

Lecture Notes in Networks and Systems 192

Vishal Goyal
Manish Gupta
Aditya Trivedi
Mohan L. Kolhe *Editors*

Proceedings of International Conference on Communication and Artificial Intelligence

ICCAI 2020

 Springer

Lecture Notes in Networks and Systems

Volume 192

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Advisory Editors

Fernando Gomide, Department of Computer Engineering and Automation—DCA,
School of Electrical and Computer Engineering—FEEC, University of Campinas—
UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering,
Bogazici University, Istanbul, Turkey

Derong Liu, Department of Electrical and Computer Engineering, University
of Illinois at Chicago, Chicago, USA; Institute of Automation, Chinese Academy
of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering,
University of Alberta, Alberta, Canada; Systems Research Institute,
Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering,
KIOS Research Center for Intelligent Systems and Networks, University of Cyprus,
Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong,
Kowloon, Hong Kong

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

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

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

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

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

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

More information about this series at <http://www.springer.com/series/15179>

Vishal Goyal · Manish Gupta · Aditya Trivedi ·
Mohan L. Kolhe
Editors

Proceedings of International Conference on Communication and Artificial Intelligence

ICCAI 2020

 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

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

Mohan L. Kolhe
Faculty of Engineering and Science
University of Agder
Kristiansand, Norway

ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-981-33-6545-2

ISBN 978-981-33-6546-9 (eBook)

<https://doi.org/10.1007/978-981-33-6546-9>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 International Conference on Communication and Artificial Intelligence (ICCAI 2020) was organized to address various issues to flourish the creation of intelligent solutions in 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 state 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 International Conference on Communication and Artificial Intelligence (ICCAI 2020) was organized at GLA University, Mathura, Uttar Pradesh, India, on September 17–18, 2020. 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 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 2020 include eminent academicians, researchers, and practitioners from abroad as well as from all over the nation.

In ICCAI 2020 proceedings, selected manuscripts are subdivided into six tracks named: Communication System, VLSI Design and its Applications, Control System and IoT, Machine Learning, Image and Signal Processing, and Soft and Cloud Computing. ICCAI 2020 received more than 196 submissions. A sincere effort has been made to make it an immense source of knowledge by including 57 manuscripts in this proceedings volume. The selected manuscripts have gone through a rigorous review process and are revised by the 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 Guest, Padma Shri Prof. Sanjay Govind Dhande, and our Guest of honor, Mr. Priyaranjan Kumar, COO, Iconic Fashion, India, for enlightening the participants with their knowledge and insights. We are thankful to our keynote speakers Prof. Nischal Verma from IIT Kanpur, Prof. J. C. Bansal from SAU New Delhi for giving their time and showering the participants with immense knowledge.

We are also thankful for delegates and the authors for their participation and their interest in ICCAI 2020 as a platform to share their ideas and innovative research work. Also, 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. Aditya Trivedi
Prof. Mohan L. Kolhe

Contents

Communication System

A Dual Band Rectangular Patch Notched Antenna with T- and E-Slots	3
Eshita Gupta and Anurag Garg	

All-Optical Priority Encoder Using Semiconductor Optical Amplifier Architecture	11
Komal Damodara, Tarun Kaushik, Sanmukh Kaur, and Nivedita Nair	

Investigation of CRN Spectrum Sensing Techniques: A Scientific Survey	21
Mohak Bansal, Kshitiz Bagora, Smriti Ahuja, and U. Ragavendran	

Frequency Reconfigurable Antenna for Energy Efficient WBAN Application	31
Ridhima Puri, Nimisha Sharma, Asmita Rajawat, and Sindhu Hak Gupta	

Characterization of Millimeter Waves RoFSO Link for 5G Applications	41
M. Vishnu Kartik, Manisha Samal, Sanya Arora, and Sanmukh Kaur	

Characterization of Millimeter Waves RoFSO Link Under the Effect of Rain	51
Manisha Samal, M. Vishnu Kartik, Sanya Arora, and Sanmukh Kaur	

A Proficient Mathematical System (PMS) Design with Improved Cross Layer Scheme for MANET QoS Support	59
Asha	

VLSI Design and Its Application

Concurrent Architecture-Based Fast Fourier Transform on FPGA IP Cores	71
Varun Maheshwari, Mahendra Singh, Pooja Pandit, Reshu Aggarwal, Arun Navputra, Shubham Sharma, and Atul Kumar Shukla	

Analysis of MTCMOS Cache Memory Architecture for Processor	81
Reeya Agrawal and Vishal Goyal	
Optimization of LNA Consisting of CCG Stage and Mutually Coupled CS Stage Using PSO Algorithm for UWB Applications	93
Manish Kumar	
Review: Parametric Variations in Analog-to-Digital Converters Using Different CMOS Technologies	101
Rohini Sharma, Aditya Goswami, and V. K. Tomar	
Power-Efficient Code Converters Using Sub-Threshold Adiabatic Logic Ultra-Low-Power Applications	111
Prashant Gaurav, Prashant, and Sangeeta Singh	
Design Optimization of Doping-less InGaAs TFET and GaAs/Si-Heterojunction Doping-less TFET for Potential Breast Cancer Sensing Applications	123
Shradhya Singh, Navaneet Kumar Singh, Sangeeta Singh, Alok Naughariya, and Neha Niharika	
Impact of Temperature on DC and Analog/RF Performance for DM-DG-Ge Pocket TFET	135
Kumari Nibha Priyadarshani, Sangeeta Singh, and Alok Naugarhiya	
Control System and IoT	
Multidisciplinary Real-Time Model for Smart City Using Internet of Things	145
Ajitesh Kumar and Mona Kumari	
Smart IOT-Enabled Battery Management System for Electric Vehicle	153
Karan Gupta and Vilas H. Gaidhane	
Admittance Measurements for Structural Health Monitoring in Metal Plates	165
T. Jayachitra and Rashmi Priyadarshini	
IoT-Based Smart Wristband	171
Sparidha Kapil, Nikita Pandey, and Pardeep Garg	
Expert System on Smart Irrigation Using Internet of Things	181
Kratika Varshney, Sweta Tripathi, and Vaibhav Purwar	
Design and Analysis of Nonlinear PID Controller for Complex Surge Tank System	189
Jitendra Kumar	

Realistic Assessment of Rural Broadband Requirements: Coverage Aspects 201
 Pranjul Kumar, Ashwani Kumar, and Sanmukh Kaur

Fractional-Order Load Frequency Control of a Two-Area Interconnected Power System with Uncertain Actuator Nonlinearities 211
 Akhilesh Kumar Mishra, Puneet Mishra, and H. D. Mathur

Robust Non-integer Control of a Nonlinear Two-Area Interconnected Power System Subjected to Large Parametric Variations 223
 Akhilesh Kumar Mishra, Puneet Mishra, and H. D. Mathur

Availability and Optimization of Continuous Manufacturing System Using Markov Modelling and Genetic Algorithm 235
 Amit Jain, Vikrant Aggarwal, Rakesh Kumar, and Harmeet Singh Pabla

Indoor and Outdoor Localization Methods for Advanced Navigation Systems 247
 P. Kanakaraja, Sarat K. Kotamraju, L. S. P. Sairam Nadipalli, S. V. Aswin Kumer, and K. Ch. Sri Kavya

Localization Algorithms and Approaches for Navigation in Advanced IoT Applications 259
 P. Kanakaraja, Sarat K. Kotamraju, L. S. P. Sairam Nadipalli, S. V. Aswin Kumer, and K. Ch. Sri Kavya

A LabVIEW-Based Resistor–Capacitor Bank Controller for Automatic Signal Generation with a Multifrequency Voltage-Controlled Oscillator (VCO) 271
 Tushar Kanti Bera

Machine Learning

Detecting Negative Emotions to Counter Depression Using CNN 285
 Pooja Pathak, Himanshu Gangwar, and Aakash Agarwal

Computer Security Based Question Answering System with IR and Google BERT 293
 Pragya Agrawal, Priyanka Askani, Ranjitha Nayak, and H. R. Mamatha

Digit Dataset Generation Using DCGAN: A ResNet Experimentation 303
 Anurag Wani, Aditya Sarwate, Sourav Sahoo, Sahil Sehgal, and Manisha Thakkar

A Novel Proof of Concept for Twitter Analytics Using Popular Hashtags: Experimentation and Evaluation 313
 Kiran Ahire, Manali Bagul, Swapnil Dhanawate, and Suja Sreejith Panicker

Investigating the Role of Machine Learning in Detecting Psychological Tension	323
Suja Sreejith Panicker and P. Gayathri	
Enhanced Movie Reviews Classification Using Precise Combination of PCA with LSTM Model	337
Kriti Bansal, Nancy Bansal, Aman Agrawal, and Bharat Mohan Thakur	
Academic Analytics Through MOOCs	349
Arpana Rawal, Shruti Gupta, Nishant Chhattani, and Jaideep Sachdeva	
Offline Voice Commerce Shopper Tracker Using NLP Analytics	359
Nikhita Mangaonkar, Vivek Venkatesh, and Maulika Arekar	
Intelligent Content-Based Hybrid Cryptography Architecture for Securing File Storage Over Cloud	371
Neeraj Ghule, Aashish Bhongade, Ashish Ranjan, Rahul Bangad, and Balaji Patil	
LegitANN: Neural Network Model with Unbiased Robustness	385
Sandhya Pundhir, Udayan Ghose, and Varsha Kumari	
Analysis of Implication of Artificial Intelligence on Youth	399
Deepika Pandoi	
The Categorization of Artificial Intelligence (AI) Based on the Autonomous Vehicles and Its Other Applications	411
S. V. Aswin Kumer, P. Kanakaraja, L. S. P. Sairam Nadipalli, N. V. K. Ramesh, and Sarat K. Kotamraju	
Context Awareness Computing System—A Survey	423
Pranali G. Chavhan, Gajanan H. Chavhan, Pathan Mohd Shafi, and Parikshit N. Mahalle	
An Adaptive K-Means Segmentation for Detection of Follicles in Polycystic Ovarian Syndrome in Ultrasound Image	431
N. S. Nilofer and R. Ramkumar	
Image and Signal Processing	
Two Layers Machine Learning Architecture for Animal Classification Using HOG and LBP	445
Sandeep Rathor, Shalini Kumari, Rishu Singh, and Pari Gupta	
Machine Learning-Based Detection and Grading of Varieties of Apples and Mangoes	455
Anuja Bhargava and Atul Bansal	
Rule-Based Approach for Emotion Detection for Kannada Text	463
C. P. Chandrika, Jagadish S. Kallimani, H. P. Adarsha, Aparna Nagabhushan, and Appasaheb Chavan	

Gait Recognition Using DWT and DCT Techniques 473
 Shivani and Navdeep Singh

A Novel Image Captioning Model Based on Morphology and Fisher Vectors 483
 Himanshu Sharma

DWT-Based Hand Movement Identification of EMG Signals Using SVM 495
 Vivek Ahlawat, Yogendra Narayan, and Divesh Kumar

A Comparative Analysis on Hyperspectral Imaging-Based Early Drought Stress Detection for Precision Agriculture in Indian Context 507
 Gajanan H. Chavhan, Yogesh H. Dandawate, and Mangesh S. Deshpande

Soft and Cloud Computing

Proposing Perturbation Application Tool for Portable Data Security in Cloud Computing 517
 Amit Kumar Chaturvedi, Meetendra Singh Chahar, and Kalpana Sharma

Horizontal Dynamic Resource Scaling by Measuring the Impacts of Scheduling Interval in Cloud Computing 529
 Amit Kumar Chaturvedi, Praveen Sengar, and Kalpana Sharma

Recapitulation of Research in Artificial Intelligence: A Bibliometric Analysis 539
 Utkal Khandelwal and Trilok Pratap Singh

A Study on Identification of Issues and Challenges Encountered in Software Testing 549
 Omdev Dahiya and Kamna Solanki

A Detailed Survey Study on Classification and Various Attributes of Fake News on Social Media 557
 G. Sajini and Jagadish S. Kallimani

Recognition and Generation of Logically Related Words for Historical Text Data using Reconstruction of Protowords 567
 G. Sajini and Jagadish S. Kallimani

Proposing Host-Based Intruder Detector and Alert System (HIDAS) for Cloud Computing 579
 Amit Kumar Chaturvedi, Punit Kumar, and Kalpana Sharma

Mutative BFO-Based Scheduling Algorithm for Cloud Environment 589
 Saurabh Singhal and Ashish Sharma

Key Management Using Particle Swarm Optimization in MANET 601
Inderpreet Kaur, Parth Pulastiya, and Vivek Anil Pandey

Author Index 609

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 18 years in teaching undergraduate and postgraduate classes. He has published more than 30 research papers publications in different international journals and in proceedings of the 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* and *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 Active Senior Member of IEEE, Life Member of ISTE, Member of IFAC (International Federation of Automatic Control), and Life Member of Indian Science Congress Association.

Dr. Manish Gupta is presently associated with the 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 the year 2006, and Ph.D. from Rajasthan Technical University, Kota

(Rajasthan). He has a very rich experience of more than 18+ years in teaching undergraduate and postgraduate classes. He has published 30+ research papers publications in different international journals and in proceedings of the international conferences of repute. He had served as Member of organizing committees for various international conferences. Moreover, he was also designated the reviewers various refereed journals. He is Life Member of ISTE and Member of IET. His research interest includes communication systems, image and signal processing.

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 MITS, 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 interests 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. Dr. Trivedi 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.

Dr. Mohan L. Kolhe is with the Faculty of Engineering and Science, University of Agder, Norway, as Full Professor in Energy Systems Engineering with focus on Smart Grid and Renewable Energy Systems. He has also received the offer of Hafslund Professorship in Smart Grid from the Norwegian University of Science and Technology (NTNU). Prof. Kolhe has more than twenty-five years' academic experience at international level on electrical and renewable energy systems. He is a leading renewable energy technologist and has previously held academic positions at the world's prestigious universities, e.g. University College London (UK/Australia); University of Dundee (UK); University of Jyvaskyla (Finland); and Hydrogen Research Institute, QC (Canada). Prof. Kolhe was Member of the Government of South Australia's first renewable energy board (2009–2011) and actively contributed to developing renewable energy policies for South Australia. Prof. Kolhe's academic work ranges from the smart grid, grid integration of renewable energy systems, energy storage, electrical vehicles, home energy management system, integrated renewable energy systems for hydrogen production, techno-economics of energy systems, solar and wind energy engineering, and development of business models for distributed generation. Prof. Kolhe has been successful in winning research funding from prestigious research councils (e.g. EU, Norwegian Research Council, EPSRC, BBSRC, Northern Research Partnership Scotland, etc.) for his work on sustainable energy systems. He has published extensively in the area of Energy Systems Engineering. He has

been invited by many international organizations for delivering expert lectures/courses/keynote addresses. He also has been Member of many academic promotion committees as well as Expert Member of international research councils.

Contributors

H. P. Adarsha Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore, India

Aakash Agarwal Department of Computer Application and Engineering, GLA University, Mathura, India

Reshu Aggarwal Electronics & Communication Engineering Department, Faculty of Engineering & Technology, Agra College, Agra, India

Vikrant Aggarwal Production, Planning and Control Manager, MSME Consulting, Ludhiana, Punjab, India

Aman Agrawal Department of Computer Engineering and Application, GLA University, Mathura, India

Pragya Agrawal Department of Computer Science, PES University, Bengaluru, India

Reeya Agrawal GLA University, Mathura, India

Kiran Ahire Department of Computer Engineering, Maharashtra Institute of Technology, Pune, Maharashtra, India

Vivek Ahlawat Department of EE, Sri Ram Group of Colleges, Muzaffarnagar, Uttar Pradesh, India

Smriti Ahuja Electronics & Telecommunication Engineering, SVKM's NMIMS (Deemed to be University), Shirpur, India

Maulika Arekar Sardar Patel Institute of Technology, Mumbai, India

Sanya Arora Amity School of Engineering and Technology, Amity University, Noida, India

Asha Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India

Priyanka Askani Department of Computer Science, PES University, Bengaluru, India

Kshitiz Bagora Electronics & Telecommunication Engineering, SVKM's NMIMS (Deemed to be University), Shirpur, India

Manali Bagul School of Computer Engineering and Technology, MIT World Peace University, Pune, Maharashtra, India

Rahul Bangad Maharashtra Institute of Technology, Pune, Maharashtra, India

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

Kriti Bansal Department of Computer Engineering and Application, GLA University, Mathura, India

Mohak Bansal Electronics & Telecommunication Engineering, SVKM's NMIMS (Deemed to be University), Shirpur, India

Nency Bansal Department of Computer Engineering (Cyber Security), National Institute of Technology, Kurukshetra, India

Tushar Kanti Bera Department of Electrical Engineering, National Institute of Technology Durgapur (NITDgp), Durgapur, West Bengal, India;
Department of Instrumentation and Applied Physics, Indian Institute of Science (IISc), Bangalore, Karnataka, India

Anuja Bhargava Department of Electronics and Communication, GLA, University, Mathura, India

Aashish Bhongade Maharashtra Institute of Technology, Pune, Maharashtra, India

Meetendra Singh Chahar CS Department, Bhagwant University, Ajmer, India

C. P. Chandrika Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore, India

Amit Kumar Chaturvedi Department of MCA, Government Engineering College, Ajmer, India

Appasaheb Chavan Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore, India

Gajanan H. Chavhan Department of Electronics and Telecommunication Engineering, VIT, Pune, Maharashtra, India

Pranali G. Chavhan Department of Computer Engineering, VIIT, Pune, Maharashtra, India

Nishant Chhattani Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, India

Omdev Dahiya University Institute of Engineering and Technology, Maharshi Dayanand University Rohtak, Rohtak, India

