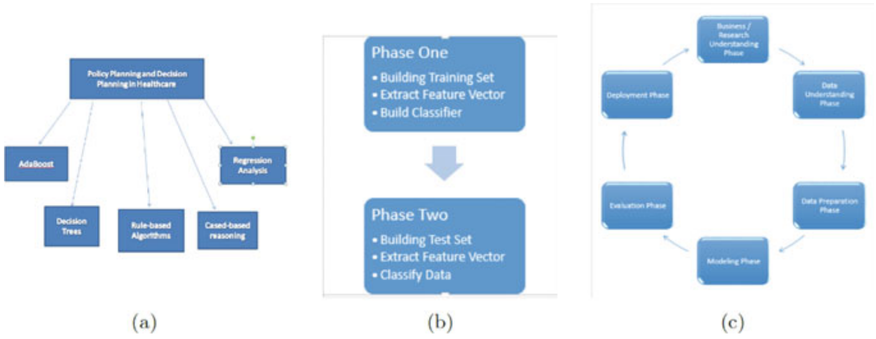


Fig. 5 **a** Policy planning and decision planning in healthcare. **b** Classification algorithms. **c** Regression analysis model

in the healthcare system. Thus, the prevention of fraudulent activities is of great concern. In this case, data mining helps in fraud detection, prevention, and response to fraudulent activities.

A few examples in the healthcare space are misinterpretation in medical data/records, health insurance, etc. These are widespread fraudulent activities in healthcare systems that need to be prevented. Applying machine learning approaches is widely used in this perspective. Figure 5a shows how different machine learning algorithms may be in the design of policy planning and decision planning in healthcare.

Reference [13] applied AdaBoosted Naive Bayes (boosted fully independent Bayesian network) scoring, which is the formulation of the weight of evidence for fraud detection. Moreover, decision trees, rule-based algorithms, case-based reasoning, and regression analysis have extensively been used for fraud detection in the healthcare system. For instance, if a patient tries to steal another patient's prescription and medication details and goes to a local medical store claiming that person initially owns the drug, it is an arduous task for pharmacies to verify the patient's credibility.

In this case, data mining techniques such as AdaBoosted Naive Bayes, decision trees, rule-based algorithms, case-based reasoning, and regression analysis can help detect fraudulent behavior and prevent them. Reference [14] introduced AdaBoost (Adaptive Boosting), a new boosting algorithm that can be applied to reduce the error of a learning algorithm. The specialty of AdaBoost is it can perform a little better than random guessing. AdaBoost reduces learning algorithms' errors from a large pool of data, which eventually sorts out the fraudulent activities, e.g., health insurance claims.

Thus, this algorithm is widely used as a data mining technique in data prediction and classification in health policy planning and decision-making.

Another algorithm that is being used in data mining is a decision tree. Reference [9] depicted decision trees "as a tree-structured classification model, which is easy to understand, even by non-expert users, and can be efficiently induced from data"

(Encyclopedia of Machine Learning). The most popular decision tree algorithms are CHID (Chi-square–Automatic–Interaction–Detection), CART (Classification-regression tree), ID3 (Iterative Dichotomiser 3), and C4.5 which are being used in data mining to plan the policy and decisions in the healthcare space [15].

Additionally, case-based reasoning (CBR) is also applied to healthcare policy planning and decision planning as a data mining technique. According to the reference, [16], case-based reasoning (CBR) refers to reasoning taken by reusing solutions from previous similar problems. That is how CBR diagnosis fraudulent activities to solve the issues in the healthcare space that are similar to previous issues.

5 Data Mining in Business

Every company is using some very familiar words while marketing their product. They seem to use "big data" for their product analysis and "big data" while selling their product. The term is so much used that we seemingly forget to question how companies and businesses manage this large amount of data.

The core of a company's economic analysis is data mining. Reference [17] reported in a survey after reporting on 500 companies in 2002 where each company had at least 35% of their data through data mining. It made their projects' effectiveness as high as $24 million worth of profit each year. After nearly a decade, it is easy to assume that the worth of profit and the use of data mining have increased. Data mining is the art of extracting information from people and using it to generate policies and actions. It is used nearly in every commercial business to generate more profit.

Making a survey based on the entire economy can be challenging, but we may understand data mining reliability by example. Walmart centralizes 20 million transaction-related data each day [18]. The transaction-related data is run and worked on to acquire different patterns and needs of shoppers. Companies are relying on data and data mining to use these to serve promotional purposes.

Data mining is true to its definition while working in the business sector as well. It is agreed that data mining is used to identify relationships among financial benefactors, extract information that can predict future insights, and help generate business plans [17]. Accessing the work process and the impactful results of data mining would be enough for us to understand that the future of business success relies on data mining.

5.1 How It Works in Business Setting

Data mining in business works similar to any setting where a vast amount of data is collected and then processed and analyzed to correct predictions. It works by using the most specific functions of data analysis. The steps are: (i) classification, (ii) clustering, (iii) association rules, (iv) summarization, and (v) regression analysis.

Classification Reference [1] suggested that classification and clustering might seem like the most basic functions of data analysis. However, it works as the core idea in data mining in business. They aim to work their way into predicting the pattern from a cluster of correctly classified data under different categories. Since classification describes classes and data concepts, big industrial data must go through this process first. The algorithms are done in two phases shown in Fig. 5b.

If we take information from customer behavior and do not keep it under organized labels, it might hinge on the prediction outcome. The classification works by naming the data under different categories.

Clustering The clustering process is more complex primarily than the traditional ways of data management in other industries. Since the work is mainly of millions of data information, it goes to a multi-machine clustering technique. The distinct choice is made as the multi-machine clustering technique allows extensive feature extraction in big data [19]. In this particular technique, the data goes through parallel clustering and MapReduce-based clustering. Similar clustering helps the data work in a reasonable time. In the MapReduce-based mechanism, the massive executions are broken into smaller tasks to help with the performance.

Association Rules This methodology is used to discover unknown and hidden relationships between big data. After simply naming and grouping the data, we should look at the association rules, which can be understood by one example. Once customers buy a mobile phone, they tend to look for a phone case and a pair of headphones. The technique helps predict an item's occurrence by understanding the event of another article in the category. These tendencies help the sellers understand the trends and patterns in a customer's need, and data mining helps to get the practices designed. It takes on the implication rule of logic; X Y.

Reference [20] describes the association rule of data mining in the business sector to understand the interestingness and usefulness of a product to the customer through the criteria of support and confidence. Rather than the entire industry, if we consider the economy as a small store, the measures will be easier to explain. The process is done in three simple steps.

- Support: Here, the data accumulates how popularity is based on the transaction amount and measures the proportion.
- Confidence: The data now runs into a probability application that helps with the configuration of how likely it is for the customer to purchase the next product.
- Lift: After assessing the interrelation among products, the result is shown.

Summarization The collected data that ignites customer behaviors' support and confidence is taken from data mining summarization. Reference [21] suggested that the key concept of data mining is summarizing the data as it provides the best way of handling the big ocean-sized data. A summary of a category works for the summary of the individual data themselves. To elaborate on that, we may think of S as a set of data where T transactions are happening. Here, each data is represented by S_i, and the transactions are represented as T_i. In the summary, every transaction. $\mathrm{T_i} \in T$ *is represented by at least one* $S_i \in S$.

Regression Analysis Regression analysis is one of the fundamental analyses that comes into play in working with economic datasets. This technique helps by looking at the matric to identify the variables and their effects. Figure 5c depicts various phases where the regression analysis model may play a role. It can strongly be related to an outlier in the collection. Reference [22] suggested that regression techniques in business analytics work as a powerful prediction function by working as a human brain by looking for relations and giving predictions and interpretation based on anomaly detection. This means that regression analysis goes over the typical concerns, goes over outliers, and includes those variables to understand the significant data change.

For example, in a cosmetics company, if there were more male customers than female in March, the regression analysis counts those outliers on the yearly data prediction. It comes up with a study so that making policies for the company would be easier.

In this way, the essential functions of data mining work to find patterns and gain knowledge over the existing data for business intelligence. Even though no one technique is used in data mining for businesses, the Ref. [23] develops a simple model that can be followed for the basics.

This cycle helps to understand how the general flow of data mining takes place for business intelligence. There are different ways that data mining is used all around us while we buy something and we do not release it. The easiest way to explain is how a loyalty card for a bank or a shop works. They take our purchase information and try to make a pattern to organize customized products that will attract more buyers.

Reference [24] explained that companies are more inclined to maximize their profit through individual customer behavior in the competitive market. The use of data does not only help individual businesses. They are helpful for real market growth. If the market information is used correctly, it will help the entire country develop economically to a progressive future. Budget planning in a country takes place after trying to hold on to a large amount of data and finding the correct pattern to seek out policies that may help. Reference [25] explained the success through data availability, saying that the excitement surrounding big data in e-commerce is because business constitutes finding new ways to make the voice of the market heard. The impact of data mining is way more significant than specific business understanding. It assists with a country's overall progress.

6 Before and After Data Mining

In the 1960s, from being called "data fishing," data mining's pragmatic work has gotten very advanced and has given a new vision to every sector possible other than just the two sectors discussed in the paper. The systematic process takes in the whole system and delivers an outstanding result. While dealing with data, it is not surprising that there are substandard risks. The significant risks are putting out tiny sparks of fear in people about their information safe and how the data is being used. Suppose we try to follow the recent news on the Huawei ban in the United Kingdom and the United States. In that case, we shall understand that the initial accusation put on Huawei started from the fear that the company is conveying the vast amount of data collected from data mining to the Communist Party of China [26]. This example can be nullified as it heavily inclines international politics rather than the actual fear of a threat. However, the very core of this problem comes from the discomfort of a country sharing such in-depth information with another sovereign as the idea of such vast data use has not been wholly guaranteed as safe.

In the healthcare space, the data prediction system has given out wrong decisions and statistics. The problem started with the lack of information, to begin with. Reference [27] suggested that the complex nature of the healthcare and privacy concern makes the amount of data doubtful and often wrong. Suppose people are afraid of how the system is going to work with their information. In that case, there is no use in having advanced technology and performing mistakes that can be solved with a sufficient amount of data coming in.

In the business analytics sector, the situation is very similar as reference [23] thinks that the shift toward an increasing problem-solving environment is prompting challenges, and data integrity and high cost are a few of them. The problem is similar here: unlike the health sector, there have been instances where customer information was wrongly used. This puts the indignity of data use into a complex situation. In a survey done on 500 financial companies in the USA, only 18% thought that data mining is cost-effective. In contrast, others thought with a yearly cost of $24 million, and it is less likely to be cost-effective [17]. This situation argues if data mining is successfully being used for the proper purposes at a fair cost.

The ocean's worth of information is now compacted into one infrastructure to help both ends of the business [28]. Data mining allows companies as it also helps a vast majority of small businesses grow as there is an enormous amount of data by governments that are being public. This move helps the integrity of data stay intact. In the healthcare space, the change is unbelievably surprising. Reference [29] states that practical applications like insights on actions and predictive analysis are possible due to the logical foundation to mine data and develop them. In contrast, previously, the data were in the record room in the form of bunches of the paper sheet [29]. Considering all the situations, it can precisely be said that data mining can be helpful for humanity more than it can cause any harm. We need to build that trust between the data-miners companies and the general public. Trust can only be structured if only laws are being preceded and practiced. The first step has already been taken in Europe commission enacted a regulation that seeks to deliver rapid progress in bringing content online through the practical industry-led solution [30]. This act will help the public to have a better understanding of how and where their information is being used.

7 Conclusions

The world has gone through a new discipline due to the scientific nature of data mining. In the larger picture, data mining necessarily increases the availability of knowledge and information. Data mining is nothing but a progressive step in the healthcare system. It makes the process of helping other people an efficient and effective trustworthy component. Data mining is a compelling technology that can help an organization to plan effective direct marketing campaigns. There are opposing sides to this technology. We also have to understand that every discovery has to go through the ethical aspect of society. The failure to gain the trust of users is a substantial offset. This condition does not mean that we should reduce the use of data mining in different scientific services. It should be acknowledged that data mining has allowed us to have a close correlation with the significant target variable and present the most accurate results in the fields mentioned above.

References

1. Jiawei H, Micheline K, Jian P (2011) Data mining: concepts and techniques, 3rd edn
2. Dennison TH, Qazi F (2005) Application of data mining to health care. In: The 2005 international conference on data mining, pp 245–248; Int J Comput Sci Appl
3. Taft M, Krishnan R, Hornick M, Muhkin D, Tang G, Thomas S, Stengard P, Berger C, Campos M, Milenova B, et al (2005) Oracle data mining concepts, 10g release 2 (10.2) b14339–01
4. Gorunescu F (2011) Data mining: concepts, models and techniques, vol 12. Springer Science & Business Media
5. Kubben P, Dumontier M, Dekker A (2019) Fundamentals of clinical data science. Springer Nature
6. Kumar DS, Sathyadevi G, Sivanesh S (2011) Decision support system for medical diagnosis using data mining. Int J Comput Sci Issues (IJCSI) 8(3):147
7. Ledley RS, Lusted LB (1959) Reasoning foundations of medical diagnosis. Science 130(3366):9–21
8. Rish I (2001) An empirical study of the naive bayes classifier. In: IJCAI-01 workshop on empirical methods in AI, international joint conference on artificial intelligence, pp 41–46
9. Sammut C, Webb GI (2011) Encyclopedia of machine learning. Springer Science & Business Media
10. Wagholikar KB, Sundararajan V, Deshpande AW (2012) Modeling paradigms for medical diagnostic decision support: a survey and future directions. J Med Syst 36(5):3029–3049
11. Gelijns AC (1990) Modern methods of clinical investigation: medical innovation at the crossroads
12. Schwartz WB, Gorry GA, Kassirer JP, Essig A (1973) Decision analysis and clinical judgment. Am J Med 55(4):459–472
13. Viaene S, Derrig RA, Dedene G (2004) A case study of applying boosting naive bayes to claim fraud diagnosis. IEEE Trans Knowl Data Eng 16(5):612–620
14. Freund Y, Schapire RE, et al (1996) Experiments with a new boosting algorithm. In: ICML, vol 96. Citeseer, pp 148–156
15. Alsagheer RH, Alharan AF, Al-Haboobi AS (2017) Popular decision tree algorithms of data mining techniques: a review. Int J Comput Sci Mob Comput 6(6):133–142
16. Grievink J, van Schijndel J (2020) European symposium on computer aided process engineering, vol 12. Elsevier

17. Calderon TG, Cheh JJ, Kim IW (2003) How large corporations use data mining to create value. Manag Account Q 4(2):1–1
18. Alexander D (2020) Data Mining, the University of Austin. http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/
19. Zerhari B, LAAMS (2015) Big data clustering: algorithms and challenges, international conference on big data, tetuan, morocco. https://www.researchgate.net/publication/276934256 Big Dat a Clustering Algorithms and Challenge
20. Tan SC, Sim BH (2014) A pragmatic approach to summarize association rules in business analytics projects. In: International conference on technologies and applications of artificial intelligence. Springer, pp 322–333
21. Chandola V, Kumar V (2007) Summarization–compressing data into an informative representation. Knowl Inf Syst 12(3):355–378
22. Mishra BK, Hazra D, Tarannum K, Kumar M (2016) Business intelligence using data mining techniques and business analytics. In: 2016 international conference system modeling & advancement in research trends (SMART). IEEE, pp 84–89
23. Sharma BR, Kaur D, Manju A (2013) Review on data mining: its challenges, issues and applications. Int J Curr Eng Technol 3(2)
24. Wedel M, Kannan P (2016) Marketing analytics for data-rich environments. J Mark 80(6):97–121
25. Chen H, Chiang RH, Storey VC (2012) Business intelligence and analytics: from big data to big impact. MIS Q pp 1165–1188
26. Helm T (2020) Pressure from Trump lead to 5G ban, Britain tells Huawei, the guardian
27. Humayun A, Waqar A (2017) A comparative study on usage of data mining techniques in healthcare sector. Int J Comput Appl 975:8887
28. Cui Z, Yang J (2019) Research on ocean big data service technology in distributed network environment. J Coastal Res 98(SI):141–145
29. Crockett D, JREB (2017) What is data mining in health care?, health catalyst. https://www.healthcatalyst.com/wp-content/uploads/2014/06/What-is-data-mining-in-healthcare.pdf
30. Communication on "Content in the Digital Single Market", European commission (2012). https://ec.europa.eu/licences-for-europe-dialogue/en/content/about-site#main-content

Factorization of AI Application in HRM

Aneesya Panicker, Avnish Sharma, and Utkal Khandelwal

Abstract Last decade has witnessed drastic change in the field of technology, resulting into transformational change in the way an organization operates. This radical technological transformation is referring to as Industry 4.0 further known as intelligent industry. This talks about the inclusion of some disruptive technologies such as Artificial intelligence (AI) for handling and managing the different business operations or managerial functions. AI applications play a very significant role in the area of human resource management. The use of AI technology facilitated the organizations to extract benefit through effective utilization of human resources. Thus it is important to find out the significant factors responsible for adopting AI in Indian context through empirical investigation. For this, 247 professionals working in NCR and UP west were surveyed through online and offline mode. Principle component analysis was used to make factorization of AI indicators in HR. Results provide four major factors responsible for adopting AI in Indian organizations namely-association, robotics, channelization, and position factors. Research findings reflect association factor as most critical factor in adopting AI in Indian organizations.

Keywords Artificial intelligence · Human resource management · Technology · Robotics

1 Introduction

In the last decade the development of technology has made a huge transformation in the way and patterns the industrial organizations operate. There is a radical shift in the way industries operate their business models. This transformation of the

A. Panicker (✉) · A. Sharma · U. Khandelwal
GLA University, Mathura, India
e-mail: aneesya.panicker@gla.ac.in

A. Sharma
e-mail: avnish.sharma@gla.ac.in

U. Khandelwal
e-mail: utkal.khandelwal@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_53

industry gave the birth to new concept called as Industry 4.0 which is known as the intelligent industry. Experts and investigators in this area believe it to be the fourth industrial revolution. Professor Schwab coined the term fourth industrial revolution which means that present day's organizations become an intelligent enough to achieve the most desirable and strategic business results. This talks about the inclusion of some disruptive technologies such as Artificial intelligence (AI) for handling and managing the different business operations or managerial functions including the most vital function, i.e., Human resource management. The adoption of these disruptive technologies transforms the people and organizational lives drastically and workplaces are hugely impacted by the emergence and development of different AI entities. This development of technology has transformed the way HR works today. This means human resource management is now operated and managed by using AI applications for making the smarter strategic and operational decisions of the organizations [9]. This study aims to understand the role and application of artificial intelligence in human resource practices from the employee' s perspective and the way AI has transformed the people related functions and decision in the organizations. This study will highlight the different experts' opinions on the applications of AI at the workplace and to find significant factors responsible for adopting AI by Indian organizations and also empirically defines and analyzes those challenges [10].

Different experts presented their opinions on artificial intelligence and its nature and application. Stanford University states that 'AI is the science of making intelligent machines, such as intelligent computer programs and using computers to recognize human intelligence, but it does not have to confine itself to biologically observable methods'. As humans use their ability to process the available information by using the cognitive processes such as perception, attention, reasoning, problems solving, make judgment, and learning, in a same manner AI uses the computational intelligence or machine intelligence [12].

2 Connecting AI and HRM

AI applications play a very significant role in the area of human resource management. Some major functions which are massively impacted by AI include recruitment and selection, on-boarding activity, performance management, process improvement, and handling administrative tasks and responsibilities. AI supports both the recruiter and the employees during the hiring process. The use of AI technology has simplified application processes by designing of easier to use forms that reduce the number of applications involved in between the process. Use of different Intelligent technologies in form of AI has automated the different processes such as employees' benefits administration, candidates' pre-screening, and scheduling of employees' interview, etc. which help to ensure the smooth HR administration and organization's strategic success, and most importantly reduce the overall burden and time consumption of HR professionals to keenly manage the HR processes and activities more smartly to make a greater organizational impact [3]. Only the major concern while using AI is safety, security, and privacy of HR data and information which create a huge pressure over the HR managers and administrators and prevent them to use AI at the workplace. The use of AI technology supported the employers, to take benefit of human resources at different timings and days via use of chatbots and remote support applications. The integration of AI in personnel management benefited the administrators by reducing their administrative burden and enabled the faster integration of different manpower activities, processes, and decisions [1].

3 Literature Review

AI becomes a very popular and important phenomenon in the present time and most of the organizations are applying it at the workplace to improve the quality, speed, and accuracy of work performed. This study is an endeavor to explore the data related to the role of AI in human resource management.

Researchers in the field of AI stated that it has developed variety of benefits for the future organizations. Research states that in the coming future there will be the huge development of AI applications in form of classy robots and there will be emergence of radical workplace transformations. There is huge probability that the manpower will be replaced with robots and only smart, skilled, trained, and updated people will be in high demands, it will also give birth to the emergence of new businesses and trades. Although experts believe that the job of technology is to make people life easy and convenient but this development of AI technology can be so terrible that people may lost their jobs in many cases [5]. Therefore, the role of AI needs to be handled and taken care by the organizational leaders more responsively [6].

In the last decade the development of AI and other advanced technologies have gained so much of importance for the industrial sector. That's why the number of researches in this field also increased significantly. Experts believe that these digital transformations may have massive impact on organizational processes including the people administration and management that's why it's duty of HR managers to look after this concern on priority and with major ethical responsibility so that it cannot affect the people life and emotions so adversely at the workplace. Many of the researchers in the area believe that AI can never replace human intelligence as human beings have cognitive abilities in form of thinking, reasoning, analyzing, and most importantly inventing and creativity development which could never be replaced by machines [13]. So even there is a continuous development in the AI field but it can never be the best substitute of human intelligence.

With the development and understanding of AI and its related implications, managers are required to initiate and develop the kind of suitable management policies driven by AI which care and benefit the employees and their expectations on priority. Workplace and HR managers are required to evaluate and understand the constraints and opportunities associated with AI at the workplace [2]. The use of AI in human resource management can help to improve the efficiency of recruitment and selection process and can assist the organizations in attracting and recruiting the quality pool. In fact, the use of AI may also diminish the discriminatory practices like nepotism and partiality in the hiring process. The effective use of AI in human resource processes and activities may also have a positive effect on the manpower development, retention, and their productive use for accomplishment of assigned duties and tasks. This way the origin and development of AI applications in HRM is always productive and very useful and can help to improve the quality of organizations' actions and decisions [4].

Human Resource administrators can use different AI technology for performing diverse human HR activities starting from people procurement to people separation. Talent acquisition and management become very easy and standardize with the integration of AI technology. Use of AI technology benefited the organizations in form of cost efficiency, quality, accuracy, speed, data management, and building competitiveness. On the other side, use of AI technology has harmful effects on safety and privacy related issues in the organizations [7].

4 Research Methodology

This research aims to find significant factors responsible for adopting AI by Indian organizations and also empirically defines and analyzes those challenges. To identify the factors responsible for adopting AI by Indian professional, 247 professionals from working in North India were surveyed. For this purpose, a well-structured questionnaire was developed by generating opinions from 12 different experts from Industry as well as academia. The questionnaire consists of 14 questions related to adaptation of AI in HR, which were extracted through telephonic interviews of 25 subject matter experts. Data were collected using a five point Likert scale, wherein 5 represents strongly agreed, 4 agreed, 3 neutral, 2 disagree, and 1 strong disagree. To test the reliability of the scale, Cronbach's α was used. The reliability value comes out to be 0.801. Data collection was done via both online and offline mode from professionals working in National capital region of India, Agra, Mathura, and Aligarh etc. A total of 247 filled questionnaires have been used. Referral and convenience sampling methods have used for selecting the sampling unit.

5 Analysis of Research Findings

As mentioned earlier this survey is regarding identifying the factors affecting the adaptation of AI in HR practices by professional working in Indian organizations, which were measured through 14 statements on five point Likert scale. The high average score indicates that the most significant element is customized personal content (PF2). Table 1 shows the descriptive statistics on number of responses, average attitude score, standard deviation, and the minimum/maximum values.

Table 1 Descriptive statistics

TT	N	Minimum	Maximum	Mean	Std. deviation
AF1	247	1.00	5.00	3.7287	0.83818
AF2	247	1.00	5.00	3.7449	0.75690
AF 3	247	1.00	5.00	3.6235	0.89711
AF 4	247	1.00	5.00	3.6640	0.82420
RF1	247	1.00	5.00	3.6721	0.77154
RF 2	247	1.00	5.00	3.6437	0.80811
RF 3	247	1.00	5.00	3.6518	0.87433
RF 4	247	1.00	5.00	3.8219	0.80689
CF1	247	1.00	5.00	3.7085	0.75172
CF 2	247	1.00	5.00	3.6842	0.86806
CF 3	247	1.00	5.00	3.6478	0.87966
PF1	247	1.00	5.00	3.6640	0.83401
PF 2	247	1.00	5.00	3.5344	0.90495
PF 3	247	1.00	5.00	3.7328	0.83706
Valid N (listwise)	247				

Integration of relevant and important content, automate administrative task, multichannel sourcing, and customized personal content are emerged as the most significant factors in the survey as the statements related to these dimensions have a high mean score. However, to bring more clarity in the results, exploratory factor analysis through varimax rotation was applied. For applying the factor analysis, suitability of data by correlation matrix, anti-image correlations, KMO (Kaiser-Meyer-Oklin) for sampling adequacy, and Bartlett's sphericity test was considered. The correlation matrix was calculated, which showed adequate correlations for the study of the factor. Because the partial correlations were weak, this meant that the data had real elements. The KMO value was obtained as 0.618, which suggested that the size of the sample was adequate for the application of the factor analysis. Bartlett's test is another indication of the strength of the relationship among variables, and it also showed a significant correlation among the variables. These two parameters revealed that data is fit for factor analysis. Factors were extracted through the principal component analysis through varimax rotation, and the number of factors was maintained through the latent root criterion, the variance was clarified, and the Scree Chart was analyzed (Fig. 1). Four factors were derived from the exploratory factor analysis with cumulative 50% of variance is explained. This figure is quite good as the predictability power of this research is 50%. Results were obtained by orthogonal varimax rotational relationship, and all factor loads greater than 0.445 (ignoring the sign) were maintained which is approximately 0.5, as shown in Table 2. The last segment in the table shows the commonalities. The percentage of the total variance is the metric used to evaluate how well the entire variable answer fits for what the

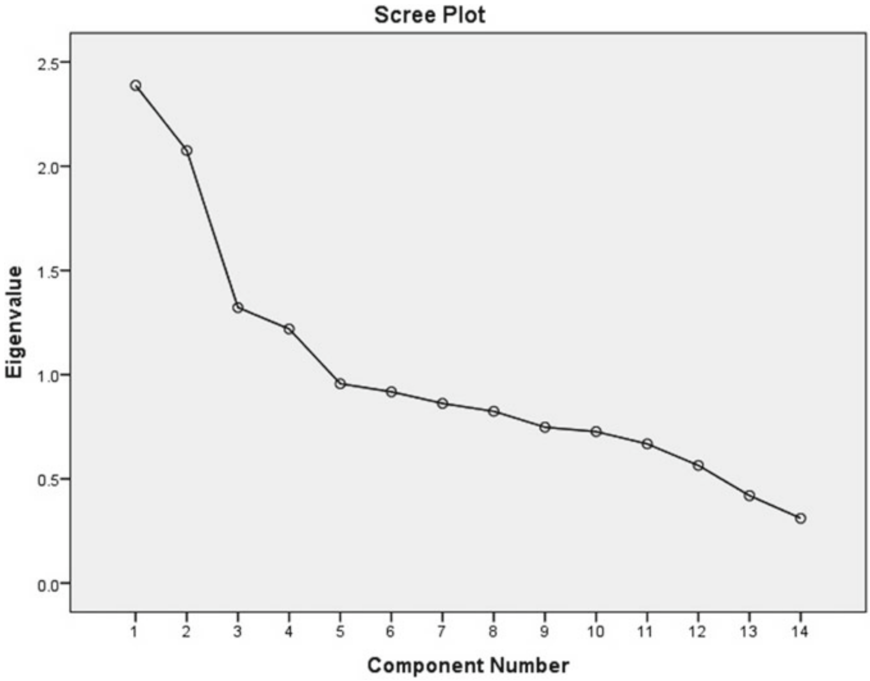


Fig. 1 Scree plot of Eigen values for adopting AI by Indian Professionals

variables collectively reflect. The terminology of the factors was based on the variables applied to each element. Component terms, marks, and their loads are listed in Table 3.

6 Discussion and Future Research Directions

The development of AI has an enormous impact on the functioning of the organizations and has a particular impact on the way different managerial and administrative functions are being performed and executed in the organizations. In the present time it has completely changed the way HR functions and activities are being performed and regulated in the organizations.

This study aims at finding the most significant factors responsible for adopting AI by Indian organizations and empirically defines and identifies the challenges faced by the professionals while adopting AI in HR practices.

The major factors responsible for adoption of AI are classified as association factors, Robotics factors, channelization factors, and position factors as shown in Table 3.

Table 2 Results of principal component analysis: varimax rotation

Label	Component				Communalities
	1	2	3	4	
AF1	0.802	−0.013	0.250	−0.022	0.707
AF2	0.743	−0.096	0.089	0.007	0.569
AF 3	0.660	0.043	−0.055	−0.102	0.451
AF 4	0.643	0.104	−0.159	0.191	0.486
RF1	0.048	0.819	0.119	−0.010	0.687
RF 2	0.019	0.758	0.075	0.035	0.582
RF 3	−0.085	0.648	0.195	−0.026	0.466
RF 4	0.061	0.445	−0.096	0.287	0.293
CF1	0.003	0.138	0.703	0.017	0.514
CF 2	−0.113	0.185	0.654	0.007	0.475
CF 3	0.220	−0.057	0.611	0.158	0.450
PF1	0.018	0.051	0.030	0.733	0.542
PF 2	−0.100	−0.096	0.016	0.667	0.465
PF 3	0.104	0.184	0.137	0.504	0.318
Eigen values	2.388	2.075	1.322	1.219	
% of variance	17.054	14.822	9.442	8.709	
Cumulative %	17.054	31.877	41.319	50.028	

The association factor which is Factor 1 emerged as the most significant when it comes to AI in HRM in Indian organizations. This factor explains 17.054% of the total variance which highlights the importance of this factor. This factor is a combination of different dimensions including application filtering, cognitive supporting decision, employee engagement, and rediscovering candidate which are very essential dimensions of HR in every organization. AI helps in managing and filtering candidates' job applications, helps in deciding the employee engagement strategies virtually, and many other aspects which support to connect and associate the employees with their organizations.

Robotics Factor has appeared as a second vital factor accounts of 14.822% of total variance. It includes different dimensions like career tracking and development, customized personal content, explaining job profile, duties and benefits and automate administrative task. AI role could be better understood by looking at the development of robotics at the workplace. Robots as AI application learn to execute different tasks from humans while using machine learning and considered as a significant element of computer programming and AI. Robotics factors in HR help to automate different administrative tasks and duties and support particularly in employee career tracking and development for successful implementation of organizational activities. This is considered as very important while look at the future of HRM.

Table 3 Factor analysis results: significant factors for adopting AI by Indian Professionals

Factor No.	Factor name	Label	HR dimensions	Factor score
1	Association factor	AF4	Application filtering	0.802
		AF3	Cognitive supporting decision	0.743
		PF1	Employee engagement	0.660
		CF3	Rediscovering candidate	0.643
2	Robotics factor	RF4	Career tracking and development	0.819
		PF2	Customized personal content	0.758
		RF3	Explaining job profile, duties, and benefits	0.648
		AF2	Automate administrative task	0.445
3	Channelization factor	AF1	Integration of relevant and important content	0.703
		CF2	Data driven recruitment	0.654
		PF3	Customized employees data	0.611
4	Position factor	CF1	Multichannel sourcing	0.733
		RF2	Document verification	0.667
		RF1	Device request	0.504

The third factor includes the channelization factor which consists of integration of relevant and important content, data driven recruitment, and customized employee data. This factor accounts for 9.442% of total variance explained. It is very much essential for the organization to keep the manpower data and content for making various strategic decisions particularly related to their recruitment, training, and performance management. This helps the department and organizations to make quick and quality HR decisions on time.