Komal Damodara Department of Electronics and Communication Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Yogesh H. Dandawate Department of Electronics and Telecommunication Engineering, VIIT, Pune, Maharashtra, India

Mangesh S. Deshpande Department of Electronics and Telecommunication Engineering, VIT, Pune, Maharashtra, India

Swapnil Dhanawate Department of Computer Engineering, Maharashtra Institute of Technology, Pune, Maharashtra, India

Vilas H. Gaidhane Birla Institute of Technology and Science Pilani, Dubai Campus, UAE

Himanshu Gangwar Department of Computer Application and Engineering, GLA University, Mathura, India

Anurag Garg Engineering College, Ajmer, Rajasthan, India

Pardeep Garg Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India

Prashant Gaurav Microelectronics and VLSI Design, National Institute of Technology Patna, Patna, Bihar, India

P. Gayathri School of Computer Science and Engineering (SCOPE), VIT University, Tamil Nadu, Vellore, India

Udayan Ghose University School of ICT, Guru Govind Singh Indraprastha University, New Delhi, India

Neeraj Ghule Maharashtra Institute of Technology, Pune, Maharashtra, India

Aditya Goswami Department of Electronics and Communication, Engineering, GLA University, Mathura, India

Vishal Goyal GLA University, Mathura, India

Eshita Gupta Engineering College, Ajmer, Rajasthan, India

Karan Gupta Birla Institute of Technology and Science Pilani, Dubai Campus, UAE

Pari Gupta GLA University, Mathura, India

Shruti Gupta Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, India

Sindhu Hak Gupta Amity University, Noida, Uttar Pradesh, India

Amit Jain Department of Computer Science Engineering, GNDEC, Ludhiana, Punjab, India

T. Jayachitra Electrical and Electronics Engineering Department, School of Engineering and Technology, Sharda University, Greater Noida, Uttar Pradesh, India

Jagadish S. Kallimani Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore, India;
Department of Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India

P. Kanakaraja Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Spardha Kapil Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Wakanaghat, Solan, Himachal Pradesh, India

Inderpreet Kaur Galgotias College of Engineering & Technology (GCET), Uttar Pradesh, Greater Noida, India

Sanmukh Kaur Department of Electronics and Communication Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Tarun Kaushik Department of Electronics and Communication Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Utkal Khandelwal GLA University, Mathura, India

Sarat K. Kotamraju Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Ajitesh Kumar GLA University, Mathura, UP, India

Ashwani Kumar Huawei Technologies India Pvt. Ltd, New Delhi, India

Divesh Kumar Department of ECE, GLA University, Mathura, Uttar Pradesh, India

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

Manish Kumar GLA University, Mathura, India

Pranjul Kumar Department of Electronics and Communication Engineering, Amity School of Engineering and Technology, Amity University, Uttar Pradesh, Noida, India

Punit Kumar Department of Computer Science, Bhagwant University, Ajmer, India

Rakesh Kumar Department of Mechanical Engineering, Ludhiana Group of Colleges, Ludhiana, Punjab, India

Mona Kumari GLA University, Mathura, UP, India

Shalini Kumari GLA University, Mathura, India

Varsha Kumari Department of CEA, GLA University, Uttar Pradesh, Mathura, India

S. V. Aswin Kumer Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Parikshit N. Mahalle Department of Computer Engineering, SKNCOE, Wadgaon, Pune, Maharashtra, India

Varun Maheshwari Electronics & Communication Engineering Department, Faculty of Engineering & Technology, Agra College, Agra, India

H. R. Mamatha Department of Computer Science, PES University, Bengaluru, India

Nikhita Mangaonkar Sardar Patel Institute of Technology, Mumbai, India

H. D. Mathur Birla Institute of Technology and Science, Pilani Campus, Pilani, Rajasthan, India

Akhilesh Kumar Mishra Birla Institute of Technology and Science, Pilani Campus, Pilani, Rajasthan, India

Puneet Mishra Birla Institute of Technology and Science, Pilani Campus, Pilani, Rajasthan, India

L. S. P. Sairam Nadipalli Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Aparna Nagabhushan Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore, India

Nivedita Nair Department of Electronics and Communication Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Alok Naughariya National Institute of Technology, Raipur, India

Yogendra Narayan Department of ECE, Chandigarh University, Chandigarh, Punjab, India

Alok Naugarhiya National Institute of Technology Raipur, Raipur, India

Arun Navputra Electronics & Communication Engineering Department, Faculty of Engineering & Technology, Agra College, Agra, India

Ranjitha Nayak Department of Computer Science, PES University, Bengaluru, India

Neha Niharika Microelectronics and VLSI Lab, National Institute of Technology, Patna, India;
Lok Nayak Jai Prakash Institute of Technology Chapra, Chapra, India

N. S. Nilofer Computer Science, Nandha Arts and Science College, Erode, India

Harmeet Singh Pabla Department of Mechanical Engineering, Chandigarh University, Gharuan, Punjab, India

Nikita Pandey Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India

Vivek Anil Pandey Galgotias College of Engineering & Technology (GCET), Uttar Pradesh, Greater Noida, India

Pooja Pandit Electronics & Communication Engineering Department, Faculty of Engineering & Technology, Agra College, Agra, India

Deepika Pandoi Institute of Business Management, GLA University, Mathura, India

Suja Sreejith Panicker School of Computer Engineering and Technology, MIT World Peace University, Pune, Maharashtra, India;
School of Computer Science and Engineering (SCOPE), VIT University, Tamil Nadu, Vellore, India

Pooja Pathak Department of Mathematics, Institute of Applied Sciences, GLA University, Mathura, India

Balaji Patil Dr. Vishwanath Karad, MIT World Peace University, Pune, Maharashtra, India

Prashant Microelectronics and VLSI Design, National Institute of Technology Patna, Patna, Bihar, India

Kumari Nibha Priyadarshani National Institute of Technology Patna, Patna, India

Rashmi Priyadarshini Electronics and Communication Department, School of Engineering and Technology, Sharda University, Greater Noida, Uttar Pradesh, India

Parth Pulastiya Galgotias College of Engineering & Technology (GCET), Uttar Pradesh, Greater Noida, India

Sandhya Pundhir University School of ICT, Guru Govind Singh Indraprastha University, New Delhi, India

Ridhima Puri Amity University, Noida, Uttar Pradesh, India

Vaibhav Purwar Department of Electronics and Communication Engineering, PSIT College of Engineering, Kanpur, U.P, India

U. Ragavendran Electronics & Telecommunication Engineering, SVKM's NMIMS (Deemed to be University), Shirpur, India

Asmita Rajawat Amity University, Noida, Uttar Pradesh, India

N. V. K. Ramesh Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh - 522502, India

R. Ramkumar School of Computer Science, VET Institution of Arts and Science Erode, Erode, India

Ashish Ranjan Maharashtra Institute of Technology, Pune, Maharashtra, India

Sandeep Rathor GLA University, Mathura, India

Arpana Rawal Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, India

Jaideep Sachdeva Department of Computer Science and Engineering, Manipal University, Jaipur, India

Sourav Sahoo Department of Information Technology, MIT World Peace University, Pune, India

G. Sajini Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore, India

Manisha Samal Amity School of Engineering and Technology, Amity University, Noida, India

Aditya Sarwate Department of Information Technology, MIT World Peace University, Pune, India

Sahil Sehgal Department of Information Technology, MIT World Peace University, Pune, India

Praveen Sengar CS Department, Bhagwant University, Ajmer, India

Pathan Mohd Shafi Department of Computer Engineering, SKNCOE, Wadgaon, Pune, Maharashtra, India

Ashish Sharma GLA University, Mathura, India

Himanshu Sharma Department of Computer Engineering and Applications, GLA University Mathura, Mathura, India

Kalpna Sharma Department of Computer Science, Bhagwant University, Ajmer, India;
MCA Department, Govt. Engineering College, Ajmer, India

Nimisha Sharma Amity University, Noida, Uttar Pradesh, India

Rohini Sharma Department of Electronics and Communication, Engineering, GLA University, Mathura, India

Shubham Sharma Electronics & Communication Engineering Department, Faculty of Engineering & Technology, Agra College, Agra, India

Shivani Computer Science and Engineering, Punjabi University, Patiala, Punjab, India

Atul Kumar Shukla Electronics & Communication Engineering Department, Faculty of Engineering & Technology, Agra College, Agra, India

Saurabh Singh GLA University, Mathura, India

Mahendra Singh Vocational Training & Skill Development, Govt. Industrial Training Institute Balkeshwar, Agra, UP, India

Navaneet Kumar Singh Department of ECE, University College of Engineering and Technology (UCET), VBU, Hazaribag, Jharkhand, India

Navdeep Singh Computer Science and Engineering, Punjabi University, Patiala, Punjab, India

Rishu Singh GLA University, Mathura, India

Sangeeta Singh Microelectronics and VLSI Lab, National Institute of Technology, Patna, Bihar, India

Shradhya Singh Microelectronics and VLSI Lab, National Institute of Technology, Patna, India;

Lok Nayak Jai Prakash Institute of Technology Chapra, Chapra, India

Trilok Pratap Singh GLA University, Mathura, India

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

K. Ch. Sri Kavya Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Manisha Thakkar Department of Information Technology, MIT World Peace University, Pune, India

Bharat Mohan Thakur Alten India Private Limited, Bengaluru, India

V. K. Tomar Department of Electronics and Communication, Engineering, GLA University, Mathura, India

Sweta Tripathi Department of Electronics and Communication Engineering, KIT, Kanpur, U.P, India

Kratika Varshney Department of Electronics and Communication Engineering, KIT, Kanpur, U.P, India

Vivek Venkatesh Sardar Patel Institute of Technology, Mumbai, India

M. Vishnu Kartik Amity School of Engineering and Technology, Amity University, Noida, India

Anurag Wani Department of Information Technology, MIT World Peace University, Pune, India

Communication System

A Dual Band Rectangular Patch Notched Antenna with T- and E-Slots



Eshita Gupta and Anurag Garg

Abstract Due to the demand of communication systems increases, here we introduce a dual band antenna with T- and E-shaped slots. This introduced antenna has small radiating rectangular patch with dimension $13 * 16 \text{ mm}^2$. The antenna is resonating at two different frequencies, i.e., 4.8325 and 7.9375 GHz with very substantial gain. At these resonating bands, the antenna has voltage standing wave ratio (VSWR) less than 2 and return loss less than -10 . This type of antenna is suitable for C-band wireless applications. The antenna is simulated by using CST suite to get optimize results.

Keywords Rectangular radiating patch · Slots · Dual band

1 Introduction

From last couple of years, the development of wireless communication has increasing exponentially, and it also increases the demand of antenna because communication is quite impossible without the antenna [1], due to which the demands like ability of multi-function, light weight, and small size from a single device also increase. For these reasons, microstrip patch antenna comes in demand [2, 3]. So, to fulfill these points, many experiments with different geometry of microstrip slot and monopole antennas are observed, and many different methods and strategies are also developed to improve the ability and functionality of the antennas.

In the field of communication, the first band of frequency which was commercially allocated for telecommunication via satellite is 'C-band' [4]. The C-band has frequency range from 4 to 8 GHz and is used in many satellite communication transmissions, in some Wi-Fi, cordless telephones, etc. It is also used in weather and surveillance radar systems. Operating frequency for different devices uses different

E. Gupta (✉) · A. Garg
Engineering College, Ajmer, Rajasthan, India
e-mail: guptaeshita96@gmail.com

A. Garg
e-mail: anurageca@gmail.com

[5] bands of frequency and that is why the demand of imprinted notched microstrip antenna increased because it has several features such as it is small in size, light weighted, cost effective, easy to fabricate, and many more.

Number of techniques has been introduced to obtain dual band antennas [6], they are loading slit technique, loading the patch with sorting pins [7], use slots in the patch, and using stacked patches [6–8]. Here, a new configuration of rectangular microstrip patch antenna is presented in which T- and E-shaped slots are cut inside the radiating patch. This antenna is also known as dual band antenna because it resonates at two different frequencies. The antenna is simulated by using CST Studio Suite 2014 to get the optimized results.

2 Designing of Antenna

The antenna configuration will consist of a ground plane, substrate, rectangular patch containing slots, and microstrip feedline. The rectangular patch is printed on one side of the substrate which is feed by microstrip line of 50Ω and slightly small ground plane on the contrary side of the substrate. The antenna is engraving on Roger RT5880 (lossy) substrate of height 1.6 mm, having loss tangent 0.0009 and permittivity 2.2.

Steps of designing

- (i) Specify some parameters, i.e., dielectric constant (ϵ_r) = 2.2, resonant frequency (f_r) = 7.2412 GHz, height of substrate = 1.6 mm, height of copper ground and patch = 0.035 mm. Now, we have to determine width and length of the patch.
- (ii) To find the practical width, we use the formula which is described in Eq. (1):

$$W = \frac{1}{2fr\sqrt{\mu_0\epsilon_0}} \sqrt{\frac{2}{\epsilon_r + 1}}$$

$$W = 16 \text{ mm} \quad (1)$$

- (iii) Now, we determine the effective dielectric constant (ϵ_{reff}) of the antenna by using Eq. (2)

$$\epsilon_{\text{reff}} = \frac{\epsilon_{r+1}}{2} + \frac{\epsilon_{r+1}}{2} \left(1 + 12 \frac{h}{W} \right)^{-0.5}$$

$$\epsilon_{\text{reff}} \approx 2.173 \quad (2)$$

- (iv) The calculation of effective length (L_{eff}) can be calculated by using Eq. (3)

$$L_{\text{eff}} = \frac{c}{2f_o\sqrt{\epsilon_{\text{reff}}}} \quad (3)$$

where c = speed of light.

$$L_{\text{eff}} \approx 14.1 \text{ mm.}$$

- (v) Total or active length of patch is obtained by Eq. (4)

$$L = L_{\text{eff}} - 2\Delta L \quad (4)$$

$$L \approx 13 \text{ mm}$$

where

$$\Delta L = 0.412h \frac{(\epsilon_{\text{reff}} + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\epsilon_{\text{reff}} - 0.258)\left(\frac{W}{h} + 0.8\right)}$$

$$\Delta L \approx 0.809 \text{ mm}$$

After calculating all the parameters, we will design it in the CST software and then simulate it. After simulating the design, we analyze the results like s-parameter, VSWR, radiation parameter, gain plot, etc.

- (vi) The antenna will resonate at two different frequencies due to the presence of slots on the rectangular radiating patch.

The geometry of antenna, i.e., its front view and back view design, is demonstrated in Figs. 1 and 2. The parameters of antenna were defined in Table 1. The length of the antenna is represented by variable 'X', and width of the antenna is represented by variable 'Y'.

3 Result and Discussion

After simulating, the design which is introduced in this paper will operate/resonates at two different bands which means it can work on these two frequency ranges one at a time or simultaneously which depend upon the ability of the antenna. Dual band antenna is very useful in communication field due to one of its biggest advantage, i.e., it is capable to provide stable and strong wireless connection to the location which is difficult to reach.

So, this antenna is very beneficial because it resonates at two different frequencies, i.e., 4.8325 GHz (4.7945–4.8692 GHz) and 7.9375 GHz (7.8365–8.0319 GHz), and the bandwidth achieved is 74 MHz and 195.4 MHz, respectively, with reference to –10 dB line. The return loss characteristic of the suggested antenna with reference to –10 dB line is shown in Fig. 3.

The voltage standing wave ratio (VSWR) of the antenna is also less than 2 at the resonant bands and is demonstrated in Fig. 4.

Fig. 1 Front view

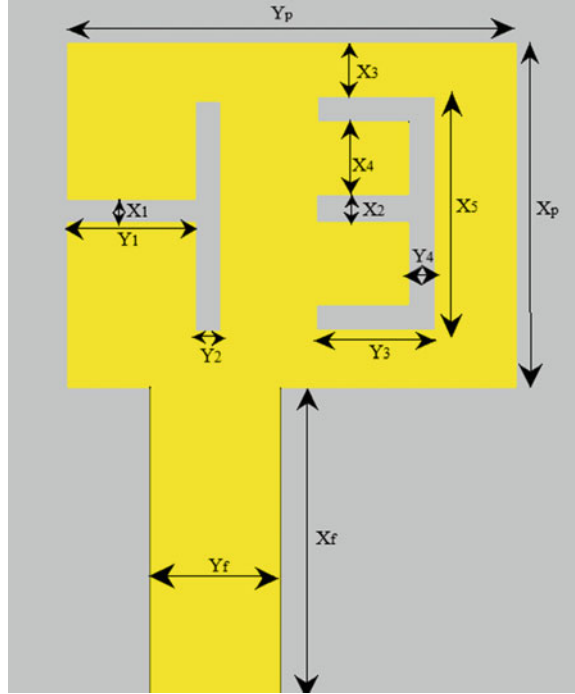


Fig. 2 Back view

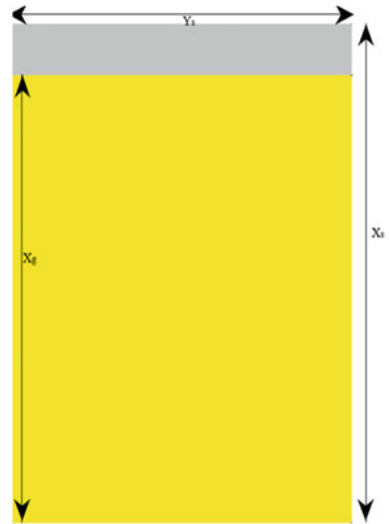


Table 1 Dimensions of antenna

Parameters	Variables	Dimensions (mm)
Patch and slots	X_p	13
	Y_p	16
	X_1	1
	X_2	1
	X_3	2
	X_4	3
	X_5	9
	Y_1	5
	Y_2	1
	Y_3	4
	Y_4	1
Feed line	X_f	10
	Y_f	4.971
Roger RT5880	X_s	20
Substrate	Y_s	24
Ground	X_g	20
	Y_g	22

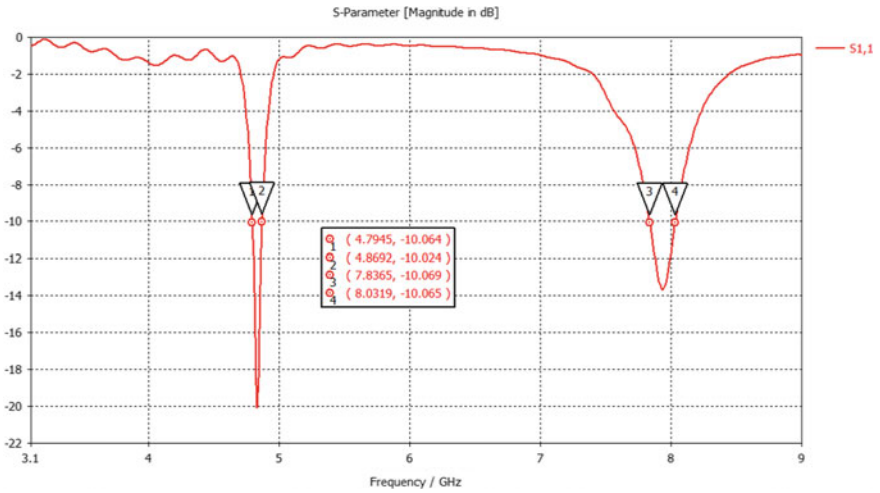


Fig. 3 Return loss plot

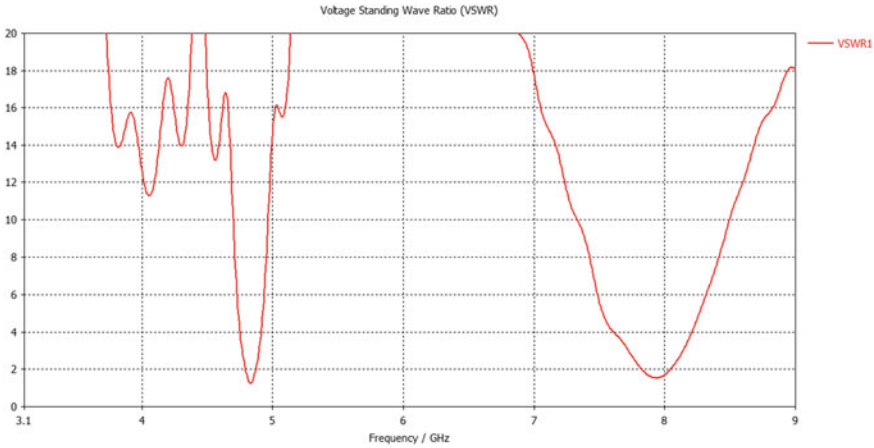


Fig. 4 VSWR plot

The antenna is resonating at two different bands, first at 4.8325 GHz with directivity 6.161 dBi and second at 7.9375 GHz with directivity 6.503 dBi. The radiation pattern for these two resonating bands is highly directional, i.e., the main lobe is in one direction only and the directivity at both resonating bands is high.

Figure 5a, b signifies the radiation pattern at two different resonant frequencies in 2-dimension as well as in 3-dimension.

Gain plot of the antenna is demonstrated in Fig. 6. The gain at resonant frequency 4.8325 GHz is approximately 5.2506 dB, and at resonant frequency 7.9375 GHz, the gain is approximately 5.6871 dB. This antenna has substantial gain.

4 Conclusion

The introduced antenna is a dual band antenna because it resonates at two different frequencies, i.e., 4.8325 GHz and 7.9375 GHz, with reference to -10 dB line and VSWR less than 2. The gain of the antenna is also high. At frequency 4.8325 GHz and 7.9375 GHz, the gain is 5.2506 dB and 5.6879 dB, respectively. The radiation pattern at these frequencies is directional, and this antenna is suitable for many C-band wireless applications.

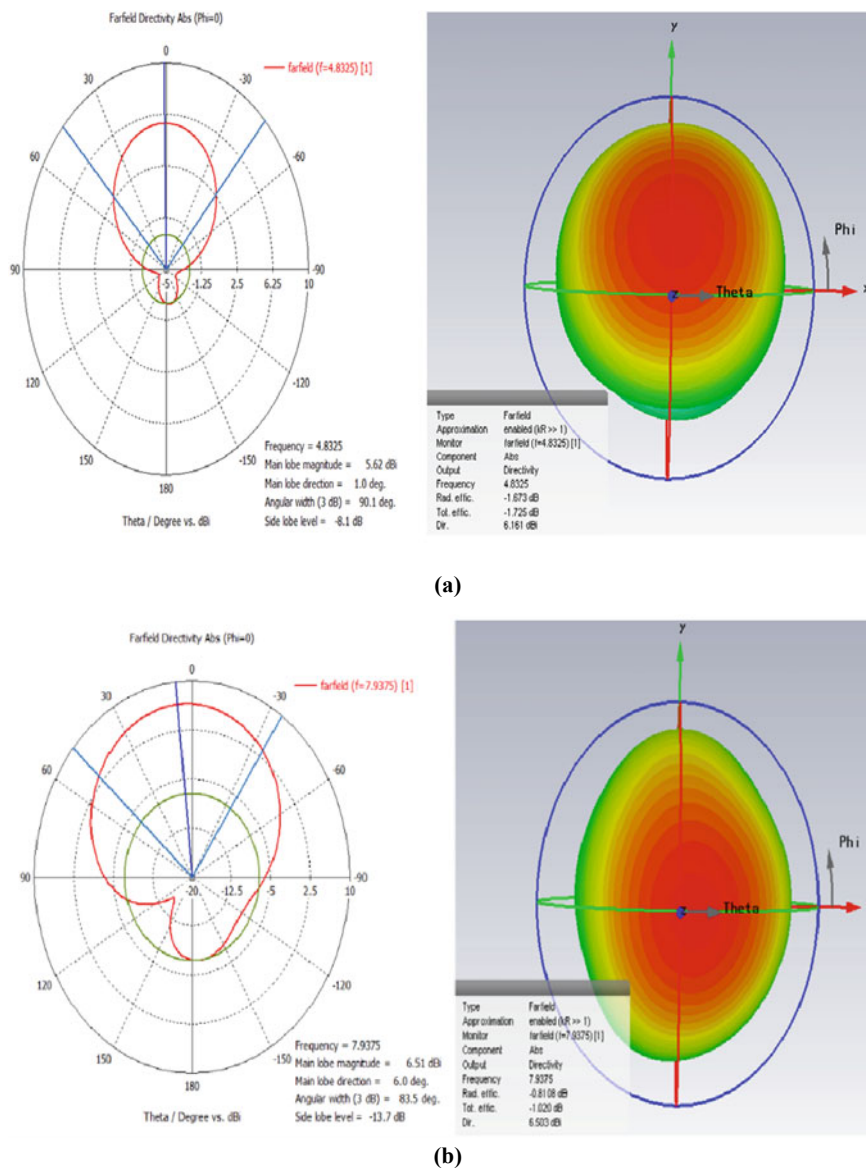


Fig. 5 Radiation pattern in 2D and in 3D at resonant frequency **a** at 4.8325 GHz and **b** at 7.9375 GHz

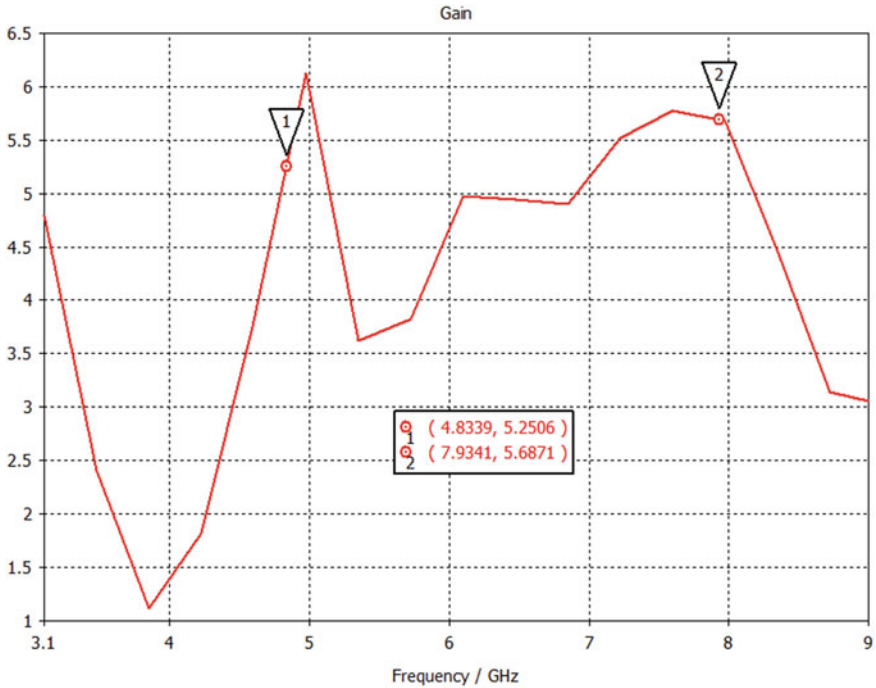


Fig. 6 Gain plot

References

1. Khan S, Daiya V, Ebenezer J, Jehadeesan R (2019) Novel patch antenna design for wireless channel monitors. In: 2019 10th international conference on computing, communication and networking technologies (ICCCNT). IEEE, New York, pp 1–6
2. Garg A, Kumar D, Dhaker PK, Sharma IB (2015) A novel design dual band-notch small square monopole antenna with enhanced bandwidth for UWB application. In: 2015 international conference on computer, communication and control (IC4). IEEE, New York, pp 1–5
3. Deshmukh AA, Singh D, Ray KP (2019) Modified designs of broadband E-shape microstrip antennas. *Sādhanā* 44(3):64
4. Singh A, Rathore K, Sharma P, Raj RK (2014) Dual band notched small square monopole UWB antenna with enhanced bandwidth. In: 2014 international conference on computational intelligence and communication networks. IEEE, New York, pp 64–68
5. Mehranpour M, Nourinia J, Ghobadi Ch, Ojaroudi M (2012) Dual band-notched square monopole antenna for ultrawideband applications. *IEEE Antennas Wireless Propagat Lett* 11:172–175
6. Yan B, Wang L, Luo Z, Deng D, Feng L, Zheng H (2016) Dual-band microstrip antenna fed by coaxial probe. In: 2016 11th international symposium on antennas, propagation and EM theory (ISAPE). IEEE, New York, pp 228–230
7. Samsuzzaman Md, Islam MT, Faruque MRI, Hueyshin W (2013) Dual wideband n shaped patch antenna loaded with shorting pin for wireless applications. In: 2013 2nd international conference on advances in electrical engineering (ICAEE). IEEE, New York, pp 273–276
8. Anitha P, Reddy ASR, Giri Prasad MN (2018) Design of a compact dual band patch antenna with enhanced bandwidth on modified ground plane. *Int J Appl Eng Res* 13(1):118–122

All-Optical Priority Encoder Using Semiconductor Optical Amplifier Architecture



Komal Damodara, Tarun Kaushik, Sanmukh Kaur, and Nivedita Nair

Abstract The digital systems or networks using photon-integrated circuits offer high-speed and energy-efficient functionalities. The optical digital devices are an integrated part of advanced signal processing which indeed is an essential requirement of next-generation optical networks and computing. A priority encoder accepts multiple input signals at the same time and after identifying the priority of the requested inputs encodes the priority of the input signals accordingly. This paper demonstrates the design and operation of a 4:2 priority encoder employing semiconductor optical amplifier-based Mach–Zehnder interferometers. The input data signals to the priority encoder have been applied at same wavelengths. The three encoded signals also appear at the output of priority encoder at the same wavelength. Simulation timing diagrams of the proposed logic device verifying the correct logic bit patterns have been obtained at 10 Gb/s.

Keywords All-optical gates · SOA–MZI · Semiconductor optical amplifier · Priority encoder

K. Damodara (✉) · T. Kaushik · S. Kaur · N. Nair
Department of Electronics and Communication Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India
e-mail: komaldamodara@gmail.com

T. Kaushik
e-mail: tkkaushik71998@gmail.com

S. Kaur
e-mail: skaur2@amity.edu

N. Nair
e-mail: nnair@amity.edu

1 Introduction

There is an ever-growing demand of data communication in communication networks. For this, high-speed digital signal processing is required to perform different computational functionalities which include header processing, switching, encryption, swapping, labeling, retiming, and reshaping [1].

Optical computing enables the realization of ultrawide band or ultrahigh speed information processing [2]. An all-optical parity encoder has been one of the important core components of optical computing systems and networks.

Ultrahigh speed logic operations realized using all-optical gates utilizing nonlinear properties of semiconductor optical amplifiers (SOA) have been reported in [3]. Many other complex optical logic devices and circuits have been demonstrated using different types of nonlinear devices [4–10]. A 1- and 2-bit magnitude comparator has been presented employing electro-optical effect in Mach–Zehnder interferometers (MZI) [4]. An all-optical data comparator and decoder have been presented in [5] using SOA-MZI based optical tree architecture. Optical arithmetic circuits for full addition [6] and subtraction [7] have also been demonstrated using logic gates employing nonlinear properties of SOA. All-optical parity generator and checker circuits utilizing electro-optical effect in MZIs and SOA based Mach–Zehnder interferometers have been proposed in [8] and [9], respectively.

An encoder circuit in electronic domain has been widely employed in performing multiplexing, demultiplexing, binary calculations, data encryption, and address recognition functions. A priority encoder accepts multiple input signals at the same time and after identifying the priority of the requested inputs encodes the priority of the input signals accordingly. An all-optical priority encoder has been demonstrated using semiconductor optical amplifiers which have been cascaded with detuned band pass filters in [10]. Photonic crystal-based ring resonators have been employed to design an all-optical four to two optical encoder in [11].

All-optical circuits and devices employing SOAs have attracted tremendous attention as a result of its high capability of integration with broad selection of active or passive devices or components, low power consumption, and stable and high speed of operation [9].

In the present work, we propose an all-optical 4:2 priority encoder using SOA-MZI based logic gates. The input data signals to the priority encoder have been applied at same wavelengths. The three encoded signals also appear at the output of priority encoder at the same wavelength. Simulation timing diagrams of the proposed logic device verifying the correct logic bit patterns have been obtained at 10 Gb/s.

2 System Design

Priority encoder works on the principle that if two or more inputs are at higher value at the same time, then the input with the highest priority takes the precedence. There

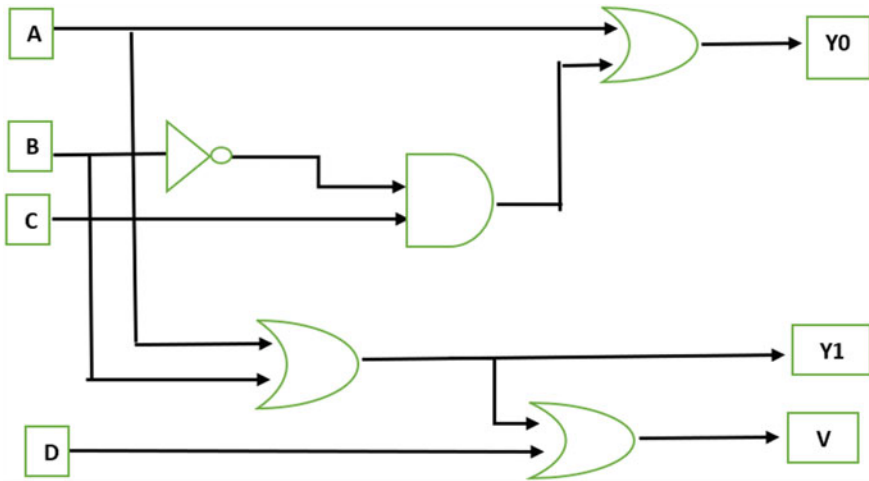


Fig. 1 Circuit schematic for priority encoder

are basically four input signals (*A*, *B*, *C*, and *D*) with two outputs Y_0 , Y_1 . The priority encoder circuit is shown in Fig. 1. The architecture has three OR gates, one AND gate, and one NOT gate being used.

The logic equations for the priority encoder are expressed as below

$$Y_1 = A + B \tag{1}$$

$$Y_0 = A + \overline{B}.C \tag{2}$$

$$V = A + B + C + D \tag{3}$$

The truth table is shown as Table 1.

The proposed architecture of 4:2 priority encoder utilizing different optical logic gates has been depicted in Fig. 2. All the gates have been realized employing SOA based MZI switches taking two identical SOAs and placing them in the path between the two input and output couplers. There are four laser diodes with wavelengths of

Table 1 Truth table

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	Y_0	Y_1
0	0	0	0	0	0
0	0	0	1	0	1
0	1	X	X	1	0
1	X	X	X	1	1

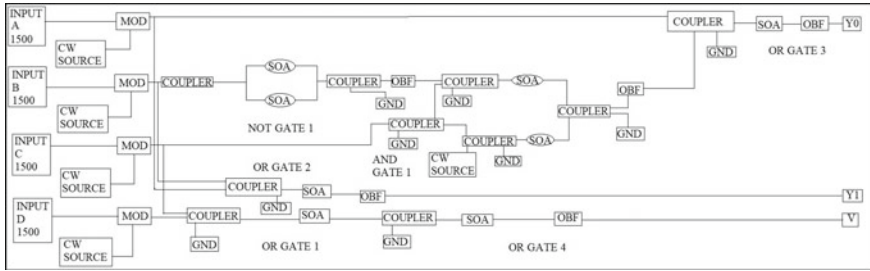


Fig. 2 Schematic diagram of 4-bit priority encoder with all-optical logic gates

1500 nm (λ_1) each and four continuous wave (CW) laser beams with a wavelength of 1550 nm (λ_2) applied as inputs to all-optical gates.