Position factor is taken as a fourth major dimension for describing the application of AI in HRM. This factor explained 8.709% of variance and covers multichannel sourcing, device request, and document verification. It's very simple for the organizations when they source the candidates from one source only but when it comes to multichannel sourcing it is quite challenging for HR managers. Here AI supports the HR administrators a lot in sourcing candidates from various channels like job boards and social networks etc. This way the use of AI makes this job very much easy and comfortable for HR managers at the workplace. It also supports HR staff in documents verification process [11].

This study helps to explore the different factors as mentioned above, which support the successful implementation of AI applications in HRM field and their role to support the various HR activities and functions adequately.

Although adopting AI applications in HRM is not an easy task and may face a lot of barriers. One of the biggest challenges from organizational perspective is the requirement of good investment for AI adoption and talent gap. Organizational and HR leaders are required to plan and make an adequate budget for installing AI through automations and robotics and also it could be difficult, costly, and tough to find well-educated and skilled people for its implementation. Apart from this maintaining the privacy of HR data, its security is another major concern. This also requires the ongoing maintenance, deep learning, and ongoing review and further updates. Therefore, it is essential for organizational leaders particularly HR managers to consider and look after these challenges well in advance [8].

Eventually, this study explored the different set of factors which are required for implementing AI in HRM field by the organizations. The future researchers are advised to make an in-depth analysis of these factors to better understand their implications from HR and organizational perspective. In present time, AI is an important concern for all the growing future organizations, that's why they need to craft and employ a sound strategy for its adoption in benefit of the people and organization. Future researchers are also advised to increase their sample size in future research to better investigate and explore the role of AI in HRM. This study could also be replicate in other functional areas of management. Future researchers are advised to expand their geographical coverage for collection of large pool of data regarding the AI application in HRM in future context.

References

1. Bondarouk T, Ruël H (2013) The strategic value of e-HRM: results from an exploratory study in a governmental organization. Int J Human Res Manag 24(2):391–414
2. Breaugh JA, Starke M (2000) Research on employee recruitment: so many studies, so many remaining questions. J Manag 26(3):405–434
3. Bhagyalakshmi R, Maria EF (2021) Artificial intelligence and HRM: an empirical study on decision-making skills of HR through AI in HRM Practices. Annals RSCB 25(6):11568–11578
4. Chang K (2020) Artificial intelligence in personnel management: the development of APM model. Bottom Line 33(4):377–388
5. Chapman DS, Gödöllei AF (2017) E-recruiting: using technology to attract job applicants. In: Hertel G, Stone D, Johnson RD, Passmore J (eds) The Wiley Blackwell handbook of the psychology of the internet at work. Wiley Blackwell, Hoboken, pp 211–230
6. Christopher A (2019) Use of artificial intelligence in human resource management. Retrieved from https://albertchristopherr.medium.com/use-of-artificial-intelligence-in-human-resource-management-ddb4e4de9c6e
7. Davenport TH, Ronanki R (2018) Artificial intelligence for the real world. Harv Bus Rev 96(1):108–116
8. Flasiński M (2016) Introduction to artificial intelligence. Springer International Publishing, Switzerland
9. Johnson RD, Stone DL, Lukaszewski KM (2021) The benefits of eHRM and AI for talent acquisition. J Tourism Futures 7(1):40–52. https://doi.org/10.1108/JTF-02-2020-0013
10. Kshetri N (2021) Evolving uses of artificial intelligence in human resource management in emerging economies in the global South: some preliminary evidence. Manag Res Rev, vol ahead-of-print No. ahead-of-print
11. Kumari PB, Hemalatha A (2019) Perception towards artificial intelligence in human resources management practices–with reference to IT companies in Chennai. Int J Recent Technol Eng 8(4S3):61–65

12. O'Connor SW (2020) Artificial intelligence in human resource management. Retrieved from https://www.northeastern.edu/graduate/blog/artificial-intelligence-in-human-resource-management/
13. Pillai R, Sivathanu B (2020) Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. Benchmarking: Int J 27(9):2599–2629

Effects of Technology and Roles of IT Softwares (SPSS and AMOS) for Factors Influencing Patients' Intention to ByPass to Higher-Level Hospitals in Northern Vietnam

Pham Van Tuan and Dinh Tran Ngoc Huy

Abstract First of all, Information technology has many applications in hospitals and helped doctors to take care of their patients, observe their treatment schedules and hence, demonstrate a great influence on many aspects of social life. Next, the applications of IT softwares such as SPSS and AMOS have helped scientists in researches in hospitals for patient treatment very much. Our study shows that because hospital overcrowding caused by actions of patients to self-transfer to higher-line hospitals recently, as a consequence, big effects on costs and cross-infection risk, hospital fees payment of Social insurance. Hence, SPSS software and AMOS software are used, for EFA discovery and CFA affirmation factor analysis. Study presented the most positive impact is attitude (0.332), then Reference Group impact positively on Awareness (0.295).

Keywords SPSS software · AMOS software · Technology applications · Reference group · Perceived behavioral control · Attitudes · Hospital self-transferring · Bypass

JEL classification O14 · O32 · O33

1 Introduction

First we recognize important roles of technology applications in hospitals, esp. under covid 19 effects. Moreover, information technology has various applications of high techniques in hospitals for taking care of patients, in training and teaching as well as

P. Van Tuan
National Economics University, Ha Noi, Vietnam
e-mail: Phamvantuan@neu.edu.vn

D. T. N. Huy (✉)
MBA, Banking University HCMC, Ho Chi Minh city, Vietnam
e-mail: dtnhuy2010@gmail.com

International University of Japan, Niigata, Japan

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_54

in medicine research. Second, we realize the vital roles of using IT software SPSS and AMOS in performing quantitative analysis on this topic [1–3].

Steps to use SPSS software in this study:

Step 1: Using SPSS 20.0 software, the author has drawn the results of the reliability of the scale (Cronbach Alpha) and the results of the EFA (Discovery Factor Analysis).
Step 2: We perform CFA (Positive factor analysis) for the scale, with support of AMOS software version 22.0.
Step 3: SEM model and Testing research hypotheses.

According to the Health Strategy and Policy Institute (2013), the cause of the overload is partly due to the steady economic growth in Vietnam. This led to the need for medical examination and treatment. Health care requirements are higher, while hospitals and health facilities at lower levels are still limited, thus causing overcrowding in higher levels of hospitals. Besides, there are numerous reasons behind the overload in higher hospitals such as inadequacies of health insurance policies, inappropriate hospital bypass mechanisms, and autonomy policy of hospitals.

In summary, there is technology applications of SPSS and AMOS softwares in order to conduct CFA and EFA analysis [4–8].

2 Methodology and Model

Approaching the theoretical framework TRA and overview of the above studies with hypotheses (H1 $\rightarrow$ H7) the research team built a research model to consider the degree of influence on the intention of patients bypassing to higher level hospital (Fig. 1).

From the basic theory of the TRA model and the above documents, the authors boldly proposed models of hypotheses that influence the intent of outpatient behavior under the impact of slicing variables as geography, patients, policy. This model has variables considered in different impacts [9–12]:

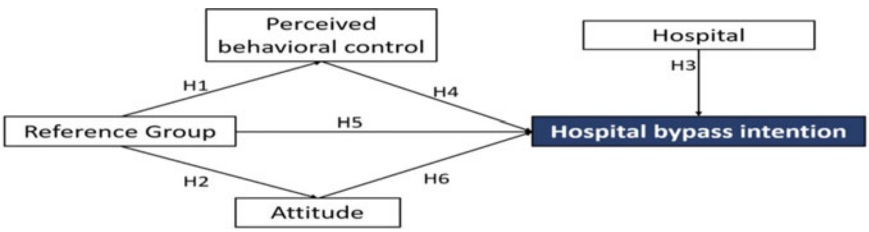


Fig. 1 Research model. *Source* Research group

In the paper, the effect of Hospital, Perceived Behavioral Control, Reference Group, Attitude, Patient, Geography, and Policy toward Intention can be classified into types of variables:

- Independent variables: Hospital, Perceived Behavioral Control, Reference Group, Attitude, Patient, Geography, Policy.
- Dependent variable: Intention to self-transfer to higher-hospitals.
- In the study the impact of the Reference Group on Perceived Behavioral Control and Attitude.
- Independent variables: Reference group.
- Dependent variable: Perceived Behavioral Control, Attitude.

2.1 Research Methods

Research is done through two steps: preliminary research and formal research. In details, in the qualitative analysis with survey, we conducted in-depth interviews with marketing, behavioral and health experts from the two National Economics University and the Hanoi School of Public Health and focus group interviews with 20 people to complete draft scale 1 and create draft scale 2.

Beside, Pham Thi Bich Ngoc, Dinh Tran Ngoc Huy, Pham Thi Hong Nhung (2021) specified we need to enhance quality of hospitals to serve patients better. Huy, D.T.N (2015) said we need to enhance better governance system in firms including hospitals. Moreover, a study on current situation of overcrowding, under-crowding in hospitals (Health Strategy and Policy Institute, 2013). And Nguyen Viet Tri, Lam Hong Son, Vo Thi Diem (2015) made assessment of the bypass situation according to Circular No 14/2014/TT-BYT at Ca Mau Obstetrics and Pediatrics Hospital. Therefore, this study differs from previous studies in a sense that it will identify effects of SPSS and AMOS softwares on Factors Influencing Patients' Intention to ByPass to Higher-Level Hospitals in Northern Vietnam. So there will come the novel of this paper.

In summary, authors use TRA model to estimate degree of influence on the intention of patients bypassing to higher level hospital [13–17].

3 Main Findings

3.1 Technology and Information Technology Effects in Hospitals

According to a report of the Ministry of Health, the application of IT in the health sector has achieved many outstanding results in recent years. In addition to building and gradually forming a legal corridor on medical IT application, almost all hospitals already have hospital information system software, initially deploying transmission and storage software images (PACS); 99.5% of hospitals have connected and linked medical examination and treatment data of health insurance (HI) with the agency for assessment and payment of health insurance, using for the task of e-health insurance assessment. Next, the National Single Window and the ASEAN Single Window were implemented. Preventive health management information system, deploying expanded vaccination software in the North, has more than 3.2 million vaccinated subjects and managed health books [18–21].

Initially forming a telemedicine network, contributing to improving the quality of medical examination and treatment at lower levels, mountainous and disadvantaged areas. Promoting the application of information technology (IT) in all hospital activities toward building a smart hospital is the responsibility of hospital leaders. Based on the actual situation of the hospital, list IT applications in order of priority that need to be implemented toward four specific goals, including: (1) constantly improving service quality patient services, (2) building a favorable working environment for staff both professionally and professionally, (3) reforming administrative procedures and the quality of hospital management, and (4) contributing data sources. General data for the health sector in state management, coordination and forecasting. Improving the hospital management information system (Hospital Information System, HIS) is a top priority in implementing IT application of the patient. The laboratory, followed by the laboratory management system (Laboratory Information System, LIS), the management system for storing and transferring images, etc. Information must be integrated in order to exchange data smoothly to form a uniform system that ensures accuracy, completeness, and meets management requirements, management, expertise, patient service [22–24].

3.2 *Using IT Software SPSS*

3.2.1 EFA Analysis

Using SPSS 20.0 software, the author has drawn the results of the reliability of the scale (Cronbach Alpha) and the results of the EFA (Discovery Factor Analysis) to eliminate some observed variables, helping the scale. Evaluate the factors more accurately. The test standard is Cronbach Alpha $> = 0.6$ and the extraction variance is greater than 50% [25, 26]. Specific results are summarized in Table 1.

Based on the result Table 1, the scales of the model are all > 0.6 and there is no case of eliminating observed variables. Therefore, all observed variables are accepted and will be used in subsequent factor analysis.

3.2.2 Positive Factor Analysis (CFA)

After preliminary testing of the scale, the author continues to use AMOS software version 22.0 to conduct CFA (Positive factor analysis) for the scale, to check the

Table 1 Reliability summary and average variance extraction

No	Scale	No. of observed variables	Cronbach's alpha	Source and calibrated scales
1	Hospital (HOS)	3	0.934	Phùng Thị H`ông Hà et al. (2012) ThS. H`ô Bạch Nhật (2014) de Cruppé W et al. (2017)
2	Attitude (TT)	6	0.893	David S. Martin et al. (2011)
3	Perceived behavioral control (PBC)	8	0.905	Christina L. Jonesa et al. (2014) Jilan Ali Ibrahim Al-Battawi et al. (2017) U.S. Public Health Service (1950)
4	Reference group (RG)	6	0.838	Kara Chan (2000) Mohamad hasnan Ahmad et al. (2014)
5	Intention (INT)	5	0.868	Ing-Long Wu et al. (2005) Rebecca Cameron et al. (2012)

Source Made by authors

appropriateness of the research model and hypotheses-research. The test criteria used include CMIN/df; Kindness of Relevant Index; Tucker and Lewis index; Relevant index comparison; Approximate squared error. The model is considered suitable when testing Chi squared with P value ≥ 0.05. CMIN/df < 5; GFI, TLI, CFI ≥ 0.9 (Bentler and Bonett 1990); RMSEA ≤ 0.08 shows the research model is suitable. CFA results of the above scales are presented in the following Fig. 2:

As in Fig. 2, we see TLI and CFI are both greater than 0.9; Chi-square/df < 5 and RMSEA < 0.08, proving that the scale fits perfectly.

3.3 *Testing Models and Research Hypotheses*

3.3.1 Model Testing

After testing the scales, the author conducted an official test for the research model:

Test results of research model (Fig. 3) show that: Chi-square/df = 2,669; TLI = 0.896; CFI = 0.907; RMSEA (model relevance to overall) = 0.073. Although the TLI index has not yet fully met the standard, but considering the TLI ~ 0.9 is very

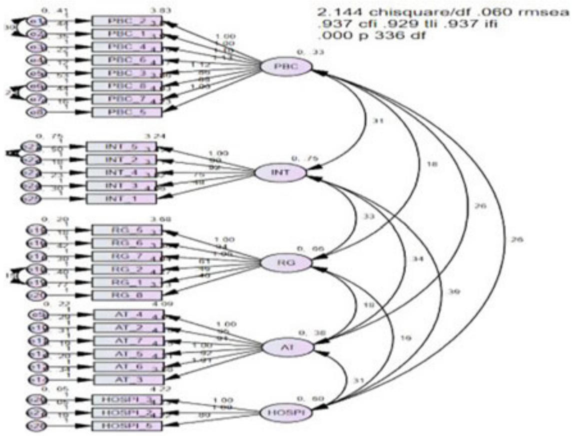


Fig. 2 CFA test results for the scale. *Source* Made by authors

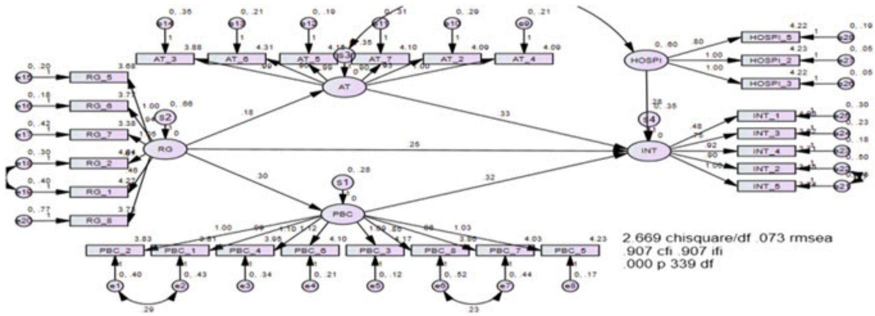


Fig. 3 SEM model results. *Source* Made by authors

close to the standards of Bentler and Bonett (1990), this model is still suitable for market data.

3.3.2 Testing Research Hypotheses

From Table 2, we recognize relationships as statistically significant (P < 5%).

In addition, Table 3 shows among the 4 factors Hospital, Reference Group, Perceived Behavioral Control, and Attitude; Attitude had the most positive impact on the intention to cross the hospital line (0.332).

In summary, authors use above softwares to make analysis and find out the model which is proper and observed variables are accepted and will be used in subsequent factor analysis.

Table 2 Results of hypothesis testing

Relationship	Estimate	S.E	C.R	P
AT ← RG	0.177	0.039	4.553	***
PBC ← RG	0.295	0.046	6.380	***
INT ← PBC	0.322	0.082	3.926	***
INT ← HOSPI	0.285	0.070	4.057	***
INT ← RG	0.249	0.061	4.077	***
INT ← AT	0.327	0.096	3.410	***

Source Made by authors
In which: Estimation: estimated average value; SE: standard deviation; CR: critical rate; P: significance level; ***: $p < 0.001$

Table 3 Compare variable model with partial in variance model according to the gender of the patients

Comparative model	χ2	df	*p*	TLI	CFI	RMSEA
Invariant	1377,6	684	0,000	0,877	0,888	0,057
Variable	1373,913	678	0,000	0,875	0,888	0,057
Value difference	3,687	6	0,000	0,000	0,000	0,000

Source Made by authors

4 Discussion

Compare market segments: Analyzing the structure of multi-group demographic variables:

Analysis of multi**-group structure of demographic variables for men and women**

The research team used the multi-group structure analysis method to compare the research model showing the impact of the hospital, reference group, perceived behavioral control, and attitude to intention to bypass local hospitals by gender group (male/female).

SEM results of the variable model for 2 groups of male and female customers: $\chi2 = 1373{,}913$; df $= 678$; $p = 0{,}000$; $\chi2/\text{df} = 2{,}026$; TLI $= 0.875$; CFI $= 0.888$; RMSEA $= 0.057$.

SEM results of partial invariant models for 2 groups of male and female customers $\chi2 = 1377.6$; df $= 684$; $p = 0{,}000$; $\chi2/\text{df} = 2{,}014$; TLI $= 0.887$; CFI $= 0.888$; RMSEA $= 0.057$.

Demonstrate that both the variable and the invariant models of both male and female clients are consistent with market data.

The results of testing the difference of compatibility criteria between the variable and partial invariants and show that the differences between the two models are statistically significant ($p = 0.00 < 0.05$). Therefore, the variable model was chosen and allowed to conclude that there were differences between male and female patients in assessing the impact of the Hospital, Reference Group, Perceived behavioral control, and Attitude on Intention to bypass hospitals within Northern Vietnam regions.

In summary, authors recognize the variable and the invariant models of both male and female clients are consistent with market data.

5 Conclusion

According to our results, the attitude has the most positive impact on bypassing higher-level hospitals behavior (0.327). This result reflects the current situation at the commune, ward, district, and provincial health centers. Along with the incomplete provision of information on medical examination and the weak ability of lower-level hospitals, patients' awareness will inevitably lead to psychological confusion and fear of patients. At the same time, with the development of the Industrial Revolution 4.0, fake news have dominated the media and become a source of information, or a reference group. The reference group has a positive influence on awareness (0.295), once again affirming that information and communication activities about health from governments, hospitals, health organizations are very important. Therefore, all stakeholders should be in charge of conveying accurate health and disease information to the public. Furthermore, hospitals seem to have monopoly power which is a barrier to patients whereas patients are customers who pay bills to be cured and taken care of. Thus, grassroots hospitals and health care facilities need to improve service quality and increase customers' satisfaction to maintain customer loyalty (patients).

In general, the thesis has affirmed the hospital factors, reference groups, awareness, positive attitude to the intention to bypass, and this result is consistent with market data and research work. Moreover, to increase the value of the research, the team conducted tomography, market segmentation analysis, conducted the NPar test that showed the age of patients who intend to exceed the highest hospital line of the group is from 31–50 years old. When applying the NPar test to geographic areas, the results show that in the mountainous areas, rural areas, provincial cities, and central cities, patients in rural areas intend to bypass the most (199,32). This also clearly reflects the situation of the age and the health care system at the grassroots level is often worse than the central level. The team continued to use the NPar test with the medical examination and treatment services used by the patient when crossing the line. Among the Internal Medicine, Surgery, Obstetrics, Pediatrics, and Other Services (geriatric, ophthalmic, psychiatric, infectious …) patients have the most desire to use Internal Medicine when they have the intention to bypass (171.64). Besides, patients who want to use Surgical services are also much higher (164.20) than patients who need to use Obstetrics, Pediatrics, and Other Services when bypassing to higher-level hospitals. Also, people need to voluntarily raise awareness and accountability to the community. Governments and hospitals should pay attention to using appropriate media for the majority of the population. For example, during the recent COVID-19 epidemic, the Ministry of Health and the

Prime Minister have sent messages to citizens daily and released the song "Ghen Cô Vy" with a catchy tune that instructed hand washing correctly. These are effective means of communication in the 4.0 revolution. If information is applied regularly to the propaganda or the dissemination of health care for infectious diseases such as tuberculosis or COVID-19, it can positively impact citizens' awareness, thereby orienting proper behavior.

Acknowledgements This research is funded by National Economics University, Hanoi Vietnam

References

1. Birk et al (2011) Patients' experience of choosing an outpatient clinic in one county in Denmark: results of a patient survey
2. General Statistics Office of Vietnam (2018) Number of patient beds in 2018 by province
3. Health Strategy and Policy Institute (2013) Study on current situation of overcrowding, under-crowding in hospitals at levels and recommended solutions for improvement
4. Huy DTN, Dat PM, và Anh PT (2020) Building and econometric model of selected factors' impact on stock price: a case study. J Secur Sustainab Issues 9(M):77–93. https://doi.org/10.9770/jssi.2020.9.M(7)
5. Hanh HT, Huy DTN, Phuong NTT, Nga LTV, Anh PT (2020) Impact of macro economic factors and financial development on energy projects-case in ASEAN countries. Management 24(2)
6. Hanh HT, Chuyen BM, Huy DTN, Phuong NTT, Phuong HTL, Nuong LN (2020) Challenges and opportunities from impacts of FDI and income on environment pollution: role of financial accounting transparency in FDI firms. J Secur Sustainab Issues 10(2)
7. Huy DTN, Trung-Hieu L, Hang NT, Sylwia G, Trung ND, Tuan PV (2021) Further researches and discussion on machine learning meanings—and methods of classifying and recognizing users gender on internet. Adv Mech 9(3)
8. Huy DTN, Nhan VK, Bich NTN, Hong NTP, Chung NT, Huy PQ (2021) Impacts of internal and external macroeconomic factors on firmstock price in an expansion econometric model—a case in vietnam real estate industry. Data Sci Financ Econometric-Stud Comput Intell 898
9. Huy DTN (2015) the critical analysis of limited South Asian corporate governance standards after financial crisis. Int J Q Res 9(4)
10. Huy DTN (2012) Estimating Beta of Vietnam listed construction companies groups during the crisis. J Integr Dev 15(1):57–71
11. Huy DTN, Loan BT, Anh PT (2020) Impact of selected factors on stock price: a case study of Vietcombank in Vietnam. Entrepreneurship Sustainab Issues 7(4):2715–2730. https://doi.org/10.9770/jesi.2020.7.4(10)
12. Ministry of Health (2013) Circular: providing for technical and professional levels in the network of health facilities, No. 43/2013/TT-BYT, 11 Dec 2013
13. Ministry of Health (2017) Guidance: How to implement a pilot model at 26 commune health stations from 2018 to 2020, No 1383/HD-BYT
14. Tho ND, Trang NTM (2007) Study on science of marketing—apply the structural equation model (SEM), published by Vietnam National University in Ho Chi Minh City
15. Tho ND, Trang NTM (2007) Research on marketing science—applying SEM linear structure model, National University Publishing House, Ho Chi Minh City. HCM
16. Tri NV, Son LH, Diem VT (2015) Assess the bypass situation according to Circular No 14/2014/TT-BYT at Ca Mau Obstetrics and Pediatrics Hospital in 2015
17. Ha PTH, Hien TTT (2012) Assess the customer satisfaction of medical service at Viet Nam—Cuba (Dong Hoi—Quang Binh province)

18. Ha PTH, Hien TTT (2012) Evaluation of customer satisfaction using medical services at Vietnam—Cuba Dong Hoi—Quang Binh hospital
19. MA. Ho Bach Nhat (2014) The inpatient satisfaction of medical and examination service in hospitals in Long Xuyen city
20. MSc. Ho Bach Nhat (2014) Inpatient satisfaction with the quality of medical examination and treatment services of hospitals in Long Xuyen city
21. Ngu DT, Huong DT, Ngoc DT, Thanh PT, Döngül ES (2021) Language teaching application to English students at master's grade levels on history and macroeconomic-banking management courses in universities and colleges. J Lang Linguist Stud 17(3):1457–1468
22. Anh P, Huy DTN, Phuc DM (2021) Enhancing database strategies for management information system (MIS) and bank sustainability under macro effects—a case study in vietnam listed banks. Acad Strategic Manag J 20(6)
23. Phuong NTT, Huy DTN, Tuan PV (2020) The Evaluation of impacts of a seven factor model on NVB stock price in commercial banking industry in Vietnam—and roles of disclosure of accounting policy in risk management. Int J Entrepreneurship 24:1–13
24. Tuan PV, Huy DTN, Duy PK (2021) Impacts of competition selection strategy on firm risk—case in Vietnam investment and finance industry. Revista Geintec-Gestao Inovacao E Tecnologias 11(3)
25. Ngoc PTB, Huy DTN, Nhung PTH (2021) Healthcare policy for patients with chronic heart failures at Nam Dinh General Hospital in Vietnam. J Pharm Res Int 33(40B):292–290
26. Ngoc TT (2016) Thematic internship report: "Study on the influence of reference group towards the behavior to choose medical and examination service of North citizen, Vietnam"; Ngoc TT (2016) Internship topic "Study on the influence of reference groups on the behavior of residents in choosing medical examination and treatment services in the North"

VLSI Design and Its Application

Comparative Analysis of Power Reduction Techniques for 6T, 8T and 9T SRAM

Gurjit Kaur, Samarth Nigam, and Ronitt Mehra

Abstract The growing demand of electronic devices particularly the mobile devices is demanding for minimal power consumption, which leads to increase in the battery life of the device and enhances its system. Static Random-Access Memory (SRAM) is operated in high-speed configurations, for example in the cache memories. SRAM occupies about 90% of silicon area. A large number of researches and studies done in the field of power reduction techniques of SRAM and a number of findings have been deduced from the research. There are majorly 2 methods of decreasing the total power consumption in SRAM; the two methods are either by lowering dynamic power or by lowering the total power in stand-by mode. Through the means of this comparative analysis our objective is to give a definitive and conclusive analysis of the numerous power dissipation control techniques in 6T CMOS-SRAM. The various techniques such as reducing the voltage swing, shared-bit line architecture, quantizing the overall structure. Our primary focus will be on 6T CMOS SRAM, for which we will be analyzing its power reduction techniques and thereby comparing it with other types of configurations like 8T and 9T CMOS SRAM. Finally, we will deduce this report by drawing the conclusions between all the mentioned configurations.

Keywords 6T CMOS SRAM · 8T CMOS SRAM · 9T CMOS SRAM · Shared-bit line · Banked organization

1 Introduction

In the current world situation, everyone is in search of electronic devices and peripherals that are held in top-notch performance. SRAM plays a crucial role in the generic structure of cache memory [1]. High SRAM power dissipation reduces the overall function productivity and contributes to integrated circuits overheating which is

G. Kaur · S. Nigam (✉) · R. Mehra
Delhi Technological University, Delhi 110042, India
e-mail: samarth109nigam@gmail.com

G. Kaur
e-mail: gurjitkaur@dtu.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_55

certainly not a desirable characteristic. So, by making use of low power SRAM is generated due to the same [2]. And if the total production of heat within the IC decreases, the energy budget along with the production costs will also decrease. In an SRAM, there are mainly two types of power consumptions: Dynamic Power and Standby Power [3]. Dynamic Power is that power which comes into the picture during the read and write operations of SRAM, whereas Standby Power is that power which comes into the picture when a bit is stored inside the SRAM cell (because of the practical limitations present in MOSFET gates that are inherently present inside the SRAM cell). Various power mitigation measures such as 'shared-bit line architecture' [4], 'reducing the voltage swing' [5] and 'reprocessing of charges' [6] in SRAM and have been suggested. In this paper, we conclude and analyze these techniques in depth and draw their inter-comparisons along with drawing a comparison with existing SRAM models like the 6T, 8T and 9T SRAM configurations [7]. Ultimately, we classify various configurations suited for different purposes for the sake of optimization.

2 Conventional 6T CMOS SRAM Cell

Figure 1 is a typical 6T CMOS SRAM cell. This configuration comprises two p-MOS and four n-MOS Field Effect Transistors (FETs). The structure resembles 2 cross coupled inverters and two n-MOS playing the role of access transistors [8].

Operations: The 6T CMOS SRAM can perform three operations in totality: 'Hold Operation, Read Operation and the Write Operation' [9]. Hold Operation is performed in the stand-by mode of the SRAM cell. In case of the Hold Operation, the Word Line is set to zero, thus inactivating the access transistors. For the Hold Operation, the 6T CMOS SRAM cell can either hold the value '0' or the value '1'.

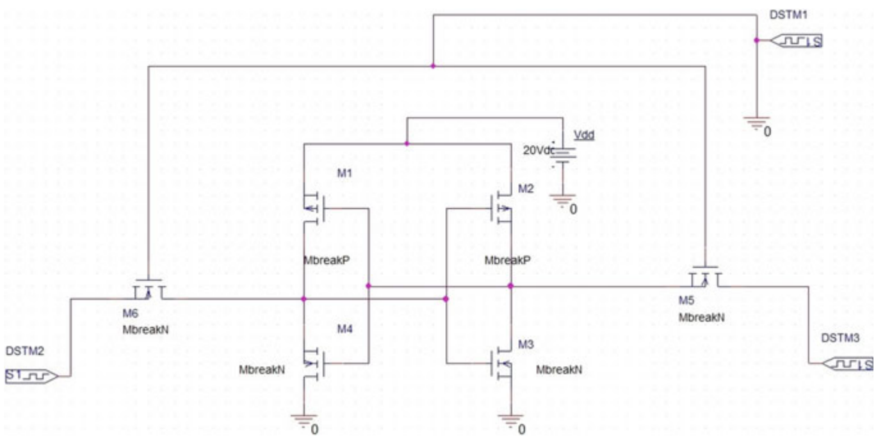


Fig. 1 Typical 6T SRAM cell

As discussed above, the ultimate cause for dissipation of power in case of stand-by mode is the leakage current.

The Read and the Write Operations are performed in the dynamic mode of the 6T CMOS SRAM cell. In case of the Read Operation, the word line gets set to one, thus activating both the access transistors. This allows for direct access of the bit lines to the cross-coupled CMOS inverters. Both the bit lines are then pre-charged. The value present inside the SRAM Cell can be either '0' or '1' [9]. Depending on the value present inside the SRAM cell, either one of the two-bit lines will find a straight path to the ground terminal thus leading to its discharging. Since, in both the cases, only one of the two-bit lines gets discharged, a potential difference is generated between the two-bit lines. This potential difference is sensed using a sense amplifier. Based on the potential difference, the sense amplifier can easily predict the value present inside the SRAM cell, thus successfully performing the 'Read' operation. For the purpose of the write operation, the value that we wish to insert into the SRAM cell is pushed in through the bit line, and after constant switching of the transistors present inside the SRAM cell, the value gets settled down, thus successfully performing the 'Write' operation. In both the Read and Write operations, due to constant switching activity of the MOSFETs, power is dissipated. This power dissipation due to switching of transistors in dynamic mode of SRAM cell is known as dynamic power consumption.