Input data signal B with a 1500 nm wavelength along with the CW beam source with a wavelength 1550 nm has been applied as inputs to the NOT gate 1, designed with the help of SOA–MZI architecture. The output of the NOT gate 1 obtained is passed through Gaussian optical band pass filter (OBF) with a wavelength of 1500 nm and with a bandwidth of 0.4 nm. The output of OBF is taken along with input data signal C with a wavelength 1500 nm, and a CW source with a wavelength 1550 nm has been applied as input signals to the AND gate 1. The output of AND gate 1 $\overline{B}.C$ so obtained has been passed through the gaussian OBF filter with a bandwidth 0.4 nm and a wavelength of 1500 nm. The output $\overline{B}.C$ along with the input data signal A with a wavelength of 1500 nm and a continuous wave beam source with a wavelength 1550 nm have been given as inputs to OR gate 1. The output of OR gate 1, $A + \overline{B}.C$ is further passed through the Gaussian OBF of bandwidth 0.4 nm and wavelength 1500 nm to obtain output Y_0 signal with a 1500 nm wavelength.

A CW source with a wavelength of 1550 nm along with an input data signal A with a wavelength 1500 nm has been provided as input signals to the coupler of OR gate 2. Likewise, an input data signal B along with CW source of wavelength 1550 nm has been applied as inputs to the coupler of OR gate 2. The output of the OR gate 2 so obtained is passed through the Gaussian OBF with a bandwidth 0.4 nm and wavelength of 1500 nm. The output obtained from OR gate 2 at terminal Y_1 is $A + B$ signal with a wavelength of 1500 nm.

Input data signal C at a wavelength of 1500 nm and a CW beam with 1550 nm wavelength have been given as inputs to the coupler of OR gate 3. Likewise, a CW source of wavelength 1550 nm and an input data signal D at 1500 nm wavelength have been applied as input signals to the coupler of OR gate 3. The output obtained from OR gate 3 of 1500 nm wavelength along with OR gate 2 output obtained at 1500 nm wavelength are taken as the input signals to the OR gate 4. OR gate 4 output has been filtered with the help of Gaussian OBF with a bandwidth of 0.4 nm and a wavelength of 1500 nm. The output obtained at a wavelength of 1500 nm from OR gate 4 at terminal V represents $A + B + C + D$ signal.

3 Results

RZ modulated signals have been used to obtain the simulated output waveforms at a data speed of 10 Gb/s. The efficiency or output power level of the proposed all-optical logic device may show some deviation as a result of the change in the states of polarization of the input signals, but the results of priority encoder are achieved without any loss in the performance of the device.

Figure 3 depicts the simulation timing diagrams of 4-bit priority encoder. Two input data signals (A and B) of peak power 1 dBm and a wavelength of 1500 nm,

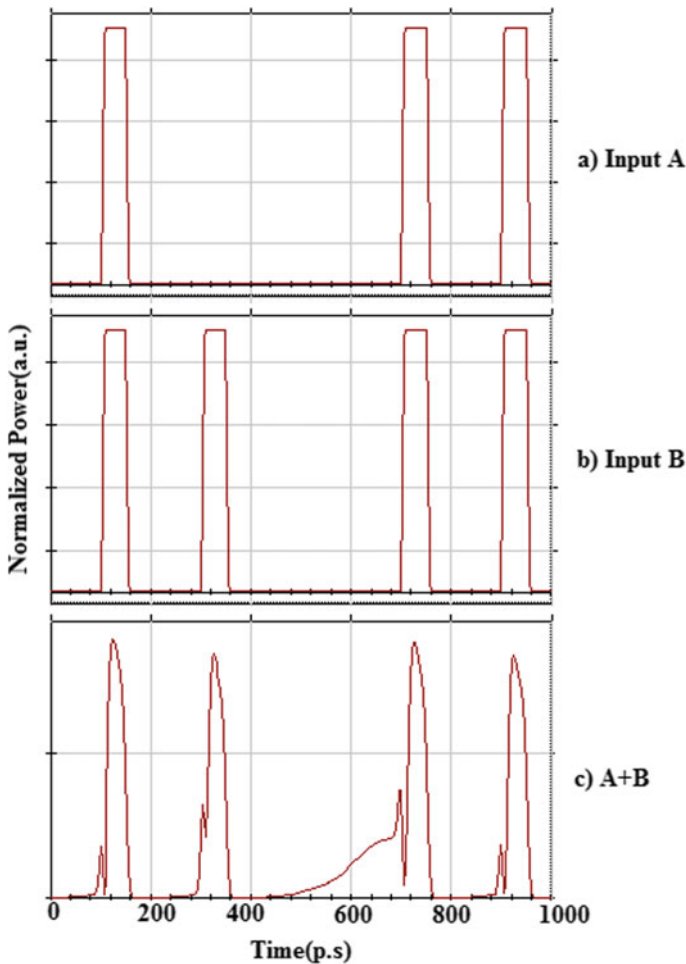


Fig. 3 Output timing diagram of Y_1 output representing $A + B$ signal. **a** Data signal A , **b** Data signal B , **c** Output $A + B$

along with CW beams with a wavelength of 1550 nm and peak power of 0 dBm have been applied as inputs to OR gate 2.

The input data signal A and input data signal B are depicted in Fig. 3a, b, respectively. Y_1 output representing $A + B$ signal with a wavelength of 1500 nm has been depicted in Fig. 3c.

Simulation timing diagrams of 4-bit priority encoder for Y_0 output representing $A + \overline{B}.C$ signal are depicted in Fig. 4. A CW beam of 1550 nm wavelength and 0 dBm peak power and a data input signal B having a peak power of 1 dBm and 1500 nm wavelength have been given as inputs to the NOT gate1. NOT gate1 output has been achieved at 1500 nm wavelength. The NOT gate output signal \overline{B} and input signal C with a wavelength of 1500 nm along with a continuous wave beam of wavelength 1550 nm have been applied as inputs to the AND gate1. Output of the gate is passed through an optical band pass filter and is observed at 1500 nm wavelength. The input data signal A with a wavelength of 1550 nm and a CW beam of wavelength 1500 nm along with the output of AND gate1, i.e., $\overline{B}.C$, has been taken as the input signal to the OR gate 1. The input data signal A, B, and C is depicted in Fig. 4a–c, respectively. Figure 4d depicts the output of NOT gate1, i.e. \overline{B} . Figure 4e depicts the AND gate output signal, i.e., $\overline{B}.C$. Figure 4f depicts OR gate1 output signal, i.e., Y_0 output representing $A + \overline{B}.C$.

The simulation results of 4-bit priority encoder for $A + B + C + D$ have been depicted in Fig. 5. The input signal C and D with a peak power of 1 dBm and 1500 nm wavelength along with a CW beam of peak power of 0 dBm and a wavelength of 1550 nm are taken as the inputs for OR gate 4. OR gate 2 and OR gate 3 outputs are taken as the input signals for OR gate 4. OR gate 4 output, i.e., $A + B + C + D$ signal, has been obtained at a 1500 nm wavelength. Figure 5a–d depicts the input data signal A, B, C, and D, respectively. Figure 5e depicts the OR gate 2 output signal, i.e., $A + B$. Figure 5f depicts the OR gate 3 output, i.e., $C + D$. Figure 5g depicts the output signal of OR gate 4, i.e., V output representing $A + B + C + D$.

4 Conclusion

In this work, we propose an all-optical 4:2 priority encoder using SOA–MZI based logic gates. We have designed and simulated the device using optisystem software. Four input data signals to the priority encoder have been applied at same wavelengths. The three encoded signals also appear at the output of the device at the same wavelength. Simulation timing diagrams of the proposed logic device verify its operation at 10 Gb/s. The device exhibits the advantages of power efficiency, integration potential, and high operational speed.

Fig. 4 Output timing diagram of Y_0 output representing $A + \overline{B.C}$ signal. **a** Data signal A, **b** Data signal B, **c** Data signal C, **d** Signal \overline{B} , **e** Signal $\overline{B.C}$, **f** Signal $A + \overline{B.C}$

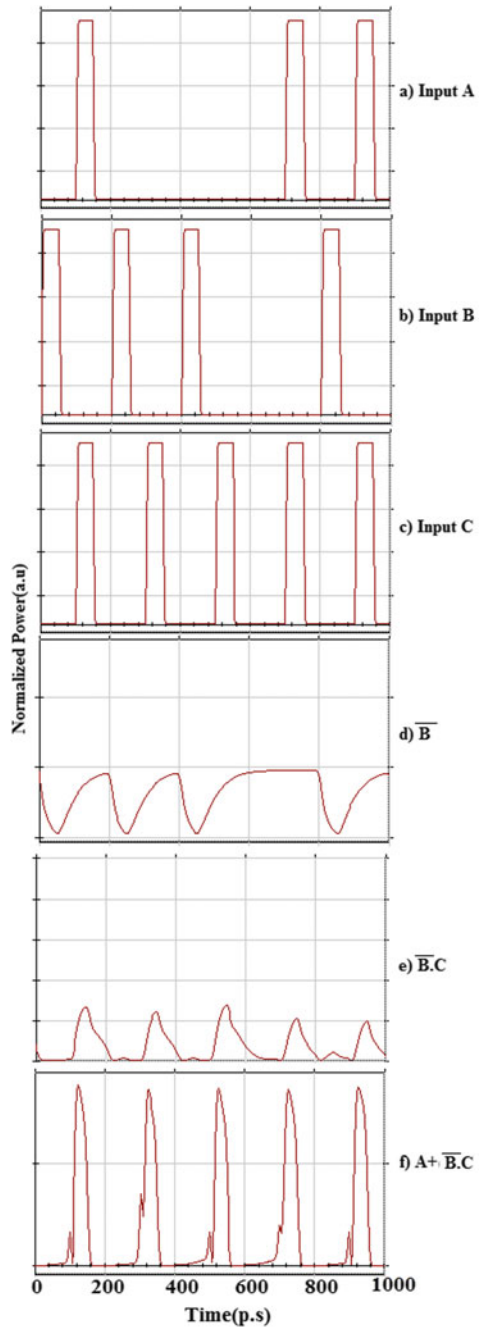
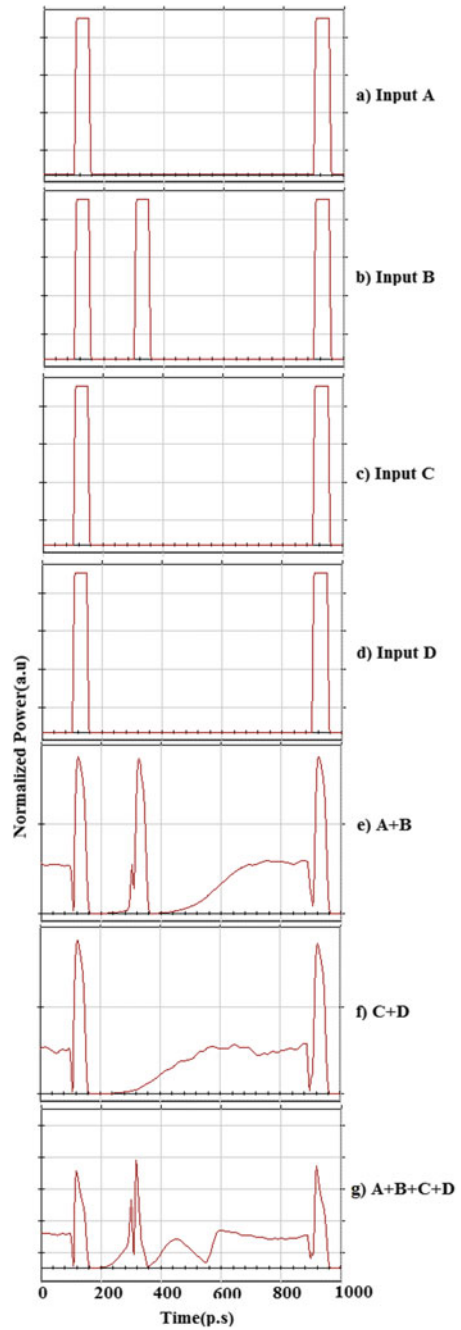


Fig. 5 Output timing diagram of V output representing $A + B + C + D$ signal. **a** Data signal A , **b** Data signal B , **c** Data signal C , **d** Data signal D , **e** Signal $A + B$, **f** Signal $C + D$, **g** Signal $A + B + C + D$



References

1. Kaur S, Shukla MK (2017) All-optical parity generator and checker circuit employing semiconductor optical amplifier-based Mach–Zehnder interferometers. *Optica Applicata* 47(2):263–271. <https://doi.org/10.5277/oa170209>
2. Anurupa SK, Malhotra Y (2019) Performance evaluation and comparative study of novel high and flat gain C+L band Raman+EYDFA co-doped fibre hybrid optical amplifier with EYDFA only amplifier for 100 channels SD-WDM systems. *Opt Fiber Technol* 53:102016. <https://doi.org/10.1016/j.yofte.2019.102016>
3. Kaur S, Kaler RS (2012) Ultrahigh speed reconfigurable logic operations based on single semiconductor optical amplifier. *J Opt Soc Korea* 16(1):13–16. <https://doi.org/10.3807/JOSK.2012.16.1.013>
4. Kumar S, Bisht A, Singh G, Choudhary K, Raina KK, Amphawan A (2015) Design of 1-bit and 2-bit magnitude comparator using electro-optic effect in Mach-Zehnder interferometers. *Opt Commun* 357(15):127–147
5. Kaur S (2013) All optical data comparator and decoder using SOA-based Mach Zehnder Interferometer. *Optik* 124(17):2650–2653. <https://doi.org/10.1016/j.ijleo.2012.07.041>
6. Kaur S, Kaler RS, Kamal T-S (2015) All-optical binary full adder using logic operations based on the nonlinear properties of a semiconductor optical amplifier. *J Opt Soc Korea* 19(3):222–227. <https://doi.org/10.3807/JOSK.2015.19.3.222>
7. Kaur S (2016) All-optical binary full subtracter using logic operations based on nonlinear properties of semiconductor optical amplifier. *J Nonlinear Opt Phys Mater* 25(1):1650003 (9 p)
8. Kumar S, Chandrakanta AA (2016) Design of parity generator and checker circuit using electro-optic effect of Mach–Zehnder interferometers. *Opt Commun* 364:195–224
9. Nair N, Kaur S, Goyal R (2018) All-optical integrated parity generator and checker using an SOA-based optical tree architecture. *Curr Opt Photon* 2(5):400–406. <https://doi.org/10.3807/COPP.2018.2.5.400>
10. Lei L, Dong J, Zhang Y et al (2012) 40-Gb/s all-optical digital 4-bit priority encoder employing cross-gain modulation in semiconductor optical amplifiers. *Chin Sci Bull* 57(10):1204–1208. <https://doi.org/10.1007/s11434-011-4957-2>
11. Seif-Dargahi H (2018) Ultra-fast all-optical encoder using photonic crystal-based ring resonators. *Photon Netw Commun* 36(2):272–277. <https://doi.org/10.1007/s11107-018-0779-3>

Investigation of CRN Spectrum Sensing Techniques: A Scientific Survey



Mohak Bansal, Kshitiz Bagora, Smriti Ahuja, and U. Ragavendran

Abstract The wireless remote traffic is growing in an unmatched strategy, which is the origin of the frequency spectrum shortage. A few overviews of the spectrum utilization show that the total range extends is not utilized reliably, such a significant number of the radio spectra are underutilized. A segment of the frequency bands in the range is empty, a part of the frequency bands less involved, and very few bands are overutilized. The cognitive radio system is a strategy that beats that spectrum underutilization. In cognitive radio, the secondary user scans for a free band to use when the primary user is not in the use of its approved band. A segment of cognitive radio is called spectrum detecting, which engages to filter for the free bands, and it helps with acknowledging the spectrum hole, which can be used by the secondary user. Detection of the empty spectrum is the first step toward a cognitive radio network (CRN). A productive and fast spectrum detecting makes cognitive radio more efficient. We studied a couple of range recognizing techniques used in cognitive radio. The spectrum which is empty is detected by the secondary users; therefore, some spectrum detecting methods are used. Spectrum detecting is a major part of cognitive radio networking that allows us to use the vacant frequency band. In this paper, we break down the cooperative spectrum detecting techniques in CRN.

Keywords Ad-hoc networks · Cognitive intelligent system · Future networking · Multiband–multiuser system · Spectrum utilization

1 Introduction

The increasing demand for wireless communicate structures has caused seeking of suitable spectrum bands for transmission of records. Because of the continuous increase in the data traffic, the scarcity of spectrum is becoming a barrier within the wireless networks. This research discovered out that the cognitive radio spectrum is used in most of the context. Cognitive radio is a kind of wireless communication

M. Bansal (✉) · K. Bagora · S. Ahuja · U. Ragavendran
Electronics & Telecommunication Engineering, SVKM's NMIMS (Deemed to be University),
Shirpur, India
e-mail: mohakbansal1000@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 192,
https://doi.org/10.1007/978-981-33-6546-9_3

where transceivers can consequently identify the accessible and empty ranges. The cognitive radios exhibit three assortments of practices which incorporates: (i) Maintaining a strategic distance from conduct where the secondary user connects with the essential range without meddling with the primary user giving rise to interference in the network, (ii) Interference controlling [1] conduct where the secondary users transmit over the equivalent range because the primary users, however, accomplish this in a way that the interference visible through the primary clients from the cognitive users is controlled to an admissible degree. (iii) Interference mitigating behavior [2, 3]: The secondary clients transmit over the indistinguishable spectrum as the licensed user, however, similar to the expertise of the channels among authorized and unlicensed users. Spectrum sensing techniques can be used to broadcast on the unused spectrum. It overcomes radio spectrum scarcity. It also avoids network jamming problems [4, 5]. CR also saves power while the transfer of data from the transmitter to the receiver. CR further improves communication and also increases the quality of service. Successful transmission of data from the transmitter to the receiver is done only because of spectrum sensing. This data transfer is one of the major importance of sensing the spectrum. Spectrum sensing permits the user to determine empty bands in the communication networks to transmit the data from that empty band. The cognitive radio network topology is determined by the ability of spectrum sensing. But in conjunction with all this, there are a few troubles in the cognitive radio spectrum such as: (i) Cognitive radio networks provide demanding situations due to fluctuations in available spectrums. Spectrum control competencies can address the ones traumatic situations for the betterment of this new community paradigm. (ii) Opportunistic verbal exchange with interference avoidance [3, 6] among the licensed and unlicensed users faces a big quantity of demanding situations inside the detection because of the presence of user precedence. (iii) Information falsification [7] where someone maliciously tries to be a primary consumer, as a result, perplexing the cognitive radio community to assume the spectrum is in use, i.e., the primary user is absent but made to think that it is present.