3 Techniques to Reduce Power

There are various techniques that are being used apparently to optimize the power consumption in an SRAM cell. But no such technique is perfect, and each has some repercussions associated with it. So, our approach is to always find the best possible technique, i.e. the one with the least repercussions. Or simply we can also use our knowledge to choose the best possible technique for the particular scenario at hand, thus broadening our horizons. Some of the techniques for Power Reduction that we wish to discuss are as follows:

3.1 Reduction of the Supply Voltage

Conceptually, inside the memory, during the charging and discharging of bit lines, one of the two-bit lines gets fully charged whereas the other one loses its charge by getting discharged into the ground. Thus, power dissipation is bound to be there. Continuous power differs directly from the potential difference between the two-bit lines (ΔVBL), which get optimized when we reduce the voltage of the power supply, recycling of charge, sense-amplifying cell etc. The Power Dissipation in an SRAM cell is in direct relation to the voltage of total power supply applied. So, in order to reduce the power consumption, we can reduce this voltage. However, it has a repercussion that reduction of this supply voltage results in the delay of gates

Table 1 Observational values of power of 6T CMOS

Supply (voltage)	Power (W)	Gate delay (ns)
1.5	11.456	2.2
3.0	15.814	0.63
4.5	21.046	0.42
6.0	26.425	0.28

to a greater extent due to a reduction in the current flowing through the transistors (according to ohm's law, which states the direct proportionality of voltage with current in a circuit). Thus, this technique is especially effective when the delay of the gates is not of a major concern. Table 1 given down below summarizes an effective comparative study for the power dissipation in different SRAM cells for various estimates of supply voltage [5]. It can be easily seen that for the lowest output of supply voltage, the power dissipation comes out to be the least.

3.2 Banked Organization

The Banked Organization approach focuses to reduce the power consumption is more of a change in layout as opposed to a structural change. It is only effective when more than one SRAM cell is present. It makes use of the fact that the power dissipation in SRAM cells is present due to the active capacitance and resistance of the Bit Lines as well as the Data Lines. So, if the wiring used is reduced, the capacitance would also reduce, thus leading to a reduced value of total power being consumed in the SRAM cell. The Banked Organization approach involves the use of the fact that access to the memory follows a tabular approach, segregated in the form of rows and columns, thus forming blocks. If the memory is divided into further sub-blocks, we would be able to reduce the length of the wire in half, thus reducing the total capacitance of the bit lines (Fig. 2).

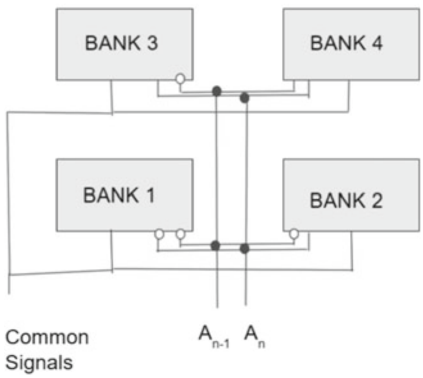


Fig. 2 Banked organization of SRAM

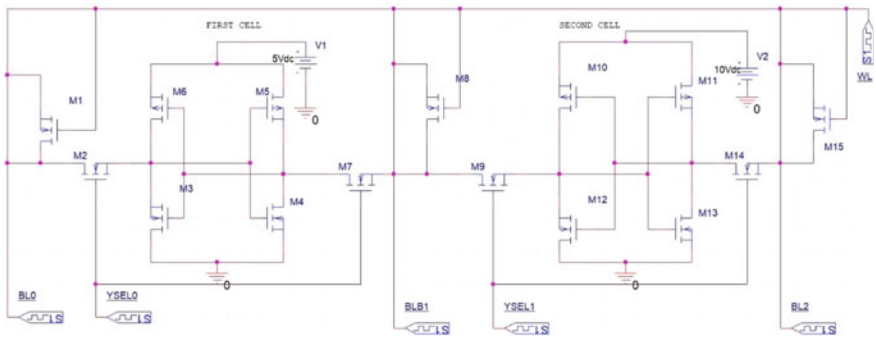


Fig. 3 Shared-bit line architecture

3.3 Architecture Involving Shared-Bit Line

This technique is an architectural change that helps in the reduction of power in SRAM. This technique involves reduction in the wastage of memory cell current, but at the same time it does not lead to a reduction in the surface area occupied by the cell architecture [4]. To design this architecture, 2 extra Access transistors get attached to the SRAM cell, and are shared by the adjacent cells. Analyzing two adjacent cells in this architecture, when the word line is set to '0', the cells are in segregation from one another. As soon as the word line is set to '1', the architecture gets activated. In this architecture, two adjacent SRAM cells share the same bit line. The presence of seven MOSFETs per SRAM cell leads to an increase in the silicon area, since one-bit line is shared between two cells, the capacitance is reduced. Even though the dynamic power gets reduced through this technique, it is ineffective while performing read and write operations because the data present in the middle bit line (the one that is shared between the two cells) will always clash. So, it is not possible to write combinations like '11' or '00' because the two-bit lines corresponding to an SRAM cell must contain inverse data (Fig. 3).

4 Disadvantages and Drawbacks

Herein we have tracked down the various shortcomings of the aforementioned techniques. Firstly, the reduction in the voltage of the power supplied causes a delay in the gates which is because of the reason that there is a reduction in the current flowing across the circuit. The banked organization approach has one small limitation that it is useful only when there are multiple SRAM cells present whereas when our focus is on a single SRAM cell, it is not applicable. The shared-bit line architecture approach, which is primarily an architectural change, also poses a major disadvantage that it limits the data that can be processed during read/write operation due to clashing of

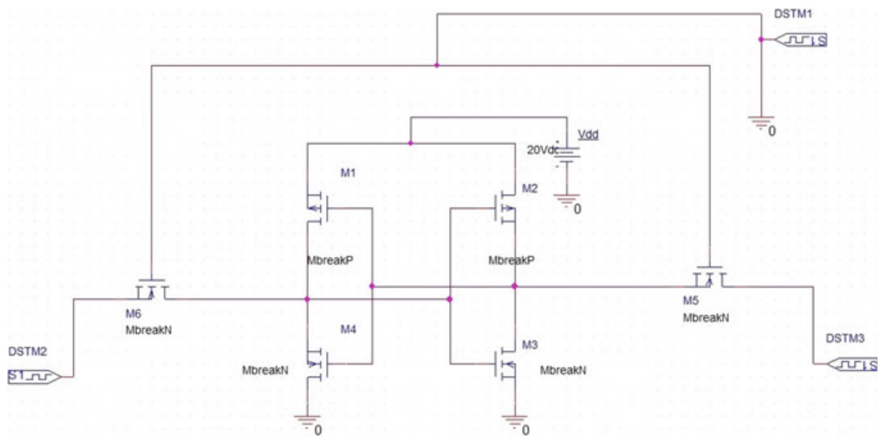


Fig. 4 6T CMOS SRAM CELL

bits in the adjacent SRAM cells. Therefore, we must consider other approaches as well. Now we shall see certain other configurations of the SRAM models like the 8T and 9T modeling. Then we shall perform a comprehensive analysis of the same.

5 Analysis of 6T, 8T and 9T

The MOSFETs used in SRAM cells are based on 90 nm technology, which implies that the length of the transistor is set to 90 nm. The length of each MOSFET is set to 90 nm. The width of the transistors is variable across different models of SRAM [10]. The structures of the six-T, eight-T and nine-T CMOS SRAM are shown in Figs. 4, 5 and 6.

6 Results

An in-depth comparison of the three SRAM models leads to the conclusion that both the read as well as the write delays of the 6T CMOS SRAM are lesser than that of the 8T SRAM model. Experimental research shows that the read delay of 6T SRAM is close to about 7 times lesser as compared to that of the 8T SRAM model [7]. On the other hand, the write delay of the 6T SRAM model is approximately 1.3 times lesser than that of the 8T SRAM model. As part of the results, we observe that the total power dissipated in the 8T SRAM models is double that of in the 6T SRAM models. The reason behind this is the greater number of MOSFETs present in the 8T SRAM model as well as the greater complexity of the working mechanism of the 8T SRAM model over the 6T SRAM model. Here in, one thing to note is that the 9T SRAM

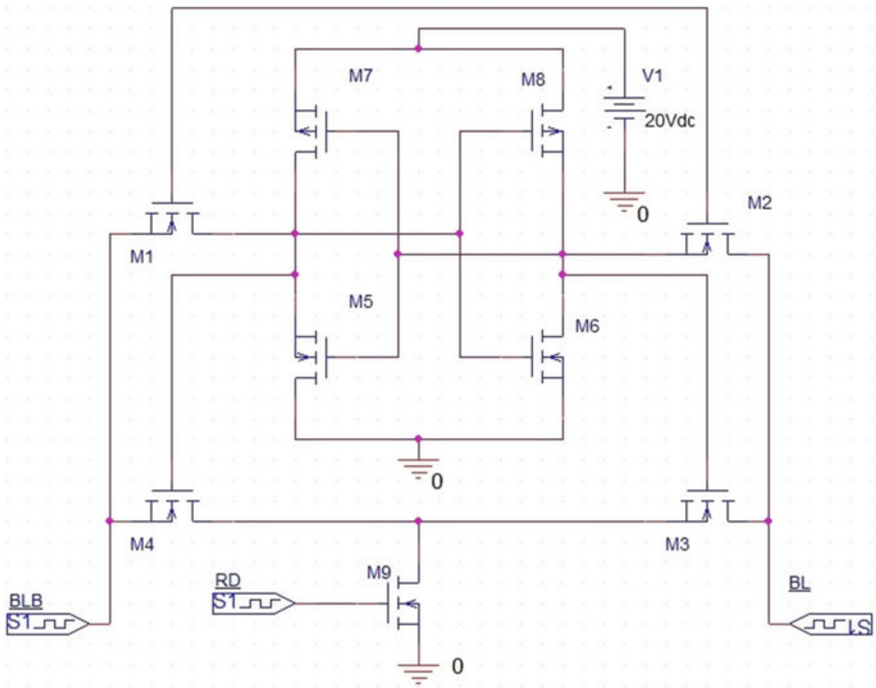


Fig. 5 9T CMOS SRAM CELL

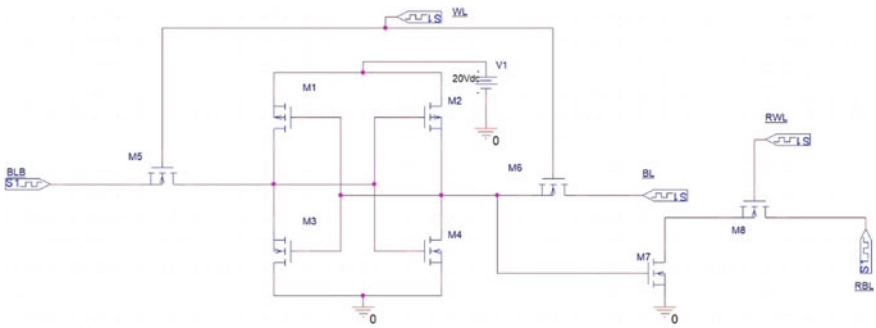


Fig. 6 8T CMOS SRAM CELL

cell configuration offers less power dissipation than the 6T SRAM configuration for the write operation, but it also offers both a greater read as well as write delay. The read delay is the highest in case of the 9T SRAM model because of the use of high voltage MOSFET which greatly improves the driving capabilities but increases the delay. This analysis can be helpful in many applications (Figs. 7 and 8).

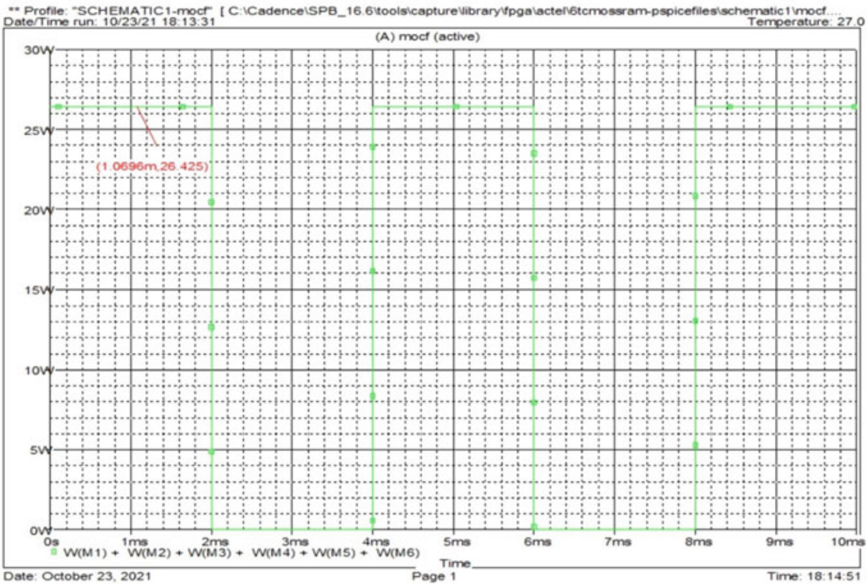


Fig. 7 Output for the power consumption in 6T SRAM

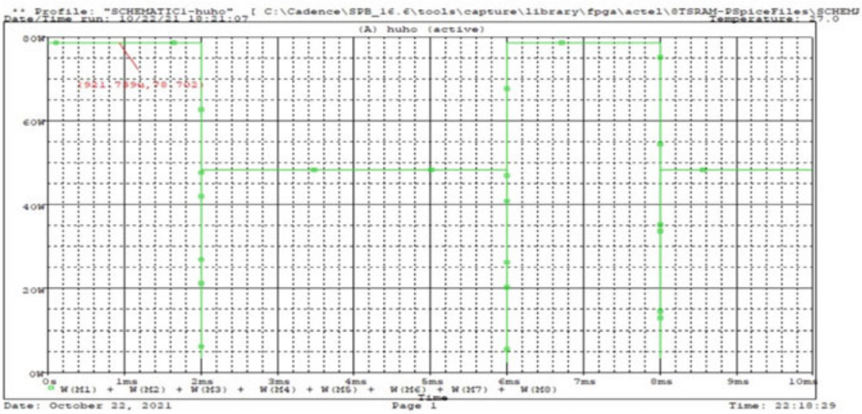


Fig. 8 Output for the power consumption in 8T SRAM

7 Conclusion

An in-depth case study of the different techniques to optimize the power dissipation in 6T, 8T and 9T SRAM memory designs in 180 nm CMOS technology node has been performed. Different design changes that are focused on optimizing the various design parameters have been worked on, some are area efficient, some are power efficient etc. We know that SRAM cells are a fundamental part of cache memory, so we can tell that in terms of power dissipation, the minimum is found out to be in 6T CMOS SRAM configuration. Out of the various power reduction techniques that were being analyzed, reducing the voltage swing gives the least power dissipation value. Henceforth it is highly suited to be used in cache memory. Since cache memory is very much close to the processor, any repercussions (heat caused due to power) in cache memory could lead to IC burn out. So, it is widely used in the electronics industry. Also, the Banked organization configuration is optimum for use in the VLSI industry, as it involves an integration of a large number of SRAM cells, wherein this structure would be beneficial. Apart from power reduction, data retention is an important characteristic when we talk about certain applications like sensor nodes. It is known that when the voltage used as supply in an SRAM cell is decreased, the power dissipation goes down, but the likelihood of encountering retention failures also increases. So, 9T SRAM cells can be used in conjunction with this technique as 9T SRAM cells are known to have the least data retention out of all the configurations and are highly preferable in case of sensor nodes. The 9T model has a better structural integrity, which allows it to achieve better immunity to process variations.

References

1. Agrawal R, Tomar VK (2018) Analysis of cache (SRAM) memory for JGCJHCore i™ 7 processor. In: 2018 9th international conference on computing, communication and networking technologies (ICCCNT), pp 1–8. https://doi.org/10.1109/ICCCNT.2018.8494063
2. Pathak A, Sachan D, Peta H, Goswami M (2016) A modified SRAM based low power memory design. In: 2016 29th international conference on VLSI design and 2016 15th international conference on embedded systems (VLSID), pp 122–127. https://doi.org/10.1109/VLSID.2016.80
3. Kumar M, Ubhi JS (2017) Performance evaluation of 6T, 7T and 8T SRAM at 180 nm technology. In: 2017 8th international conference on computing, communication and networking technologies (ICCCNT), pp 1–6. https://doi.org/10.1109/ICCCNT.2017.8204092
4. Morimura H, Shigematsu S, Konaka S (1999) A shared-bitline SRAM cell architecture for 1-V ultra low-power word-bit configurable macrocells. In: Proceedings. 1999 international symposium on low power electronics and design (Cat. No.99TH8477), pp 12–17. https://doi.org/10.1145/313817.313829
5. Kumar A, Rabaey J, Ramchandran K (2009) SRAM supply voltage scaling: a reliability perspective. In: 2009 10th international symposium on quality electronic design, pp 782–787. https://doi.org/10.1109/ISQED.2009.4810392
6. Kim K, Mahmoodi H, Roy K (2008) A low-power SRAM using bit-line charge-recycling. IEEE J Solid-State Circ 43(2):446–459. https://doi.org/10.1109/JSSC.2007.914294
7. Joshi VK, Lobo HC (2016) Comparative study of 7T, 8T, 9T and 10T SRAM with conventional 6T SRAM cell using 180 nm technology. In: Choudhary R, Mandal J, Auluck N, Nagarajaram H (eds) Advanced computing and communication technologies. Advances in intelligent systems and computing, vol 452. Springer, Singapore. https://doi.org/10.1007/978-981-10-1023-1_3
8. Wang Y, Wang Z, Zhang L (2012) A new 6-transistor SRAM cell for low power cache design. In: IEEE 11th international conference on solid-state and integrated circuit technology, pp 1–3. https://doi.org/10.1109/ICSICT.2012.6467902

9. Majumdar B, Basu S (2011)Low power single bitline 6T SRAM cell with high read stability. In: 2011 international conference on recent trends in information systems, pp 169–174. https://doi.org/10.1109/ReTIS.2011.6146862
10. Athe P, Dasgupta S (2009)A comparative study of 6T, 8T and 9T decanano SRAM cell. In: 2009 IEEE symposium on industrial electronics and applications, pp 889–894. https://doi.org/10.1109/ISIEA.2009.5356318

A Review on CMOS Down Conversion Mixer for High Frequency Applications

Dheeraj Kalra

Abstract In this paper, various topologies of down conversion mixer for high frequency applications are discussed. Common source topology is beneficial for providing the high linearity at the cost of high NF and more area. Common gate topology is helpful in achieving the good input impedance matching but having the low gain. Various circuits of mixer such as Square-law mixer, High isolated square-law mixer, Dual gate Mixer, Double-Balanced Dual Gate Mixer, Single-Balanced Mixer, and Gilbert Cell Mixer are discussed with their advantages and disadvantages.

Keywords CMOS · Mixer · Noise figure

1 Introduction

The receiver converts RF signals to IF signals at the receiver end of communication networks. The quality of a receiver system's performance is controlled by frequency translation technology, conversion gain, filtering type, and frequency selection [1]. Designing a high-quality receiver system is difficult due to the trade-off between mixer parameters like as linearity, noise figure, conversion, and power consumption. In a typical wireless system, the modulated signal is accompanied by extra fake (interferences, noise, picture, etc.) signals at the receiver end [2]. And all of these phoney signals have a lot in common with the favored signal. As a result, receiver blocks should provide certain assured functionality such as (i) expand signal to decrease the influence of other signal present signals (ii) separating required signal from other fake signal and (iii) fine-tuning capability to gather required signal and improving the figure of merits; the receiver's primary determining factor [3]. The super heterodyne receiver system's block diagram is shown in Fig. 1.

The low power RF signal is initially received at antenna, and then it is transformed into an electrical signal before given to band pass filter. The band select filter separates out other signals from the intended signal. The frequency components are present

D. Kalra (✉)
GLA University, Mathura, India
e-mail: dheeraj.kalra@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_56

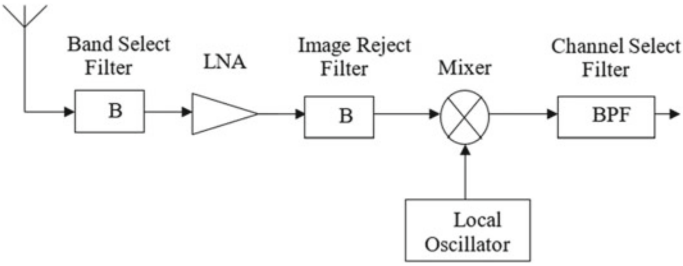


Fig. 1 Block diagram of super heterodyne receivers

in the intended signal. The RF signal is next boosted by low noise amplifier (LNA), used to minimize noise from the subsequent stages within receiving lane [4]. Image-reject filter, which is a BPF, is now linked to the LNA's output. Image-reject filter's job is to exclude from the necessary channel signal any image signal which is equal to twice of intermediate frequency (IF) signal. Finally, RF signal is converted into a lower frequency signal known as IF signal, which is then given to the other channel select filter with the assistance of a mixer [5]. The demodulation procedure at the channel select filter's output can be used to retrieve the required signal. This is known as a single stage down conversion mixer, and it creates a trade-off in the receiver's performance parameters [6]. If IF value is high, picture signal's frequency will be outside of the necessary signal range, and image signal can be eliminated using a BPF [7]. The channel selection filter's Q-factor should be quite high in this situation. The center frequency divided by the 3 dB bandwidth is the Q-factor of a filter (frequency difference). In general, designing a high Q-factor filter is difficult. If the IF value isn't too high, the channel selection filter isn't required. However, because the frequency of the picture signal is so close to the required signal, removing the correct image becomes more difficult in this instance. By multiplying the RF signal and the local oscillator signals, the mixer converts the frequency [8]. The frequency addition and subtraction of the two signals are produced by product of two signals. Up-conversion mixers are used at the transmitter, and down-conversion mixers are used at the reception. The product of two signals is given by Eq. 1.

$$\begin{aligned} &A\cos(w_{\mathrm{LO}}t) * B\cos(w_{\mathrm{RF}}) \\ &= \frac{\mathrm{AB}}{2}[\cos(w_{\mathrm{LO}} + w_{\mathrm{RF}})t + \cos(w_{\mathrm{LO}} - w_{\mathrm{RF}})t] \end{aligned} \tag{1}$$

In a perfect mixer, there is no noise, no maximum amplitude limit, no reliance on LO signal strength, and no intermodulation distortion at the output signal. In addition, there should be no LO or RF signal components in the waveform at the IF output. However, real mixer performance differs from ideal mixer performance and even for a modest number of signals at the input port, IF output can be rather complex. The quantity of conversion gain of passive and active mixer is essentially different. Passive mixers offer high linearity, isolation, and speed at the tradeoff of increased

Table 1 Normal comparison for active and passive mixer

Mixers	Passive	Active
Gain	Low	High
Noise figure	High	Low
Power supply	High	Low
Linearity	High	Low
Power consumption	High	Low
LO Power	High	Medium

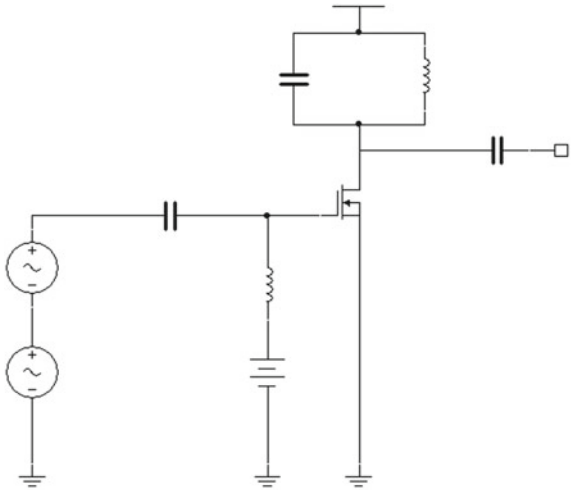
LO power. In comparison to passive mixers, active mixers provide higher conversion gain and need less local oscillator power. Mixer circuits are often employed in the RF industry and are made up of transistor circuits. Active CMOS mixers are simple to incorporate into IC and give excellent IF-LO and RF-LO isolation. Passive mixers, followed by amplifiers, have completely interpreted the CMOS receiver system. Table 1 gives the comparison among active and passive mixers [9].

2 CMOS-Based Mixer Topologies

The following discussion is focused on various mixer topologies. Transconductance, drain, and resistive mixers are the three standard modes of operation for single transistor mixers. The transconductance mixer operates by sending a low-level signal at gate of MOSFETs, which results in gate-source voltage to fluctuate, shifting the transistor's quiescent point between saturation and cutoff zones. The noise figure, linearity, and conversion gain all change when the LO frequency components change. The following are the requirements for a successful mixing procedure and excellent transconductance: To begin, the transistor should be powered at its threshold voltage and biased with a 50% duty cycle transconductance wave. The conversion gain for the transconductance mixer will be maximal if this condition is maintained [10]. In addition, the voltage between drain-source must be continuous and high enough so transistor's quiescent point does not reach the linear area. The RF signal is applied to gate of MOS, while LO signal is supplied to drain terminal. The drain-source voltage changes as a result, and the transistor's operating point shifts from saturation to linear [11]. The mixing of frequencies is caused by variations in the transconductance gm and the drain-source transconductance gds. The resistive mixers use a big LO signal to change the rds (drain to source resistance) while keeping the transistor's operating point in the linear zone. The transistor's gate-source and drain-source are not necessary to keep it in the linear area. Because the transistor operates in the linear zone all of the time, the resistive mixer is also known as a passive mixer.

Mixer's behavior must follow square law and non-linear square terms are desired, while other high order non-linear terms are undesirable. Figure 2 depicts a square law mixer with a gate that receives a series combination of RF and LO signals. At drain, the tuned circuit is needed to keep drain-source voltage constant throughout

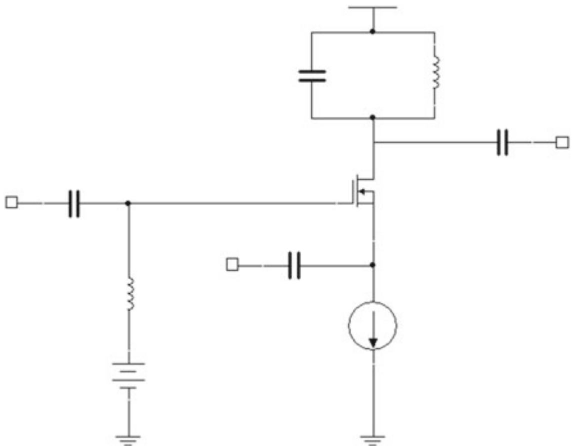
Fig. 2 Square-law mixer



LO signal. Major benefit of this topology is that the high order non-linear components are different from IF signal and hence simply be eliminated [12]. This topology has high conversion gain and as well as low noise figure. However, both the LO and RF signals are sent through the same port, this arrangement has poor RF-LO, LO-RF isolation. The IF-LO isolation of this topology is poor. Figure 3 shows another structure with the same purpose. The LO signal is at source of MOS to offer strong isolation between RF and LO.

The LO signal and RF signal are applied at gate of MOS in dual gate mixers. In a dual gate mixer, the cascade topology is employed. Figure 4 depicts the circuit. The major purpose of the second MOS is to regulate the first MOS transconductance, which aids in obtaining gain. The capacitance impact between LO and RF is decreased

Fig. 3 High isolated square law mixer



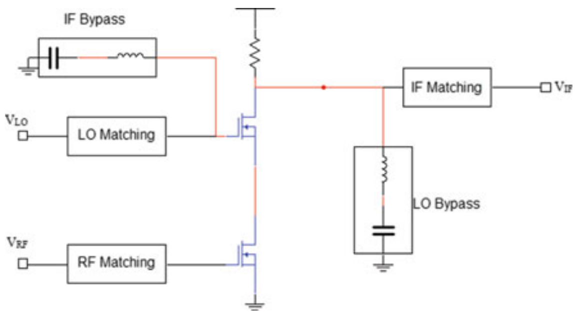


Fig. 4 Dual gate mixer

as well. The benefit of this topology is that the LO and RF ports may be matched separately [13].

This mixer operates in the same manner as a drain mixer. Small amplitude of the RF signal, gate to source i.e. v_{gs} voltage of the lower MOS is nearly constant. The drain-source voltage i.e. v_{ds} alters the lower transistor's operating point in the linear and saturation region. Throughout LO signal, upper transistor's operating point remains in saturation zone. Twin gate mixer's conversion gain is reduced due to the usage of passive component [14]. It is less appropriate for low frequency applications due to passive component. Because there are no passive components, the double-balanced mixer improves port-port isolation while still providing high conversion gain.

LO and RF signals are connected at distinct ports in a single-balanced mixer to ensure good isolation across all three ports. Because a differential LO signal is used in this mixer and the RF signal is single ended, it is referred to as a single-balanced mixer. Figure 6 depicts a single balancing mixer with an RF supplied to bottom

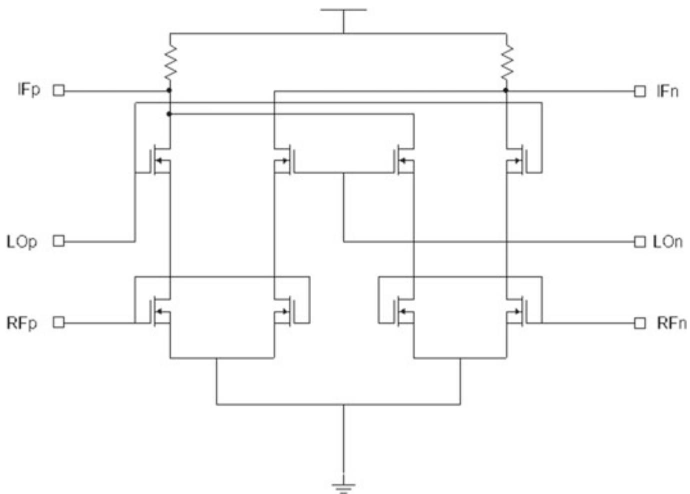


Fig. 5 Double-balanced dual gate mixer

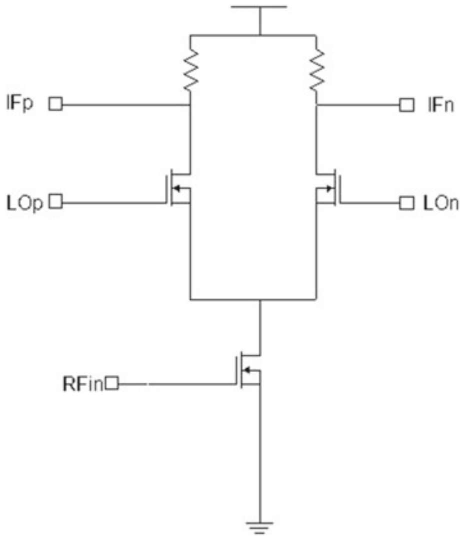


Fig. 6 Single-balanced mixer

MOS and transformed to a current signal. LO transistors have a v_{ds} voltage that is somewhat higher than the threshold voltage. This aids in switching on and off. The operational point of the RF transistor must be in the saturation zone, with only one LO transistor off or on.

The LO might be square, but amplitude must be large enough for the LO transistors to switch properly. The product of the two signals is obtained by multiplying RF signal by LO signal. The leakage between the IF and LO ports is a drawback of this mixer. The LO signal is likewise amplified at the output, necessitating the use of a filter there. Two single-balanced mixer are combined to form the double-balanced mixer. The RF signal, like the LO signal, is differential. Figure 7 shows a double-balanced mixer known as a Gilbert Cell Mixer.