Cognitive radio attempts to determine the areas of the used or unused spectrum by way of determining if a primary user is transmitting in its location. In the cognitive radio community, cognitive radio users analyze the radio spectrum continuously. The three important forms of spectrum sensing techniques for detecting primary user licensed spectrum band: (i) Cooperative spectrum sensing method or collaborative spectrum sensing method. (ii) Transmitter or non-cooperative spectrum sensing approach. (iii) Interference primarily on the basis spectrum sensing technique.

In this paper, we have examined different spectrum detecting procedures utilized in cognitive radio. The void frequency ranges are first dissected by utilizing cognitive radio users, and consequently, various spectrum detecting techniques are utilized. Spectrum sensing characteristics of cognitive radio provide us the arrangement of empty bands. There are three principal methods for range detection: primary transmitter identification, cooperative detection, and obstruction-based detection. In this review, we had examined cooperative spectrum sensing techniques in cognitive radio in detail.

The remaining paper proceeds as follows: the detailed introduction of the proposed techniques for sensing the spectrum under the cooperative spectrum sensing method in Sect. 2. Finally, the conclusion is presented in Sect. 3.

2 Proposed Approaches

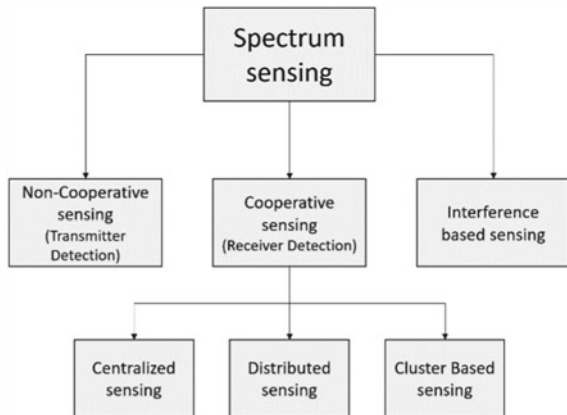
Cooperative detection points out to spectrum sensing methods in which data from the licensed users are accumulated for the detection of the primary user. Cooperative detection can be applied in three types of sensing which are: centralized, distributed, and cluster-based manner.

In the centralized approach, the primary base station plays a role to build up all sensing information from the primary users to detect spectrum holes; however, the distributed method requires sharing of data between the licensed users. In the clustering approach [8], all CRs are grouped right into a cluster, in which one CR acts as cluster head (CH), and all the different CRs act as cluster participants which provide their information to the fusion center (FC) (Fig. 1).

2.1 Coordinated Multiband Spectrum Sensing (CMSS)

The spectrum consists of orthogonal frequency subbands with the particular bandwidth that is the transmission channels for the primary users (PUs). Various secondary users (SUs) and an immobile base stations together make a secondary user network. These SUs and base station have direct communication between them. The SUs are equipped with an antenna that can perform sensing and transmission but at a time. These SUs are capable of sensing a PUs channel per sensing with the help of energy

Fig. 1 Classification of spectrum sensing techniques [9]



detectors employed by radio frequency front end of secondary. The movement of SUs is based on a random waypoint mobility model [10]. At the starting of every time period, SU decides a destination in the network consistently and begins to move toward the destination at a constant velocity. As the SU reaches its destination, it waits for the new time period to begin which is following the same rule. This technique is useful in the case of movable and geographically spread cognitive radio networks.

The problem of broadly spread networks is that if location detecting technology is not provided to SUs, fusing the sensing outcomes to determine the suitable spectrum sensing allocation for the subsequent sensing time becomes difficult for the base station (BS). This complication is resolved using the matrix based on sensing outcomes of SUs and a low-complexity clustering algorithm [11].

2.2 Compressive Spectrum Sensing

Various RF frontends are used by the secondary user to sense each band, so to speed up the spectrum scanning process and minimize the processing time, compressive sensing has been proposed. It permits decreasing the amount of samples required for high-dimensional signal accession whereas maintaining crucial information. Its mechanism is based on directly getting a sparse signal in its compressed form having the maximum information applying a minimum evaluation and then recovering the spectrum with a suitable probability and getting rid of all of the null coefficients to speed up the spectrum scanning. The number of measurements is always associated with the spectrum sparsity level, concluding that the information of spectrum sparsity level will be required to determine an appropriate number of measurements in CR networks. It works on the reconstruction algorithm [12] which combines sampling and sensing processes. It involves three main operations: sparse representation, encoding, and decoding.

The hassle in this method is the uncertainty of sparsity degree in sensible cognitive radio networks because of both the instantaneous activities of licensed users or the time varying fading channels between licensed users and cognitive radios. To overcome this issue, increment in measurements is executed to increase the recovery rate inflicting extra unnecessary energy intake (Fig. 2).

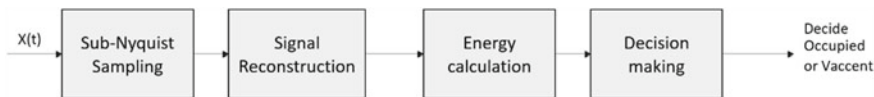


Fig. 2 Block diagram for compressive sensing

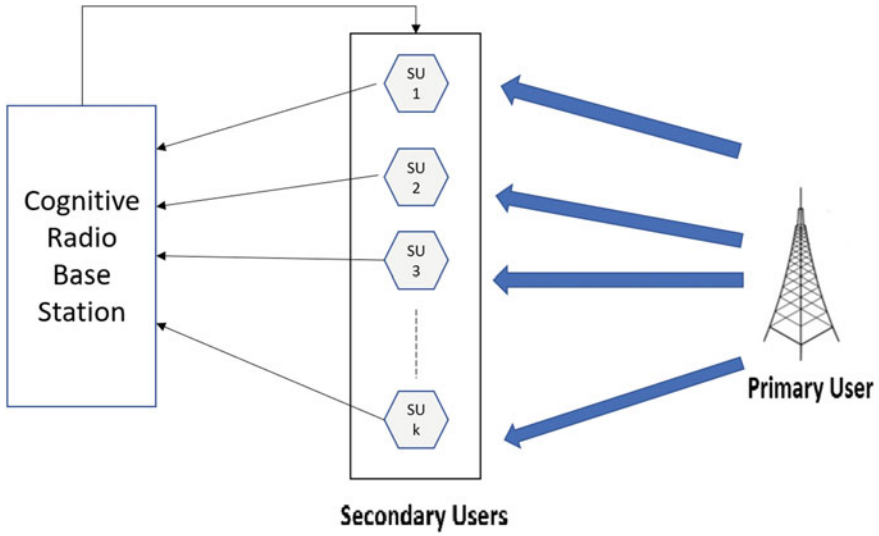


Fig. 3 Representation of centralized spectrum sensing

2.3 Centralized Spectrum Sensing

In centralized spectrum sensing (CSS), fusion center or a base station (BS) is present in the cognitive radio network which can be used in the spectrum sensing mode for decision making. In this, the channel is sensed by a particular secondary user for a consecutive time period to observe a baseband-equivalent signal. Then, BS assembles this information through noisy channels. This information enables in figuring out the available spectrum. In the CSS model, the baseband-equal signal transmitted by the primary user is transferred to secondary users over a flat-fading and time-invariant channel [13] in order to control the cognitive radio traffic.

The process of centralized cooperative sensing takes place in three steps. Firstly, the channel is selected by the BS, so that secondary users can individually perform the local sensing. Secondly, the sensing results are being reported by the users through a control channel. Then, finally, the BS amalgamates the received sensing information to determine the existence of licensed users and informs the unlicensed user about the position of the spectrum (Fig. 3).

2.4 Robust Collaborative Spectrum Sensing Schemes

The purpose of collaborative spectrum sensing (CSS) is to deliver the preferred detection accomplishment underneath noise uncertainty for a massive wide variety of customers. This method complements the possibility of detecting the transmission

of primary customers as it should be. Then, trust value is being calculated for every unlicensed user to determine its suspicious level and reduce its negative impact on cooperative sensing with the aid of temporal and spatial correlations [14], and with the data of the secondary users, the value is decided. This technique enables to gain high detection and false alarm charge. This will further help to achieve high power efficiency via the temporal correlation among alternate points.

2.5 Deep Cooperative Sensing

It is a cooperative spectrum sensing gleaned from convolutional neural networks [15, 16]. The idea of deep cooperative detecting utilizes helpful spectrum detecting in which numerous secondary users work together to get aware of the primary user. Deep cooperative sensing depends on convolutional neural systems. On this, no mathematical displaying of the cooperative spectrum sensing is utilized rather going before information and outcomes are utilized to figure the likelihood [15]. Both spectral and spatial connection of character detecting results is contemplated with the end goal that a situation-specific CSS is empowered in the deep cooperative sensing. This method assists in improving the sensing execution.

2.6 Energy Detection-Based Cooperative Spectrum Sensing

The energy detection procedure can be implemented in the time domain as well as frequency domain. When cooperative spectrum sensing is done using energy detection, secondary users (SUs) provide their results to fusion center [17] in either of the following techniques; (i) Data fusion [18]: Each SU simply passes on the obtained signal from the primary user (PU) to the fusion center. At the fusion center, various fusion methods can be carried out, (ii) Decision fusion [18]: The decision is made by each secondary user on the basis of PU activity, and these results are communicated to the fusion center via communicating channel. Cooperative sensing scheme [19] in the energy detection technique is in the way that first of all, every SU senses the specific frequency spectrum independently with energy detection techniques. Then, it provides its sensing information through the above approaches. After this, a broadcast message is received. The sensing results are updated if this message is useful.

2.7 Multi-objective Clustering Optimization for Multi-channel Cooperative Spectrum Sensing

The inspiration driving behind cooperative spectrum sensing policy for massive scale heterogeneous cognitive radio technology includes many primary channels and an enormous number of secondary users (SUs). These issues occur due to macro and micro point of view. As indicated by macro attitude agencies, the secondary users into groups with the objectives: (i) general force admission minimization; (ii) general throughput expansion, and (iii) between bunch force and throughput decency. This problem is solved by using the subservient sorting genetic algorithm II [20]. This process works as a sub-process on the formation of the bunch determined by way of the macro perspective. According to the micro outlook, the cluster head (CH) yields: (i) these are the best cluster head in which total multi-hop error rate is minimum; (ii) the ideal directing ways from SUs to the cluster heads. Despite using Poisson-binomial distribution, an exclusive and comprehensive K-out-of-N balloting rule [20] is designed for a heterogeneous cognitive radio network that permits the SUs to have uncommon neighborhood detection. Then, an optimized framework is an equestrian to cut the intra-group power cost by together obtaining the top of the line detecting spans and edges of highlight identifiers for the proposed casting a voting rule. Also, as a substitute of a not unusual fixed pattern length check, we likewise have referenced a weighted example size test for quantized delicate determination combination to get a greater energy-efficient proficiency rule underneath heterogeneity. This proves that the mixture of proposed CH choice and cooperation methods offers more energy efficiency and is very robust.

2.8 Green Cooperative Spectrum Sensing and Scheduling

As the cognitive radio networks are comprising of the primary channel (PC), it includes diversified features and secondary users which include different sensing techniques for outstanding PCs. Thereafter, measurement of usual information rate and its fee is achieved, because of the energy consumption brought on from detecting, announcing, and channel switching operation. It previously frames a joint range disclosure and vitality proficiency target to limit the vitality spent per unit of information rate. At that point, a blended whole number nonlinear programming issue is figured to decide: (i) the SU task set for every scheduled PC; (ii) the ideal division of PCs to be arranged for detecting, and (iii) detecting spans furthermore, recognition limits of each secondary user on PCs it is doled out to sense [21]. Later, an identical arched structure is created for explicit occasions of the above combinatorial issue. Aimed at assessment, best detection and sensing limits are inferred scientifically under the homogeneity assumption. An organized requesting heuristic is created based on these in order to arrange channels under the range, vitality, and range vitality restricted systems. From that point onward, task heuristic and scheduling are

presented, and it appeared to perform exceptionally near the comprehensive ideal arrangement. At long last, the performance of the cognitive radio network is numerically dissected under these systems as for various numbers of SUs, PCs, and detecting characteristics.

2.9 Utility-Based Cooperative Spectrum Sensing

Considering the issue in cooperative spectrum sensing in a cognitive radio network (CRN) when there is a presence of many primary channels. Each SU has the freedom to take part in sensing. If SU does not take part in sensing, then it becomes free rider [22] so basically, it saves the energy which is used for sensing, and hence, due to this, efficiency is also increased. While performing spectrum sensing, we all focus on two very common processes, i.e., which channel is to be sensed and what action can be used to sense the channel. This technique uses an algorithm known as the coalition formation algorithm [22]. In this algorithm, the SU picks the coalition which draws out the most data in regards to the status of the comparing channel. SU will choose which channel is to be selected, and it will only select those channels which will give the most information. This method will reduce the uncertainty of the channel status and will further increase the utility of the spectrum. For a free rider F ($F \in F$), the utility function can be interpreted as [22]

$$U(F) = E(F) - H(F) \quad (1)$$

where $E(F)$ is the return regarding stored energy for not engaging in spectrum sensing and $H(F)$ is the compensation for not contributing. Similarly, a contributor C ($C \in C$) is defined as [22]

$$U(C) = R(C) - E(C) \quad (2)$$

2.10 Distributed Spectrum Sensing by Exploiting Sparsity

This technique is based upon the concept of sensing the frequency bands by power spectral density (PSD) [23]. In this technique, by visualizing or by analyzing the power spectral density map, we can find out which frequency band is used and which frequency band is not used, and by analyzing, we can make a rough estimation for choosing a specific spectrum. In this technique, there are generally two methods. The first method is to introduce a narrow band nature of usable spectrum in comparison with unusable spectrum, [23] and the second is finding sparsely located active radios in the operational space. This technique is performed by using the lasso algorithm [23]. In this algorithm, an estimator is developed which can find the undetermined

positions of transmitting cognitive radios. This type of algorithm and PSD simulation in frequency and space minimizes the sensing function for the estimation of a scattered frequency band.

3 Conclusion

Cognitive radio network technology gives the platform to unlicensed users or secondary users to access an empty spectrum band without interrupting licensed users. In this method, the empty spectrum is utilized by the users which further increases spectrum utility. Spectrum sensing is required to find out the availability of the spectrum band for the allocation of secondary users. There are various sensing techniques which we have mentioned in this paper. There are major three important sensing techniques for detecting primary user licensed spectrum bands which are cooperative spectrum sensing method or collaborative spectrum sensing method, the transmitter or non-cooperative spectrum sensing approach, and interference primarily based on the spectrum sensing technique. The void frequency spectrums are analyzed by using cognitive radio users, and for this reason, numerous spectrum sensing techniques are used. In this paper, we have performed comparative studies on different types of cooperative spectrum sensing method, and by analyzing all types of cooperative spectrum sensing method, we concluded that this technique is more efficient and can further be used in spectrum sensing to gain or achieve the highest possible results.