The RF input signal and the LO input signal are linked in parallel. Because the two single-balanced LO signals have opposite phase so LO signal at the output port gets cancelled. The LO-IF ports are more isolated because to this arrangement. When compared to a single-balanced mixer, the primary disadvantage is the high power consumption.

3 Conclusion

As discussed in paper, there are various mixer topologies for improving the performance. There is always trade-off between the mixer parameters. The Gilbert Cell mixer shows the better performance in terms of port to port isolation and the conversion gain. Gilbert cell mixer also gives the optimum value of IIP3. Gilbert cell mixer

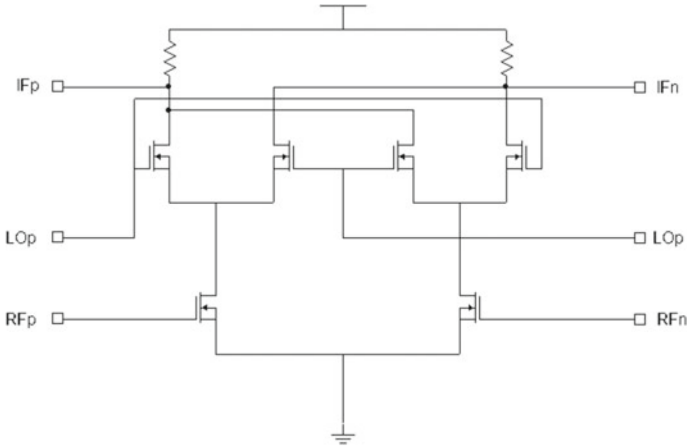


Fig. 7 Double-balanced mixer (Gilbert Cell Mixer)

linearity can be improved by using the source degeneration technique and input matching can be improved by using the π matching at the input side.

References

1. Jouri M, Golmakani A, Yahyabadi M, Khosrowjerdi H (2010) Design and simulation of a down-conversion CMOS mixer for UWB applications. In: Electrical engineering/electronics computer telecommunications and information technology (ECTI-CON), 2010 international conference, pp 937–940. 19–21 May 2010
2. Anh-Tuan, Kim C-W, Kang M-S, Lee S-G (2004) A high performance CMOS direct down conversion mixer for UWB system. In: IEEE Proceedings of the IEEE, vol 88, pp 609–640
3. Chiou H-K, Chou H-T (2013) An ultra-low power v-band source-driven down-conversion mixer with low-loss and broadband asymmetrical broadside-coupled Balun in 90-nm CMOS technology. Microwave Theor Tech IEEE Trans 61(7):2620–2631
4. Gilbert B (1998) The multi-tanh principle: a tutorial overview. Solid-State Circ, IEEE J 33(1):2–17
5. Katsuji K, The super-multi-tanh technique for bipolar linear transconductance amplifiers. Analog Integr Circ Sig Process
6. Gilbert B (1997) The MICROMIXER: a highly linear variant of the Gilbert mixer using a bisymmetric Class-AB input stage. Solid-State Circ IEEE J 32(9):1412–1423
7. Wang X, Weber R (2003) A novel low-voltage low-power 5.8 GHz CMOS down-conversion mixer design. In: Radio and wireless conference, 2003. RAWCON '03. Proceedings, pp 301–304, 10–13 Aug 2003
8. Kim M-G, An H-W, Kang Y-M, Lee J-Y, Yun T-Y (2012) A low-voltage, low-power, and low-noise UWB mixer using bulk-injection and switched biasing techniques. In: Microwave theory and techniques, IEEE transactions, vol 60, no 8, pp 2486–2493
9. Chuan Chen I, Yang J-R (2010) 2–13GHz broadband CMOS low voltage mixer with active balun designed for UWB systems. In: Electron devices and solid-state circuits (EDSSC), 2010 IEEE international conference, pp. 1–4, 15–17 Dec 2010

10. Gholami S, Shamsi H, Mirzaie H, Mirzahosseini D (2011) Design and analysis of a high-linear UWB CMOS mixer. In: Electrical engineering (ICEE), 2011 19th Iranian conference, pp 1–4, 17–19 May 2011
11. Fashtali ZG, Maghsoodi M, Atani RE, Mahdavi M (2012) A 1-V, 1.3-mW mixer for MB-OFDM UWB receivers by 0.18-μm CMOS technology. In: Telecommunications (IST), 2012 sixth international symposium, pp 338–342, 6–8 Nov 2012
12. Mohamed SAS, Manoli Y (2013) Design of low-power direct-conversion RF front-end with a double balanced current-driven subharmonic mixer in 0.13μm CMOS. Circ Syst I: Reg Pap, IEEE Trans 60(5):1322–1330
13. Tavassoli M, Kargaran E, Tous SI, Nabovati H (2012) Design of 0.5V, 1.28mW CMOS UWB mixer using the body effect. In: Circuits and systems (MWSCAS), 2012 IEEE 55th international midwest symposium, pp 202–205, 5–8 Aug 2012
14. Wu C-Y, Lo Y-K, Chen M-C (2006) A 3.110.6 GHz CMOS direct-conversion receiver for UWB applications. In: Electronics, circuits and systems, 2006. ICECS ‘06. 13th IEEE international conference, pp 1328–1331, 10–13 Dec 2006

Use of Sense Amplifiers for SRAM in Both Conventional Voltage and Charge Transfer Mode

Thakurendra Singh and V. K. Tomar

Abstract This paper focuses on the study and analysis of various sense amplifier circuits used in memory architecture. The sense amplifier plays a very prominent role during read operation in static random access memory. In this work, voltage mode sense amplifier and charge transfer sense amplifier are simulated at 45nm technology node in cadence virtuoso tool. It has been observed that charge transfer sense amplifier gives better response than voltage mode sense amplifiers in respect to power consumption and performance.

Keywords Sense amplifier · Memory · Voltage scaling · SRAM

1 Introduction

In this paper, the design and analysis of the voltage-mode sense amplifiers (VMSA) is discussed [1–3]. These types of amplifiers are designed with CMOS-Memristive shown in Fig. 1. The CMOS, which is based on Von Neumann architecture, uses a huge power. But that would not be an issue till the big data came into the focus. The traditional sense amplifiers are becoming very much speed inefficient and power consuming. So as an alternative, memristors have been introduced, shortening memory registers. Now, the memristor-based sense amplifiers are fast, having a low power consumption, and has an ability to store a huge amount of data in $I - V$ pinched hysteresis [4]. Till now, the memristors are being improved time to time and now four types of memristor devices have been found, i.e., phase-change memristors, resistive switching memristors, spintronics memristors, and ferroelectric tunnel junction memristors. Use of resistor improves the performance as the memristors identify values by resistive reading, rather than capacitive reading [5].

SRAM stands for Static Random-Access Memory that works as a non-volatile memory and stores information till the power is on. The sense amplifier allies the power and operates the SRAM when it requires to read stored data from memory. The

T. Singh · V. K. Tomar (✉)
Department of ECE, GLA University Mathura, Mathura, India
e-mail: vinay.tomar@gla.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_57

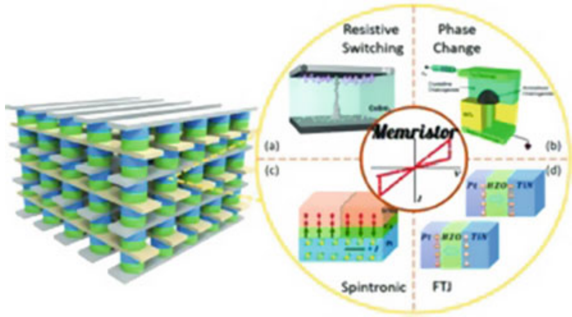


Fig. 1 Memristors

sense amplifier is capable of detecting the very little bitline voltage and amplifying it into the maximum voltage wave before the bitline goes fully charged or discharged. The entire process circuitry is fast and processes as soon as there is a voltage applied, it does not require full charge stage. In the traditional method, it is required to make the circuitry fully charged to determine whether it is in 0 or 1 position. Here, only a small spark voltage is required to determine it's operation state [4].

2 Sense Amplifiers

There are two types of sense amplifiers that can be used for voltage mode sense amplifiers and charge transfer sense amplifiers [6, 7].

2.1 Voltage Mode Sense Amplifier

This type of sense amplifier operates when there is a voltage difference found in the bitline. The main circuit of the amplifier is based on some cross coupled inverters with the help of these inverters; the sense amplifier converts the bitline voltage to a full swing output.

Figure 2 demonstrates that the voltage mode sense amplifier circuit and its implementation. There are two inputs called BLandBLB. Those are coupled in the bitlines column into the cell. There are two inverters $P6N1$ and $P7N2$ in the middle which are used for amplification of differential voltage so that we can get a full wave output. The $P3$, $P4$, and$P5$ inverters are used to precharge the circuit as the amplifier only works when there is power. These inverters precharge the X and $\overline{X}$ nodes through the bitlines. $P1$and$P2$ is connected to the memory via BLandBLB. The inverter $N3$ is used for enabling and disabling the amplifier.

The output inverters isolate the internal nodes of the sense amplifier from the external load. There are two phases of the operation of the voltage mode sense amplifier [8]. At the first operation mode precharge has been done where the P

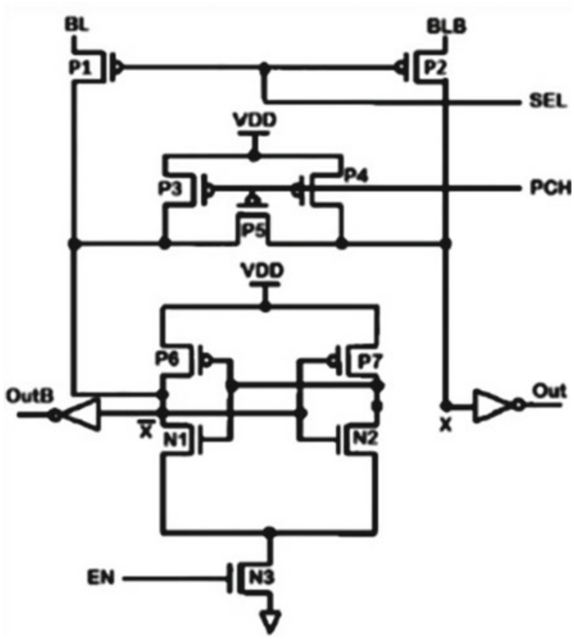


Fig. 2 Voltage mode sense amplifier

channel (PCH) is kept active low. That is how the bitlines and the X and $\overline{X}$ nodes are highly precharged. The second phase is the evaluation phase where the sense amplifier is connected to the memory cell by select-line pulling down. This process is done for the bitlines BLandBLB will develop a voltage difference due to the stored data in the memory cell. The voltage across the bitline BLB will slightly decrease if the data stored in the memory cell is a '1'. Similarly, the voltage across the bitline BL will slightly decrease if the data stored in the memory cell is a '0'. After reaching the required differential voltage between bitlines, EN will increase and the sense amplifier will be activated. The bitline capacitance of voltage mode sense amplifiers will make a differential discharge which is used for sensing the voltage difference and convert it as its inputs to a full swing wave outputs.

2.2 Charge Transfer Sense Amplifier

This type of sense amplifier operates when a charge distribution happens from high bitline capacitance to the low bitline capacitance. Figure 3 discusses a high capacitance to the low capacitance node as X and $\overline{X}$.

The charger distribution process of the charge transfer sense amplifier will be responsible for its high-speed operation. There are two parts in the circuit the first defines the $P6$, $P9$, $N1$and$P7$, $P10$, $N2$ forms a common gate cascade with PMOS.

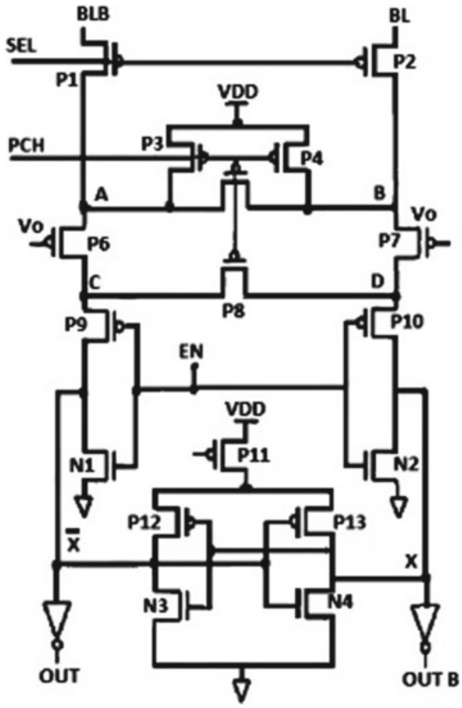


Fig. 3 Charge transfer sense amplifier

The $P6$and$P7$ will be biased at V_0. In the second part, cross coupled inverters will be formed from $P12$, $P13$, $N3$ and$N4$.

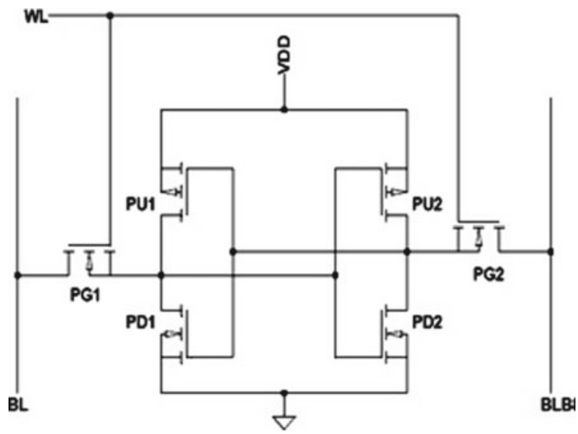
Charge transfer sense amplifier is operated in two phases. The first one is precharge phase where high precharging has been done on the bitlines along with its intermediate nodes A, B, $C$and$D$. The output of the common-gate amplifier is predischarged to low by keeping the EN high.

In the second phase, which is the evaluation phase, P channel (PCH) is kept high and select input (SEL) is grounded to select a column [8–10]. The charge transfer sense amplifier becomes enabled when EN low. The bitline BLB is also going to be low and its voltage goes immediately to $V_0 + |V_{TP}|$. The $P6$ will go into the sub-threshold region of operation to avoid getting charged with the output node. On the other hand, the bitline BL will remain high and charge the X output node to high.

3 Working Principle of 6T SRAM

The 6T SRAM is worked by the previously mentioned sense amplifiers. The 6T SRAMs are the most used SRAMs because they are very thin and longer in size. This feature will allow the SRAM to use longer bitline and reduce the capacitance. For that, a short work line will be enough for a faster performance which also reduces

Fig. 4 6T SRAM



the cost. SRAM consists of 6 transistors as shown in Fig. 4. Out of the 6 transistors, 4 creates two cross coupled inverters.

There are two NMOS transistors called access transistors which are connected as a bridge between cross coupled structure and bitlines BLandBLB. A word line WL operates the transistors.

Now there are three modes of operation on the SRAM which are standby mode, read mode, and write mode. In the first mode which is standby, the SRAM will keep the information stored previously till the power is on. The word line WL is grounded during this mode. In the read mode, first the precharging of the bitline is done to Vdd. After that, the word line WL will make the access transistor turn on. By this process, the access transistor is allowing the bitline to connect to the memory. The logic value of memory shorting will be '0', meaning at low voltage. It makes the pull-down transistor and access transistor to discharge the bitline. For that, a voltage difference will be developed between bitlines. The sense amplifier then senses this differential voltage and shows the stored information in the cell. This voltage difference should be detectable but not extremely high because the high voltage will change the state of the inverter and those will be flipped.

The write operation in the SRAM cell is similar to the reading operation. Here, the bitlines are also being precharged to VDD. Again, the address decoder on the word line WL is enabled by the access transistor. The initial condition of the SRAM cell is stored logic '1' at first node BL and is written with a logic '0'in the other.

4 Simulation and Results

The 6T SRAM cell consists of two inverters that are connected in a cross-coupled manner and formed a core latch. Figure 5 shows the cell layout of 6T SRAM.

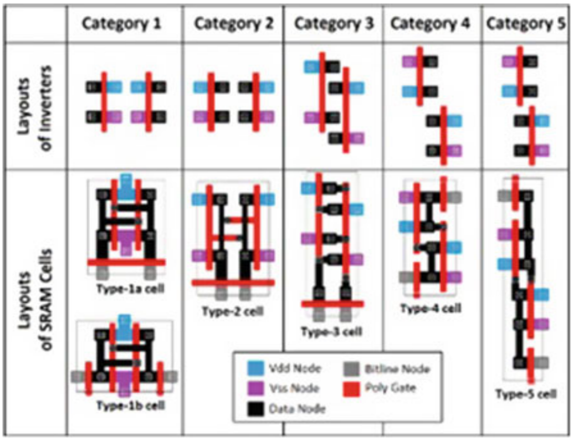


Fig. 5 6T SRAM cell layout topologies

In Fig. 5 (1b) is suitable for profound nano scaling. On the basis of state-of-art, the type 2 has been developed and it is the most popular design till now. But it has a limitation and it won't work above 90 nm technology. To overcome this drwaback, the type 4 has been introduced and has been popularized as it is thinner than previous types. Recently, the type 5 has been introduced which has an ultra-thin design. It has an improved notch less design to make a greater resistivity and lower capacitance. It is just a matter of time to have its acceptance over industries [11, 12].

To determine the output, the amplifiers have been designed into the TANNER EDA simulator. The Figs. 6 and 7 are the simulation diagram of voltage mode sense amplifier and charge transfer sense amplifier respectively. The simulation is based on the industry standard of 65 nm.

The average power consumption has been measured in both the cases and compared with power consumption and sense delay has been renowned between the amplifiers. Tables 1 and 2 shows the power consumption and sense delay comparisons respectively.

Figures 8 and 9 will represent the different voltage output of the voltage mode sense amplifier (VMSA) and charge transfer sense amplifier (CSMA) respectively at 65 nm technology scale.

5 Conclusions

Conclusion of this paper is based on voltage mode sense amplifier and charge transfer sense amplifier along with the operation of 6T SRAM. The operation in terms of power consumption and sense delay is being discussed in brief and the possible outcome has been derived by simulation process. It has been proven that charge transfer sense amplifier gives better response than voltage mode sense amplifiers in respect to power consumption and sense delay.

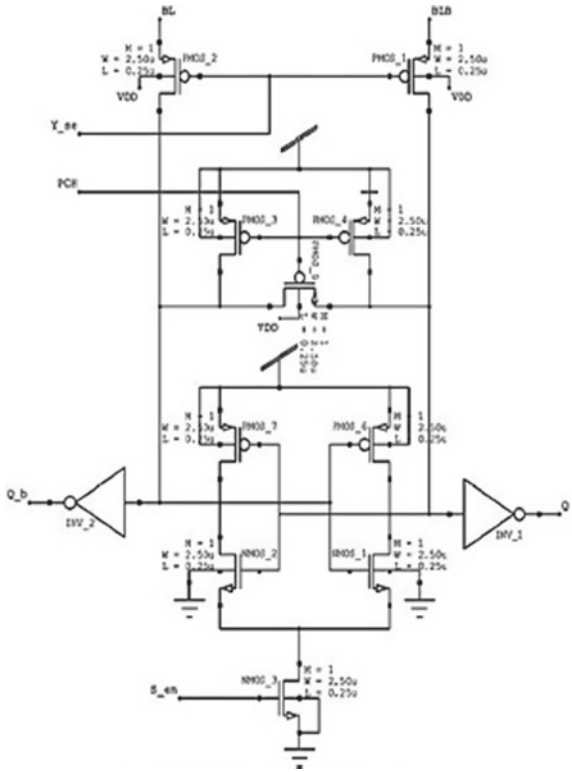


Fig. 6 Schematics of voltage mode sense amplifier

The research work has been based on the current microelectronic industry scenario. The results discuss about different variants of SRAM that have been designed and developed for better architecture layout. This can provide much better output response and work with a higher efficiency under different supply voltages for technical standard. The future research work will be based on the study of effect on sense amplifier and other SRAMs like 8T, 9T etc.

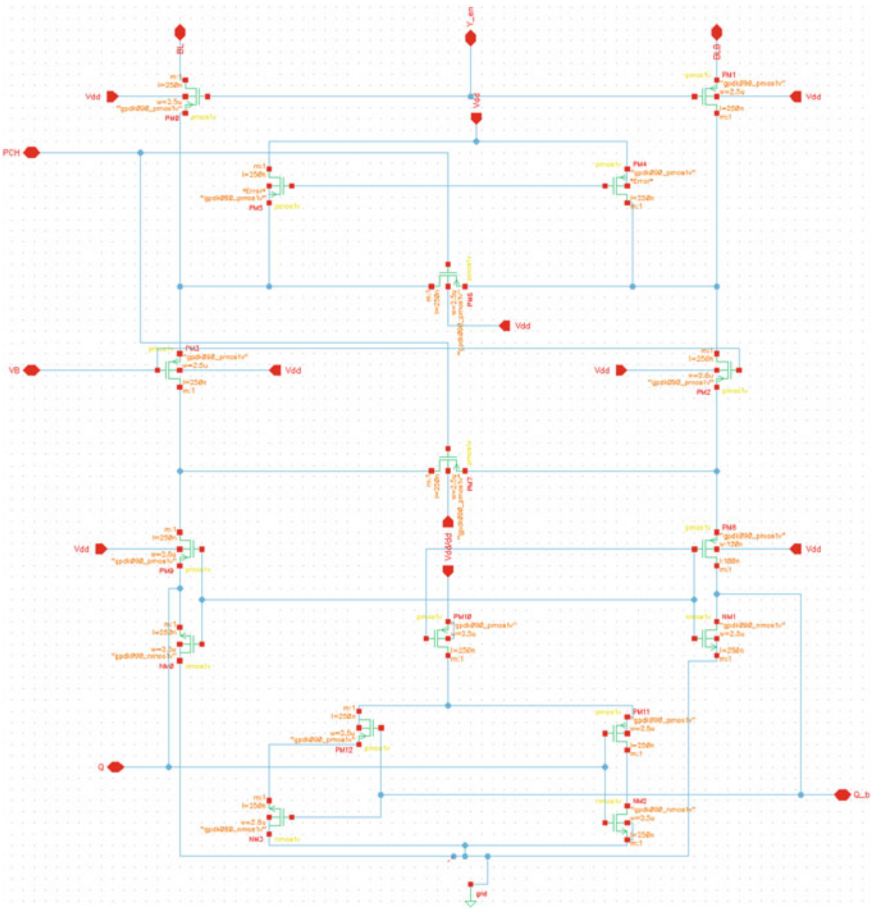


Fig. 7 Schematics of charge transfer sense amplifier

Table 1 Comparison of power consumption with different supply voltage

S No.	Voltage (V) or V_{dd}	Power consumption (μW)	Power consumption (μW)
		VMSA	CSMA
1	1.2	15.65	2.71
2	1.6	29.25	11.26
3	1.9	54.65	29.2
4	2.2	158.45	64.1
5	2.5	492.95	173.7

Table 2. Comparison of sensing delay with different supply voltage

Table head	Voltage (V) or V_{dd}	Sensing delay (ps)	Sensing delay (ps)
		VMSA	CSMA
1	1.2	151.45	75.5
2	1.6	121.6	35.45
3	1.9	105	25.26
4	2.2	96.55	20.16
5	2.5	84.73	18.78

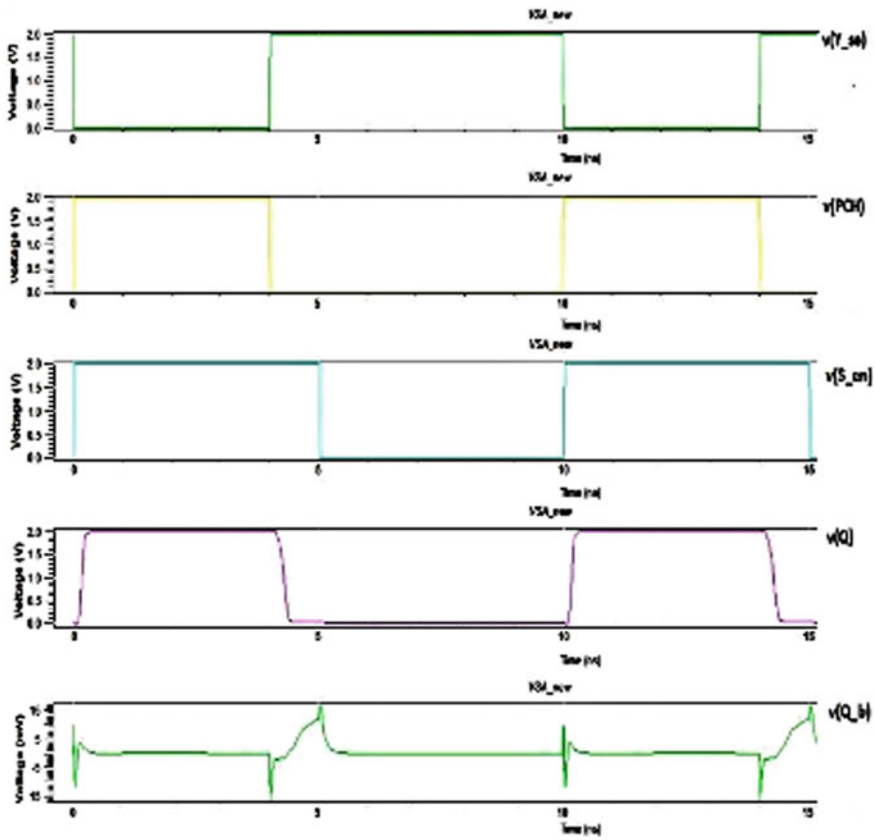


Fig. 8 Waveform of voltage mode sense amplifier

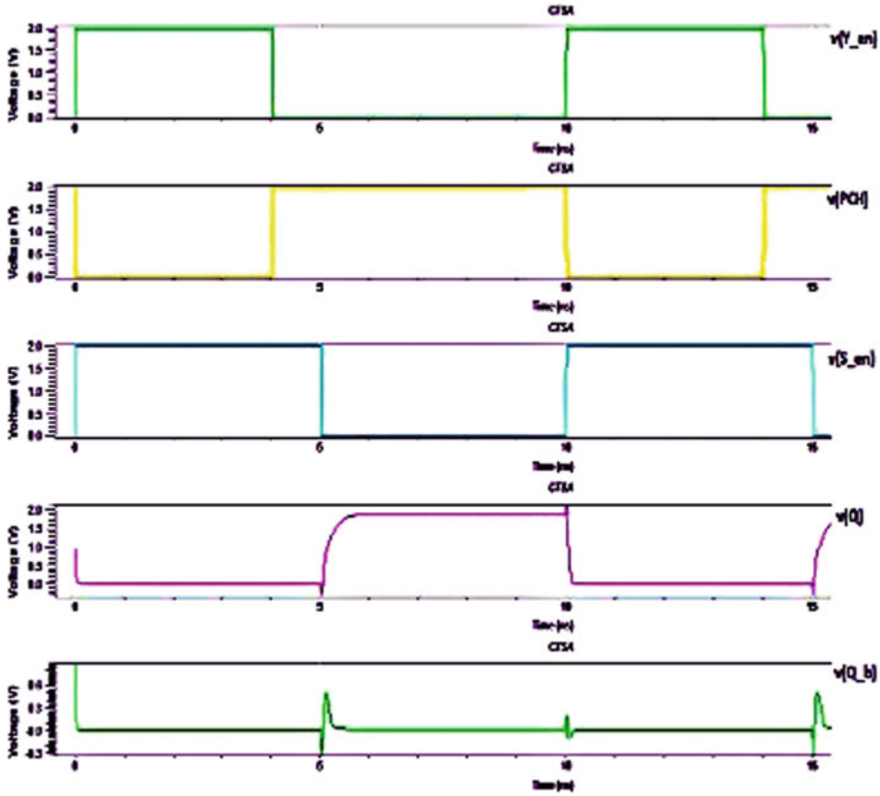


Fig. 9 Charge transfer sense amplifier

References

1. Rani TS, Srinivasulu A, Ravariu C, Appasani B (2021) Low power, high performance PMOS biased sense amplifier. In: 2021 12th international symposium on advanced topics in electrical engineering (ATEE), pp 1–4
2. Mishra AK, Pal S, Vaithiyanathan D (2021) Implementation and analysis of couple suppress current sense amplifier at 45 and 65 nm regime. In: 2021 sixth international conference on wireless communications, signal processing and networking (WiSPNET), pp 421–425
3. Surkar A, Agarwal V (2019) Delay and power analysis of current and voltage sense amplifiers for SRAM at 180 nm technology. In: 2019 3rd international conference on electronics, communication and aerospace technology (ICECA), pp 1371–1376
4. Mittal D (2020) Normal and abnormal spillage suppression methodologies for the SRAM sense amplifiers. In: 2020 IEEE international conference on advent trends in multidisciplinary research and innovation (ICATMRI), pp 1-6
5. Rawat B, Mittal P (2021) A 32 nm single-ended single-port 7T static random access memory for low power utilization. Semicond Sci Technol
6. Sachdeva A, Tomar VK (2021) A multi-bit error upset immune 12T SRAM cell for 5G satellite communications. Wireless Personal Commun 120:2201–2225. https://doi.org/10.1007/s11277-021-08462-8 (SCIE- Impact factor 1.671)) Publisher: Springer ISSN**:** 0929-6212

7. Yadav V, Tomar VK (2021) A low leakage with enhance write margin 10T SRAM cell for IoT applications. In: proceedings of "international conference on micro/nanoelectronics devices, circuits and systems (MNDCS-2021), National Institute of Technology, Silchar, pp 201–211. Print ISBN: 978-981-16-3766-7
8. Kumar H, Tomar VK, Simulation and Analysis of schmitt trigger based 9T SRAM cell with expanded noise margin and low power dissipation. In: Dhar S, Mukhopadhyay SC, Sur SN, C-M Liu (eds) Advances in communication, devices and networking. Springer Singapore, pp 341–347
9. Jain PU, Tomar VK (2020) FinFET technology : as a promising alternatives for conventional MOSFET technology. In: Wagh KS, Sharma N, Kurkute S, Itole DA (eds) IEEE international conference on emerging smart computing and informatics (IEEE–ESCI 2020) will be held on 12–14 March 2020 at AISSMS institute of information technology. IEEE, Pune, pp 43–47. https://doi.org/10.1109/ESCI48226.2020.9167646, ISBN:978-1-7281-5264-6
10. Rastogi A, Tomar VK (2020) Comparative analysis of 6T, 9T and 10T SRAM cells using 18nm FinFet technology. In: Misra R, Kesswani N, Rajarajan M, Bharadwaj V, Patel A (eds) Internet of things and connected technologies , 5th international conference on internet of things and connected technologies (ICIoTCT), at IIT Patna, India, pp 287–296, ISSN 2194-5357 https://doi.org/10.1007/978-3-030-76736-5_27. Springer, Singapore,2021 Print ISBN: 978-3-030-76735-8
11. Krishna H, Tomar VK (2021) Design of low leakage 9T SRAM cell with improved performance for ultra-low power devices. In: Accepted for publication in Journal of circuits, Systems and Computers (2021), https://doi.org/10.1142/S021812662250027X (SCIE- Impact factor 1.337) ISSN No. 0218-1266
12. Sachdeva A, Tomar VK (2021) Characterization of stable 12T SRAM with improved critical charge. J Circ Syst Comput https://doi.org/10.1142/S0218126622500232 (2021) (SCIE-Impact factor 1.337) ISSN No. 0218-1266

Correction to: Multifactor Biometric Authentication for Cloud Computing Security

Dhanshree More, Bhushan Deore, and Surendra Bhosale

Correction to:
Chapter 32 in: V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_32

In the original version of the book, the belated corrections have been incorporated in References 8 and 14 of Chapter 32. The book and the chapter have been updated with this changes.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-19-0976-4_32

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_58

Author Index

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4