References

1. Jiang C, Chen Y, Ray Liu KJ, Ren Y (2012) Analysis of interference in cognitive radio networks with unknown primary behavior. In: IEEE international conference on communications (ICC)
2. Zhou Y, Yao Y-D (2015) Secondary user scheduling in cognitive radio networks with transmit beamforming for interference mitigation. In: 36th IEEE Sarnoff symposium
3. Min R, Qu D, Cao Y, Zhong G (2008) Interference avoidance based on multi-step-ahead prediction for cognitive radio. In: 11th IEEE Singapore international conference on communication systems
4. Balogun V, Krings A (2014) An empirical measurement of jamming attacks in CSS cognitive radio networks. In: IEEE 27th Canadian conference on electrical and computer engineering (CCECE)
5. Slimeni F, Scheers B, Chtourou Z, Le Nir V (2015) Jamming mitigation in cognitive radio networks using a modified Q-learning algorithm. In: International conference on military communications and information systems (ICMCIS)
6. Kachroo A, Ekin S (2018) Impact of secondary user interference on primary network in cognitive radio systems. In: IEEE 88th vehicular technology conference (VTC-Fall)
7. Ngomane I, Velepini M, Dlamini SV (2008) The detection of the spectrum sensing data falsification attack in cognitive radio Ad Hoc networks. In: Conference on information communications technology and society (ICTAS)

8. Bai Z, Wang L, Zhang H, Kwak K (2010) Cluster-based cooperative spectrum sensing for cognitive radio under bandwidth constraints. In: IEEE international conference on communication systems
9. Omer AE (2015) Review of spectrum sensing techniques in cognitive radio networks. In: International conference on computing, control, networking, electronics and embedded systems engineering
10. Shahrabi B, Rahnavard N, Vosoughi A (2017) Cluster-CMSS: a cluster-based coordinated spectrum sensing in geographically dispersed mobile cognitive radio networks. *IEEE Trans Veh Technol* 66(7)
11. Perez J, Santamaria I (2018) Advanced signal processing group, University of Cantabria. Adaptive Clustering Algorithm for Cooperative Spectrum Sensing in Mobile Environments. In: IEEE international conference on acoustics, speech and signal processing (ICASSP)
12. Candès EJ, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inform Theory* 52(2)
13. Shinde SC, Jadhav AN (2016) Centralized cooperative spectrum sensing with energy detection in cognitive radio and optimization. In: IEEE international conference on recent trends in electronics information communication technology, May 20–21, 2016
14. Li H, Cheng X, Li K, Hu C, Zhang N, Xue W (2014) Robust collaborative spectrum sensing schemes for cognitive radio networks. *IEEE Trans Parallel Distrib Syst* 25(8)
15. Lee W, Kim M, Cho D-H (2019) Deep cooperative sensing: cooperative spectrum sensing based on convolutional neural networks. *IEEE Trans Veh Technol* 68(3)
16. Zheng S, Chen S, Qi P, Zhou H, Yang X (2020) Spectrum sensing based on deep learning classification for cognitive radios. *China Commun* 17(2)
17. Fan R, Jiang H (2010) Optimal multi-channel cooperative sensing in cognitive radio networks. *IEEE Trans Wireless Commun* 9(3):1128–1138
18. Atapattu S, Tellambura C, Jiang H (2011) Energy detection based cooperative spectrum sensing in cognitive radio networks. *IEEE Trans Wireless Commun* 10(4)
19. Xuping Z, Jianguo P (2007) Energy-detection based spectrum sensing for cognitive radio. In: IET conference on wireless, mobile and sensor networks
20. Celik A, Kamal AE (2016) Multi-objective clustering optimization for multi-channel cooperative spectrum sensing in heterogeneous green CRNs. *IEEE Trans Cogn Commun Network* 2(2)
21. Celik A, Kamal AE (2016) Green cooperative sensing and Scheduling in heterogeneous cognitive radio networks. *IEEE Trans Cogn Commun Network* 2(3)
22. Li H, Xing X, Zhu J, Cheng X, Li K, Bie R, Jing T (2017) Utility-based cooperative spectrum sensing scheduling in cognitive radio networks. *IEEE Trans Veh Technol* 66(1)
23. Bazerque JA, Giannakis GB (2010) Distributed spectrum sensing for cognitive radio networks by exploiting sparsity. *IEEE Trans Sign Process* 58(3)

Frequency Reconfigurable Antenna for Energy Efficient WBAN Application



Ridhima Puri, Nimisha Sharma, Asmita Rajawat, and Sindhu Hak Gupta

Abstract The paper presents dual-mode reconfigurable antenna that can toggle between 2.4 GHz (ISM Band) and 4.43 GHz (UWB) with the help of a PIN diode. The antenna resonating in medically approved ISM band and UWB band along with the judicious use of energy makes it appropriate for WBAN application. The main focus is laid on WBAN applications and power management techniques with the help of concepts of STEM and duty cycle. The value of S_{11} obtained when the antenna resonates at 2.4 GHz is -15.216 dB, and the value of S_{11} obtained when the antenna resonates at 4.43 GHz is -35.08 dB. Measured gain for antenna when PIN diode is switched to ON state and OFF state is 0.1068 dBi and 2.345 dBi, respectively. The performance of the antenna is assessed on the basis of return loss, gain of the antenna, energy, and power consumption values of the antenna for ISM and UWB.

Keywords ISM · Reconfigurable antenna · STEM · UWB · WBAN

1 Introduction

Wireless body area networks (WBANs) are a rising trend in today's time. WBAN is formed on the radio frequency, which connects the small sensor nodes to a sink node, i.e., energy constrained with additional processing capabilities. These sensor nodes deployed on/in human body help in identifying multiple abnormalities in early detection of diseases [1]. On account of characteristics like minimum consumption of power, no delay, high reliability, and independent functioning of the node, WBAN is

R. Puri · N. Sharma (✉) · A. Rajawat · S. H. Gupta
Amity University, Noida, Uttar Pradesh, India
e-mail: nimisha.17.sharma@gmail.com

R. Puri
e-mail: ridhima.17.puri@gmail.com

A. Rajawat
e-mail: arajawat@amity.edu

S. H. Gupta
e-mail: shak@amity.edu

used in the field of regular patient monitoring, bio feedback, defense, healthcare, and assisted living [2]. Reconfigurable antenna is the key feature in WBAN applications as they support wide area communication. In recent research, significant attention has been paid to reconfigurable antennas on account of their ability to resonate at various frequencies. Reconfigurable antenna is best used in WBAN applications due to their ability to improve antenna performance of RF system and change the geometry of the antenna with changing ecological conditions in a controlled manner and provide a compact design. In addition to this, reconfigurable antenna attracts the researchers as it holds the capability to tune the operating frequency of the antenna in order to change the operating bands and filter out interference signals [3]. Several reviews and publications have been published on the design of a reconfigurable antenna. The existing method provides ways to accurately and efficiently determine the type of antenna which can be used in the design to provide ease in fabrication. The usage of microstrip patch antenna in the antenna design has gained significant importance. The authors suggest that on account of planar configuration, ease in integration, robust design, and light weight, MSPA can be used over other types of antennas [4]. The procedure used for making the antenna operate on dual frequency is one of the most commonly used method of using a PIN diode which is a lumping element, as it provides ease of fabrication and reduced cost [5]. The authors suggest that good radiation pattern can be obtained in dual frequency antenna if the voltage biasness is kept in control. A series of research has been carried out to investigate the approach of providing power to the microstrip patch. It is observed that on account of characteristics like simplicity in modeling and impedance matching, inset feed can be incorporated in the design [6]. However, in this paper, the main aim is to utilize the antenna parameters to generate a cost effective, low power/energy consuming antenna by incorporating the protocols and concepts of STEM, which is a step toward creating sustainable technology.

The PIN diode in the antenna switches between ON mode and OFF mode at frequencies of 2.4 and 4.43 GHz. The first resonant frequency (2.4 GHz) is chosen in ISM band because it supports high data rate applications. The second resonant frequency is chosen to lie in UWB range as the band uses low energy level for short range communication. UWB is also useful in medical applications as it has low penetration radiation and consumes low processing energy [7].

The feature of reconfigurable antenna, i.e., redirecting the radiation pattern of the antenna toward saving the power with the help of energy management techniques, is reviewed in the paper. STEM technology which stands for sparse topology and energy management holds major applications in energy and power consumption, namely two channels: wake-up and data channel which is used in STEM. These channels switch between monitor state and transfer state depending on whether the data is to be transmitted or received by the sensor node. The data channel sits idle and operates only when the wake-up channel senses some input signal. STEM improves the network lifetime by exploiting the fact that most of the time, the network is asleep and wakes up only after sensing its environment [8]. Another basic and most commonly used power management techniques is duty cycling, i.e., the ratio of listen period/ (listen period + sleep period) [9]. The two parameters, namely S_{11} and gain,

associated with the patch of the antenna are used to record the overall performance of the system. In order to simulate the antenna design, computer simulation technology (CST) studio software is used. The flow of the paper is as follows: the antenna design is briefly discussed in Sect. 2, the antenna simulation results is in Sect. 3 of the paper, reconfigurable antenna in WBAN scenario is discussed in Sect. 4, and the conclusion of the work is presented in Sect. 5.

2 Antenna Design

Figures 1 and 2 show the design of the microstrip patch antenna. MSPA offers three-fold design, consisting of microstrip patch, dielectric substrate, and the ground plane.

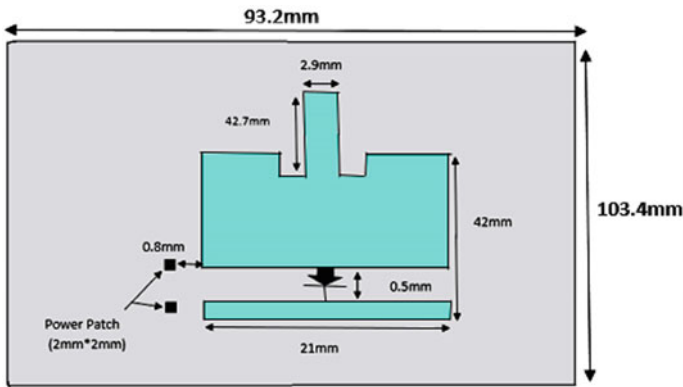
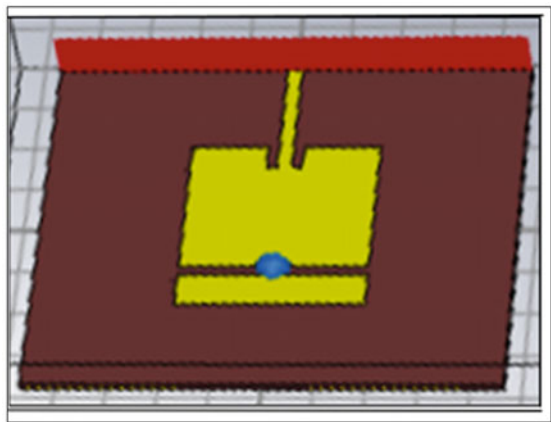


Fig. 1 Antenna dimensions

Fig. 2 Antenna in CST software



Microstrip patch antenna (MSPA) is used in the design because of its characteristics like low profile, ease of fabrication, and high-speed data transfer [10]. To supply power to the antenna, inset feeding technique has been incorporated which is easy to fabricate. The sinusoidal current distribution is more toward the center of the patch as compared to the edges. As there is an inverse relation between amount of current and impedance ($Z = VI$), thus the feed line has been placed toward the center of the patch to provide impedance matching to the antenna. The work on the antenna [11] has been extended by doing suitable modifications in the design to yield beneficiary results, suitable for WBAN applications. Two copper patches of size (2 mm * 2 mm) have been incorporated on the side of the patch at a distance of 0.8 mm away from the patch, for easy power supply to the design for further fabrication and testing. The defected ground structure is removed from the design of the antenna to significantly reduce the gain of the antenna. The patch dimension is (21 mm × 42 mm). The height of the substrate helps in reducing the amount of fringing in the antenna and also provides mechanical strength to the antenna. The dimensions of FR4 dielectric substrate is (93.2 mm × 103.4 mm). FR-4 has a dielectric constant of 4.4. The dimensions of copper ground plane is (93.2 mm × 103.4 mm). The antenna feed is of width 2.9 mm and length 42.7 mm.

Following equations are taken into consideration for calculating the dimensions of the design [12]:

$$W = \frac{C}{2f_0\sqrt{\frac{\epsilon_r+1}{2}}} \quad (1)$$

$$L = \frac{c}{2f_0\sqrt{\epsilon_{\text{eff}}}} - 2\Delta L \quad (2)$$

$$\epsilon_{\text{eff}} = \frac{\epsilon_r+1}{2} + \frac{\epsilon_r-1}{2} \left[1 + 12 \frac{h}{w} \right]^{-\frac{1}{2}} \quad (3)$$

where

$c = 3 \times 10^8$ m/s, speed of light,

$f_0 = 2.4$ GHz, resonance frequency;

$\epsilon_r = 4.3$, relative permittivity of substrate;

$h = 1.6$ mm, height of the substrate;

ϵ_{reff} = effective dielectric constant;

Antenna is simulated for different values of feed line width, and an optimized value of 2.9 mm is used to provide the best impedance matching at the desired frequency bands. A gap after three-fourth length of the microstrip patch is introduced in the antenna in order to insert the lumping element, i.e., PIN diode. The circuit model of PIN diode at RF frequencies for ON/OFF state is shown in Fig. 3. The pin diode behaves like a variable resistor at ON/OFF state. Figure 3 shows that the ON state has low resistance R_s which is a major reason for insertion loss. The OFF state circuit is reverse biased and has a parallel combination of reverse bias resistance R_P and total capacitance C_T , which leads to isolation of the circuit.

Fig. 3 RLC equivalent circuit for ON state and OFF state, respectively

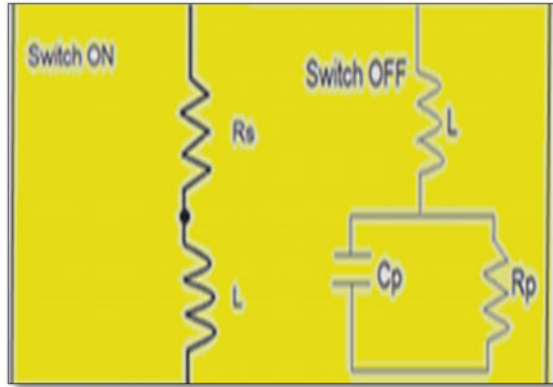


Table 1 Measurements of elements used in designing the antenna

Element	Length (mm)	Width (mm)	Thickness (mm)
Ground plane	93.2	103.4	0.4
Dielectric substrate	93.2	103.4	1.6
Microstrip copper patch	21	42	0.1
Feed line	42.7	2.9	0.1

3 Antenna Simulation Results

Return loss (S_{11})

The scatter parameter of an antenna describes the relationship between the input and output terminals. The S_{11} parameter varies with frequency. It is defined as the power reflected from the antenna, and thus, it is also referred to as the return loss [13]. When S_{11} equals to zero, it is an indication that no power is radiated. When S_{11} is obtained as -10 dB, out of the value, -7 dB is reflected, and 3 dB is radiated. The value of S_{11} obtained for antenna operating in ON state is -15.216 dB where antenna resonates at a frequency of 2.4 GHz as shown in Fig. 4a, and value of S_{11} obtained for antenna operating in OFF state is -35.08 dB where antenna resonates at a frequency of 4.43 GHz (UWB band) as shown in Fig. 4b.

Gain

The amount of power transmitted in the direction of the peak radiation of an isotropic source is known as gain of antenna. Gain in ON and OFF state is found to be -0.1068 dBi and 2.345 dBi, respectively. The value of gain signifies the amount of power transmitted and the amount of RF energy which is being radiated by the antenna. The antenna has a very low gain in ON body state which prevents the adverse biological effects that can be caused due to high RF radiations. Tissue rupture can occur in

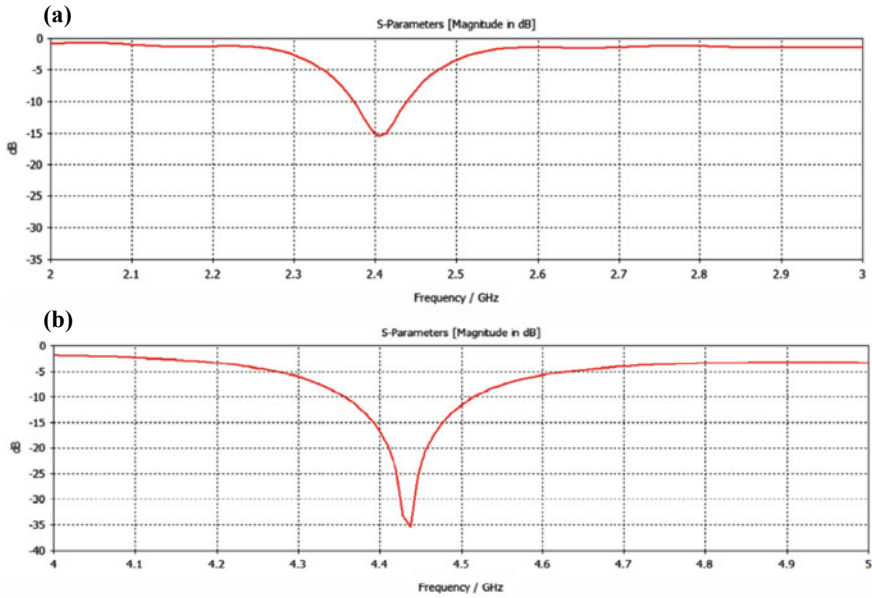


Fig. 4 a S_{11} obtained for ON state, b S_{11} obtained for OFF state

the body due to the inability of the body to cope up with the high thermal heat if the wearable antenna used has significantly high gain. Sensitive areas such as eyes are vulnerable to relative heating because of less blood flow, which is required to overcome the heating effect. However, for negligible value of gain, for which the body is not adversely affected by the RF radiation and the antenna can be said to produce ‘non-thermal’ effects. The gain obtained for antenna for ON state is 0.1068 dBi which is significantly low to have non-thermal effects on the body, making the antenna suitable to be used in wireless body area network (WBAN) applications. The value of gain obtained from the simulation of the microstrip patch antenna is shown in Fig. 5a, b.

When the antenna is resonating at 2.45 GHz, the main lobe magnitude is -0.106 dBi, the main lobe direction is 1.0 deg, and the side lobe level is -20.0 dB. When the antenna is resonating at 4.43 GHz, the main lobe magnitude is 2.35 dBi, the main lobe direction is 44.0 deg, and the side lobe level is -10.7 dBi as shown in Fig. 6a, b, respectively.

Surface Current

Surface current of an antenna is an approximation of the amount of surface charge that exists on the surface of the conductor to charge the antenna with a specific electric potential. The plots for surface current have been shown in Fig. 7a for ON state and Fig. 7b for OFF state.

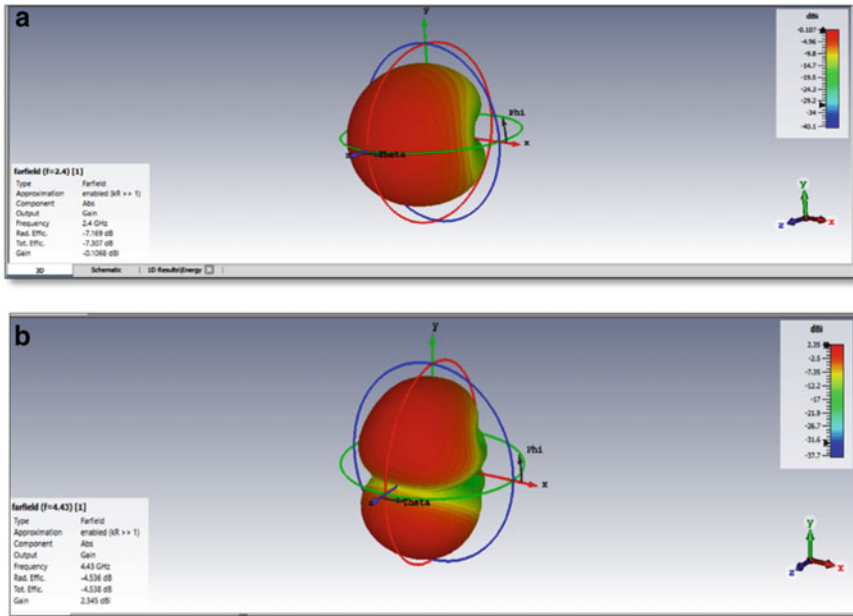


Fig. 5 a ON state gain for antenna, b OFF state gain for antenna

4 Reconfigurable Antenna in WBAN Scenario

Wireless body area network (WBAN) is a key technology to enable information communication in patient-centric tele-health [14]. STEM innovation that represents sparse topology and energy management holds significant application in the antenna design. It uses wake-up and data channel to switch between monitor and transfer state depending on whether the data is to be transmitted or received by the sensor node. Wake-up channel transmits the date to the receiver after informing the receiver. Data channel transmits data using various MAC protocols. This technology extends the system life [15]. Duty cycling is another technique to manage. Duty cycle is the ratio of listen period/ (listen period + sleep period). Ideal duty cycle value helps in avoiding high delay and transient energy that can be caused due to startup cost [16]. The average duty cycle calculated for 10 s is observed as 0.15 in the paper. To calculate the transmitted power by the antenna, the Friss equation as shown in Eq. (4) is used.

$$P_t = \frac{(E_b/N_o) * N_o * R(4\pi*)^2 * d_o^2 * L * (d/d_o)^\gamma}{G_t G_r \lambda^2} \tag{4}$$

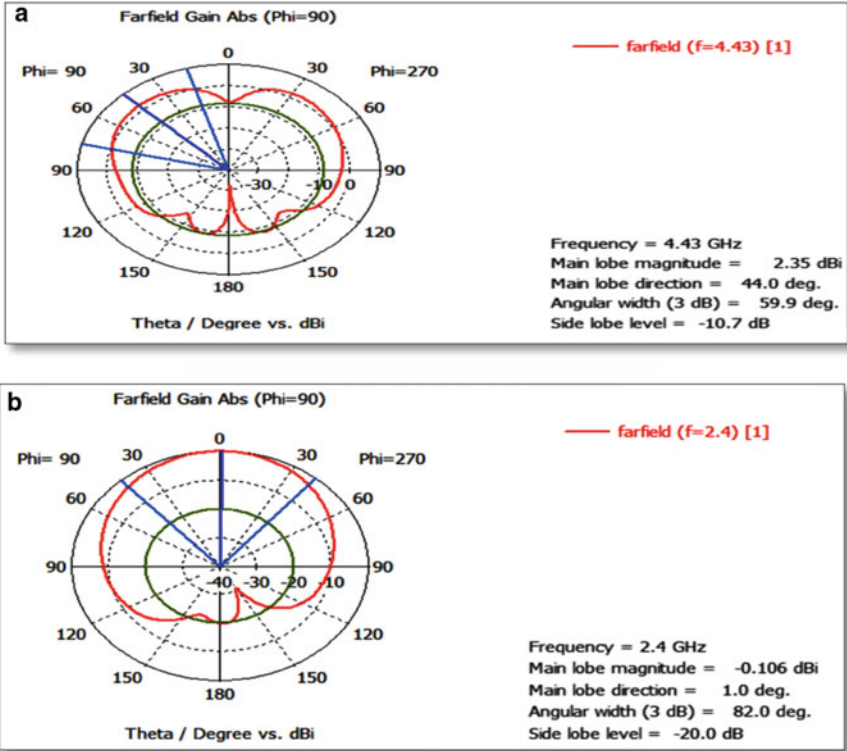


Fig. 6 a Polar plot for radiation pattern of ON state, b Polar plot for radiation pattern OFF state

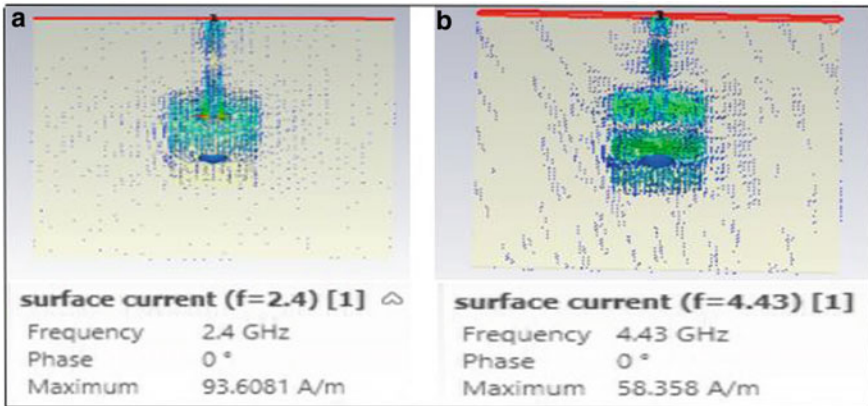


Fig. 7 a ON state surface current, b OFF state surface current

where:

G_t, G_r = gain of transmitter and receiver of antenna, respectively, R = data rate, (E_b/N_o) = Ratio between energy per bit and the noise level for different M-ary modulation schemes, λ = Path loss component, P_t = Power transmitted.

The values taken of the above-mentioned parameters in the paper are:

$(E_b/N_o) = 14$, $N_o = -201$ dB, $d_o = 10^{(-4)}$, $d = 50$, $\gamma = 3.5$, G_t, G_r = ON/OFF state gain achieved, i.e., 0.1068 dBi for ON state and 2.345 dBi for OFF state.

Using Eq. (4), the power and energy which is required by the antenna is calculated. ON state power is 1.0693×10^{35} W, and energy is 1.0693×10^{36} J. OFF state power is -2.0041×10^{30} W, and energy is -2.0041×10^{31} J.

Using Eq. (4) and the above-calculated results, it is observed that the power and energy consumption of the antenna can be reduced if the antenna mode is switched to ON state via PIN diode only when the information is being received, and the power flow to the antenna is stopped in OFF state when there is no input signal, as compared to an antenna which is sensing at all the time. This inference can also be observed by taking an example of patient monitoring system [17]. In this system, the antenna enters a low duty cycle in ON state. Other transmitting/ receiving antenna continues to work in OFF state (sleep cycle) till the time an alarming situation occurs. Once the alarming situation occurs, the antenna goes back to a mode less than the threshold. The antenna switches back to the ON state that is low duty cycle and goes back to the OFF state (sleep cycle) when the body conditions are normal.

5 Conclusion

The paper presents the design of a reconfigurable antenna which works on the principle of energy conservation, operating at 2.4 GHz in ISM band for ON state and 4.43 GHz in ultrawide band for OFF state in WBAN application. It is observed that the mode in ISM band is suitable for a long-range communication while the mode in the ultrawide band range is suitable for short-range communication. The antenna resonates at a frequency of 2.4 GHz (ON state) when sensing data and resonates at 4.43 GHz (OFF state) during no data transmission, in order to save the power or reduce the battery consumption. The S_{11} parameter evaluated for ON state is -15 dB and in OFF state is -35 dB. The antenna gain achieved is -0.1068 dBi when PIN diode is in ON state and 2.345 dBi when diode is switched to OFF state. With the help of Friss equation, the power consumed by the antenna when diode is switched between the two states is also calculated. The power which is required by the antenna when the PIN diode is switched ON is 1.0693×10^{35} W. The ON state acts like data sensing cycle. When diode is switched to OFF, the power consumed is 2.0041×10^{30} W. The OFF state with less power consumption as compared to ON state acts like data channel and sits idle until an input signal is received.

References

1. Ullah S, Higgins H, Braem B, Latre B, Blondia C, Moerman I, Kwak KS (2012) A comprehensive survey of wireless body area networks. *J Med Syst* 36(3):1065–1094
2. Boulemtafes A, Badache N (2016) Design of wearable health monitoring systems: an overview of techniques and technologies. In: *mHealth ecosystems and social networks in healthcare*. Springer International Publishing, Berlin, pp 79–94
3. Mohamadzade B, Simorangkir RBVB, Hashmi RM, Shrestha S (2019) Low-profile pattern reconfigurable antenna for wireless body area networks. In: *International conference on electromagnetics in advanced applications (ICEAA)*, pp 546–547
4. Parikh R, Joshi P, Rawat A (2018) Designing of rectangular microstrip patch antenna at 2.48 Ghz frequency for IRNSS application. In: *2018 2nd international conference on electronics materials engineering & nano-technology (IEMENTech)*, pp 1–3
5. Ojaroudi Parchin N, Jahanbakhsh Basherlou H, Al-Yasir YIA, Abdulkhaleq AM, Abd-Alhameed RA (2020) Reconfigurable antennas: switching techniques—a survey. *Electronics*, pp 546–547
6. Hakanoglu BG, Turkmen M (2017) An inset fed square microstrip patch antenna to improve the return loss characteristics for 5G applications. In: *General assembly and scientific symposium of the international union of radio science (URSI GASS) 2017, XXXIInd*, pp 1–4
7. Jin G, Deng C, Xu Y, Yang J, Liao S (2020) Differential frequency-reconfigurable antenna based on dipoles for sub-6 GHz 5G and WBAN applications. *Antennas Wireless Propagat Lett IEEE* 19(3):472–476
8. Saraswat J, Bhattacharya PP (2013) Effect of duty cycle on energy consumption in wireless sensor networks. *Int J Comput Networks Commun (IJCNC)* 5(1)
9. Yoo H, Shim M, Kim D (2012) Dynamic duty-cycle scheduling schemes for energy harvesting wireless sensor networks. In: *IEEE communication letters*, vol 16, No 2. Kyungpook National University, Korea. IEEE, New York
10. Tong X, Liu C, Chen Y, Zhu J, Yang X, Guo H, Liu X (2019) A dual-mode multi-polarization millimeter wave wearable antenna for WBAN applications. In: *2019 IEEE MTT-S International Microwave biomedical conference (IMBioC)*, vol 1, pp 1–3
11. Sharma N, Puri R, Rajawat A, Hak S (2020) Performance analysis of reconfigurable antenna in WBAN applications. In: *International conference on intelligent computing and control systems (ICICCS 2020)*
12. Balanis C (2005) *Antenna theory, analysis and design*. Wiley, New York
13. Singh I, Tripathi VS (2011) Micro strip patch antenna and its applications: a survey. *Int J Comp Tech Appl* 2(5):1595–1599
14. Garg R, Bartia P, Bahl I, Ittipiboon A (2001) *Microstrip antenna design handbook*. In: 253316 Artech House Inc. Norwood, MA, pp 168
15. Ullah S, Higgins H, Braem B, Latre B, Blondia C, Moerman I, Kwak KS (2012) A comprehensive survey of wireless body area networks. *J Med Syst* 36(3):1065–1094
16. Charfi F, Mohamed Bouyahi M (2012) Performance evaluation of Beacon-enabled IEEE 802.15.4 under NS2. *Int J Distrib Parallel Syst (IJDPS)* 3(2):67–79
17. Yang L, Cheng B, Zhu Y, Li Y (2016) Compact antenna with frequency reconfigurability for GPS/LTE/WWAN mobile handset applications. *Int J Antennas Propagat* 1

Characterization of Millimeter Waves RoFSO Link for 5G Applications



M. Vishnu Kartik, Manisha Samal, Sanya Arora, and Sanmukh Kaur

Abstract Radio over free-space optics (RoFSO) communication system is useful for different applications where optical fibers are not possible, like in rural areas, to provide wireless services easily and more effectively. Radio over free-space optics framework is one of the developing advancements toward the 5G networks. With the consistent increase of data and media associations, radio frequency (RF) range has additionally expanded, and subsequently, there emerges a need to move from RF carrier to optical carrier. Additionally, the RoFSO innovation can be applied as an all-inclusive technology for both the fiber and free-space optical communication systems and in this manner stretching out broadband network to underserved regions. In this work, we characterize the performance of a millimeter waves RoFSO link considering different modulation techniques for 5G applications. Quality of received signal has been analyzed as function of transmission range, transmitted power, data rate, and beam divergence considering different modulation formats.

Keywords Free-space optics · Q-factor · Data rate · Beam

1 Introduction

Free-space optics (FSO) manages the exchange of data through free-space medium between transmitter and receiver. Data is being transferred by propagation of light in this environment. It works like optical fiber cables, with the main difference that transmission of data through FSO does not require guided medium links [1]. As

M. Vishnu Kartik (✉) · M. Samal · S. Arora · S. Kaur
Amity School of Engineering and Technology, Amity University, Noida, India
e-mail: vishnukartk123@gmail.com

M. Samal
e-mail: manishasamal2000@outlook.com

S. Arora
e-mail: arorasanya40@gmail.com

S. Kaur
e-mail: skaur2@amity.edu

a result of wireless communication technique, it also faces a set of difficulties in reliable transmission of signals as compared to optical fiber cables. In this technique, a narrow beam of light launched at the transmitter station goes through the air and afterward received at the receiver's station. Transmitters and receivers need to be in an appropriate line of sight (LoS) view for avoiding loss of information [2].

As of now, huge development and progression has been seen in information exchange and data communication. As the use of video conferencing, fast Web, and so forth has expanded to a more noteworthy degree, the transfer speed requirements have also been increased [3]. With the consistent increase in data and media associations, utilization of radio frequency (RF) range has additionally expanded, and subsequently there have emerged a need to move from RF carrier to optical carrier [4]. Optical carrier does not require any spectrum authorizing and hence is an attractive possibility for high data transfer capacity applications [5].

Radio over fiber (RoF) and radio over free-space optics (RoFSO) are the two innovations that can be applied for optical fiber and free-space optical communication systems, respectively [6]. In addition, RoFSO communication is one of the preferred one compared to other accessible choices for executing 5G [7]. It has numerous points of interest like permit free, high transfer speed and information rate, simple sending, less working cost, and high security [8, 9].

RoFSO is a minimal effort, more effective, high transmission capacity, and free-space optics (FSO)-based framework for giving rapid communication joins (100 Gb/s) to associate versatile base stations without the need of fiber links [10]. This innovation proves to be useful where the links are not possible because of significant expenses.

In this paper, the characterization of millimeter waves RoFSO link has been performed for 5G applications. Quality of the received signal has been analyzed as a function of input power, bit rate, propagation range, and beam divergence considering different modulation schemes.

2 System Design

OptiSystem is used to design the RoFSO framework. The design incorporates a transmitter, FSO channel, and a receiver. The target of the design is to generate modulated RF signal, and the resulting from the RF signal and the optical carrier generated from CW laser is modulated and is then fed to FSO channel. The transmitter section comprising of user bit sequence generator, NRZ modulator, CW laser, Mach-Zehnder modulator, and DPSK sequence generator. The bit sequence generator generates 10 GB/s. Then, the duo-binary phase shift keying (DPSK) modulated scheme is utilized for extending the spectral efficiency of the framework. The signals coming from the DPSK modulator is then sent to M-ray pulse generator for converting into M-ray pulse and afterward is modulated using 60 GHz RF carriers. The output RF signal is modulated with optical carrier (laser), i.e., CW laser with frequency 193.1 THz by the Mach-Zehnder modulator (MZM) and is then fed to FSO channel

under the downpour climate conditions. The laser diode has wavelength of 1550 nm. For the FSO channel, the range is set up to 1 km as a default value. The signal transmits through FSO, transmitter power being 0 dBm, and divergence angle being 2 mrad. The receiver comprises of a PIN photodetector with 1 A/W responsivity, and 10 nA dark dull current is utilized for detection of the optical signal prior to transmission through the filter headed by following the PIN photodetector. A low pass Bessel filter cuts off the higher frequency components and thereby the filter noises from the signal. The BER analyzer analysis the performance and indicates the quality of the signal by of the minimum BER or Q-factor. Modulation techniques play a vital role in the characterization of RoFSO link [11].

2.1 Modulation Techniques Chosen for Analysis

- **DPSK:** In differential phase shift keying (DPSK), the phase of the regulated signal is moved comparative with the past signal component. No reference signal is considered here. The signal phase follows the high or low condition of the past component. The DPSK technique need not require a reference oscillator.
- **QPSK:** The quadrature phase shift keying allows signal to transfer double the information as normal PSK using the same bandwidth, which sends two bits of computerized data one after another called as bigits.
- **BPSK:** Binary phase shift keying (BPSK) utilizes two stages which are isolated by 180° , thus can likewise be named 2-PSK. It is generally DSBSC modulation technique, for message being the digital information.
- **QAM:** Quadrature amplitude modulation (QAM) is a method for combining two amplitude modulated (AM) signals into a single channel, in this way doubling the effective transmission. QAM is utilized with pulse amplitude modulation (PAM) in advanced frameworks, particularly in remote applications (Fig. 1).

3 Simulation Results

The aim of this work is to study characterization of RoFSO link considering different modulation techniques and internal parameters of the system. The global parameters of the system have been listed in Table 1.

The Q-factor has been analyzed for different modulation techniques (DPSK, QPSK, QAM, BPSK) as function of transmission range, transmitted power, data rate, and beam divergence.

We can observe from plot in Fig. 2 that the quality factor varies with the range for different modulation techniques. As it is observed, Q-factor constantly decreases with the increase in range. The least quality factor is obtained in BPSK modulation scheme. The range is taken from 0 to 5000 m. Maximum quality factor is obtained in DPSK modulation, and it is shown with the red curve in plot. Q-factor is about

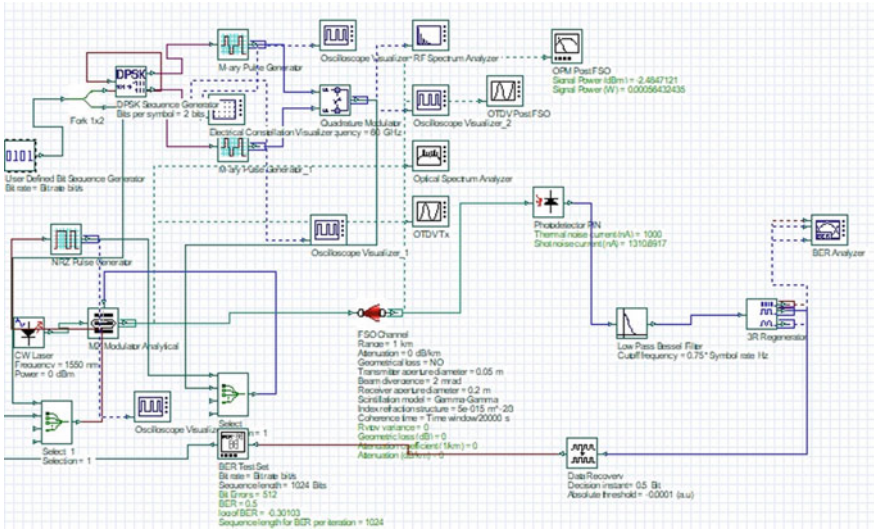


Fig. 1 A block diagram of RoFSO link

Table 1 Simulation parameters

Parameter	Value
Bit rate (Gbps)	10
Range (km)	1
Wavelength (nm)	1550
Modulation scheme	DPSK, QPSK, BPSK, 8-QAM, 16-QAM
Responsivity (A/W)	1
Input power (dBm)	-10 to 0
Sequence length (bits)	512
Samples per bit	64
Samples rate (GHz)	3200
Beam divergence (mrad)	2-4
Dark current (nA)	10
Cut off frequency (GHz)	2.4

1500 at a range of 1000 m, and an acceptable signal quality is achieved at a distance of 5000 m with this modulation scheme. Minimum is seen in BPSK with a value of Q-factor 700 at 1000 m and about 300 at 2000 m.

We can observe from the plot in Fig. 3 that the value of Q-factor varies with power for different modulation techniques. As it is observed, Q-factor constantly increases with power. Q-factor is observed by varying input power from -10 to 0 dBm. Here, DPSK shows the maximum value of quality factor, with increase up to a Q-factor

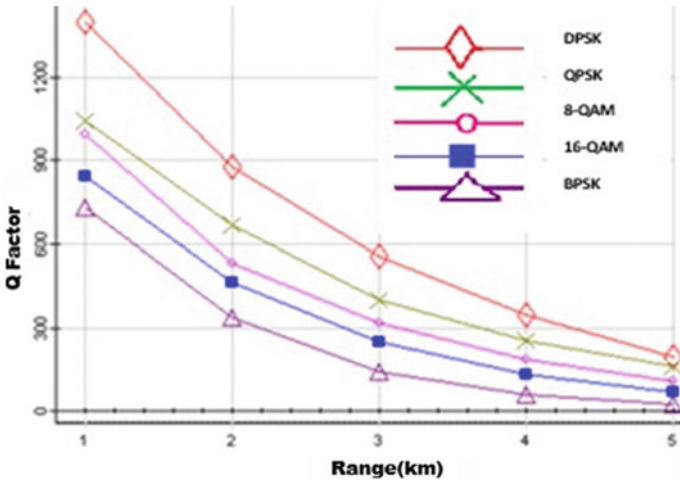


Fig. 2 Variation of Q-factor with range for different modulation techniques

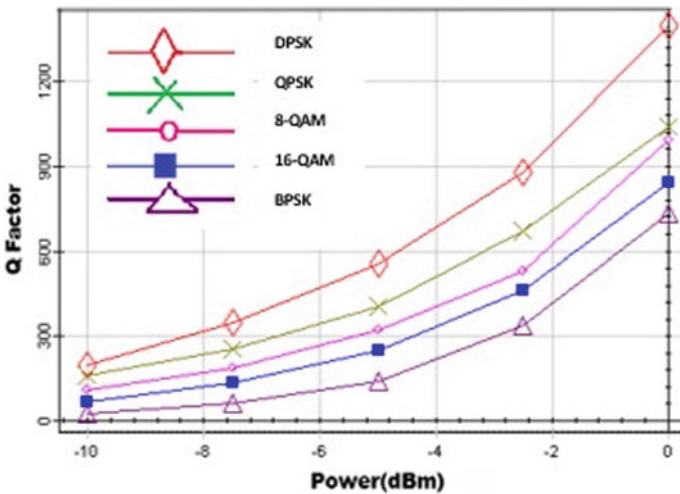


Fig. 3 Variation of Q-factor with power for different modulation techniques

1300 at 0 dBm power. QPSK shows peak value of Q-factor of 950. Least value of Q-factor is observed in case of BPSK with peak value of 700 at 0 dBm power. Hence, the effective modulation technique is DPSK as it shows the highest quality factor.

We can observe from the plot in Fig. 4 that the value of Q-factor varies with the change in data rate for different modulation techniques. As it is observed, Q-factor constantly decreases with data rate. The least value of Q-factor among all modulation techniques is BPSK, and 8-QAM, 16-QAM, QPSK modulation techniques performed better than this. The data rate is taken from 1 to 10 Gbps. Maximum value of quality

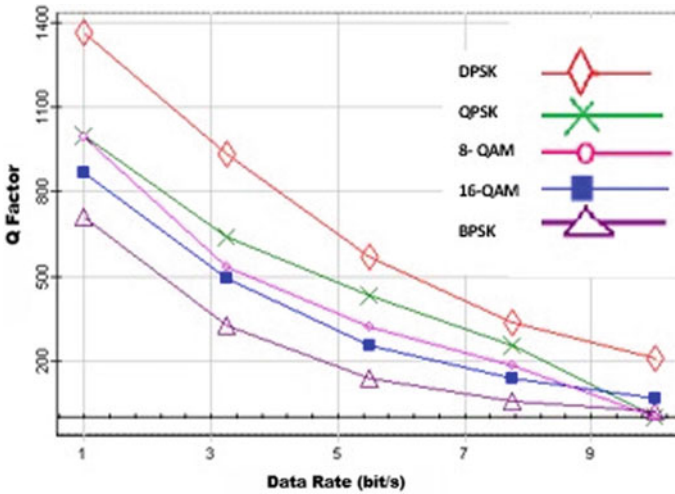
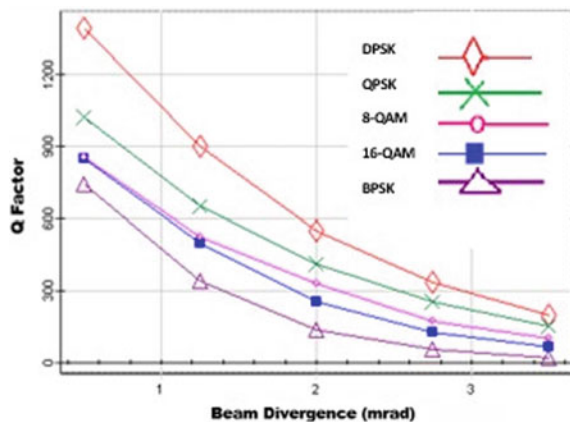


Fig. 4 Variation of Q-factor with data rates for different modulation techniques

factor is obtained with DPSK modulation shown with the red line in the plot. The value of quality factor is about 1400 at data rate 1 Gbps with DPSK. QPSK, 8-QAM, 16-QAM, and BPSK types of modulation techniques show the variation of Q-factor with respect to bit rate. Minimum is observed in case of BPSK which shows Q-factor 750 at 1 Gbps and about 300 at 3 Gbps.

We can observe from the plot in Fig. 5 that the value of quality factor varies with the beam divergence for different modulation techniques. As it is observed, Q-factor constantly decreases with the increase in beam divergence. Least value of quality factor among all modulation techniques is observed in case of BPSK modulation technique. The beam divergence is taken from 0.5 mrad to 3.5 mrad. Highest Q-factor is observed with DPSK modulation, which is shown with the red line in the

Fig. 5 Variation of Q-factor with beam divergence for different modulation techniques



graph. The value of Q-factor is more than 1300 at 0.5 mrad beam divergence for this scheme.

QPSK, 8-QAM, 16-QAM, and BPSK types of modulation techniques show variation of Q-factor with respect to beam divergence. Minimum is observed in case of BPSK which shows the value of Q-factor 700 at 0.5 mrad and about 330 at 1.3 mrad.

Figure 6 indicates the eye diagram of received signal in case of BPSK modulation scheme at a data rate and range of 1 Gbps and 1 km, respectively.

The clear opening of eye of BER analyzer with a Q-factor 176.08 indicates successful transmission of data with acceptable BER and Q-factor.

Figure 7 indicates the eye diagram of received signal in case of BPSK modulation scheme at a data rate and range of 10 Gbps and 5 km, respectively. The quality of the signal has deteriorated in this case as compared to Fig. 6 as distance as well as data rate has been increased in this observation.

The clear eye opening with a Q-factor 21.91 indicates successful transmission of signal with acceptable BER and Q-factor in this case.

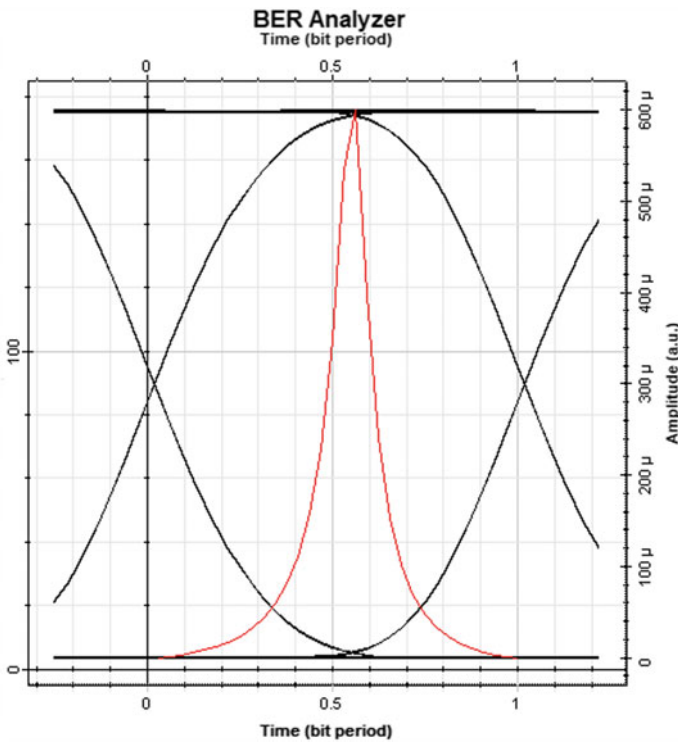


Fig. 6 Eye diagram of BPSK modulation format at 1 Gbps, with a transmission range of 1 km

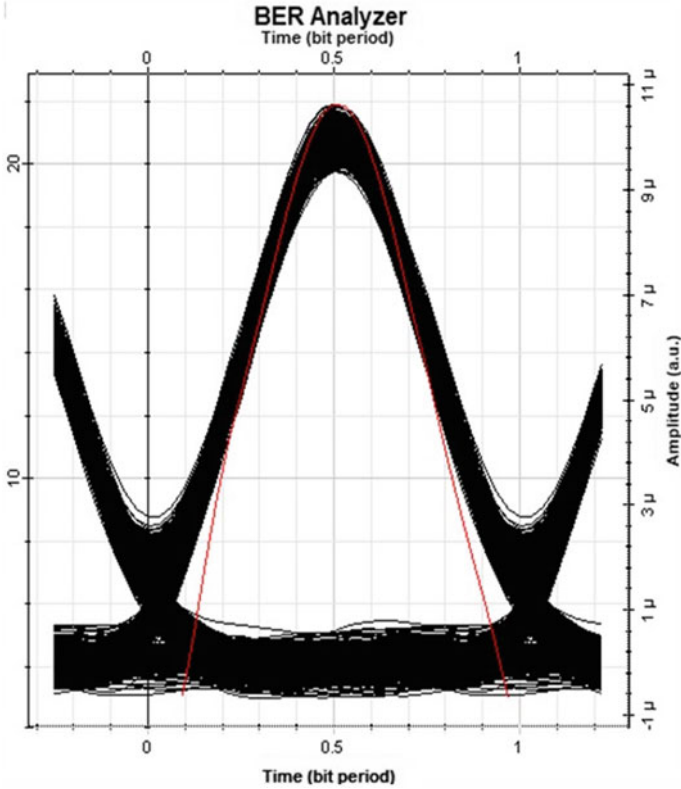


Fig. 7 Eye diagram of BPSK modulation format at 10 Gbps, with a transmission range of 5 km

4 Conclusion

As to conclude the paper, characterization of millimeter waves RoFSO link for 5G application has been analyzed. Five different modulation techniques have been studied to analyze the performance of the FSO system with specific visibility range, data rate, power, and beam divergence as system design parameters.

The Q-factor was evaluated and plotted for each of the modulation schemes, and it is observed that DPSK and BPSK are maximum and least efficient techniques in terms of quality of received signal. Q-factor observed is about 1400 at a data rate of 1 Gbps with DPSK technique. Yet, minimum is seen in case of BPSK which shows Q-factor 750 at 1 Gbps and about 300 at 3 Gbps. 8-QAM, 16-QAM, and QPSK modulation techniques performed better than BPSK in all cases.

The clear eye opening with a Q-factor 21.91 indicates successful transmission of signal with acceptable BER and Q-factor in case of BPSK modulation format at 10 Gbps, with a transmission range of 5 km.

References

1. Tabassum N, Franklin N, Arora D, Kaur S (2018) Performance analysis of free space optics link for different cloud conditions. In: 4th international conference on computing communication and automation (ICCCA), August 2018, pp 98–101
2. Kazaura K, Wakamori K, Matsumoto M, Higashino T, Tsukamoto K, Komaki S (2010) RoFSO: a universal platform for convergence of fiber and free-space optical communication networks. *IEEE Commun Mag* 48(2):130–137. <https://doi.org/10.1109/MCOM.2010.5402676>
3. Salehi J (1989) Code division multiple-access techniques in optical fiber networks. I. fundamental principles. *IEEE Trans Commun* 37(8):824–833, 45–89
4. Saito Y, Kishiyama Y, Benjebbour A, Nakamura T, Li A, Higuchi K (2013) Non-orthogonal multiple access (NOMA) for cellular future radio access. In: IEEE 77th vehicular technology conference (VTC Spring), June 2013, pp 1–5
5. Kizilirmak RC, Rowell CR, Uysal M (2015) Non-orthogonal multiple access (NOMA) for indoor visible light communications. In: 4th international workshop on optical wireless communications (IWOW), Sep 2015, pp 98–101
6. Bloom S, Korevaar E, Schuster J, Willebrand H (2003) Understanding the performance of free-space optics. *J Opt Network Conf* 2(6):105–196
7. Mishra B, Gajal L, Kaur S (2019) Analysing the performance of terrestrial FSO link for different internal parameters of the system. In: SPIN 6th International conference on signal processing and integrated networks, August 2019, pp 56–75
8. Manickam S, Kaur K, Chaudhary S (2018) Millimeter waves over free space optics system for 5G application. *J Opt Commun*, pp 28–56
9. Kaur S, Kakati A (2018) Analysis of free space optics link performance considering the effect of different weather conditions and modulation formats for terrestrial communication. *J Opt Commun*. <https://doi.org/10.1515/joc-2018-0010>
10. Adnan SA, Ali M, Ali A (2018) Characteristics of RF signal in free space optics (RoFSO) considering rain effect. *J Eng Appl Sci* 13(7):1644–1648, pp 14–86
11. Dang J, Zhang Z, Comparison of optical OFDM-IDMA and optical OFDMA for uplink visible light communication

Characterization of Millimeter Waves RoFSO Link Under the Effect of Rain



Manisha Samal, M. Vishnu Kartik, Sanya Arora, and Sanmukh Kaur

Abstract Free-space optics, being an optical wireless communication, uses the concept of light propagation in free space for data transmission between two points and is progressing as competitive and efficient alternative to optical fiber links or RF system. Being a favorable optical wireless communication technology, radio over free-space optics (RoFSO) has proved to be of great significance in various applications in telecommunication including 5G networks. High speed and data rate, large bandwidth, low consumption of energy, and unlicensed spectrum are some of the advantages of this technology. As the technology is based on an unguided medium, atmospheric weather affects the transmission of data through the FSO wireless communication. Varying climatic depletion factors including cloudiness, downpour, and haze conditions attenuate the transmitted signal and degrade its quality. Rain is considered as the significant cause of constriction causing degradation in the transmission of signal. This paper mainly focuses on the characterization of attenuation in a millimeter waves RoFSO link for 5G applications. Received signal quality has been evaluated in terms of Q factor and eye diagram patterns under different downpour conditions.

Keywords Weather conditions · RoFSO · Attenuation · Free-space optics · Q factor · Mm waves

M. Samal (✉) · M. Vishnu Kartik · S. Arora · S. Kaur
Amity School of Engineering and Technology, Amity University, Noida, India
e-mail: manishasamal2000@outlook.com

M. Vishnu Kartik
e-mail: vishnukartk123@gmail.com

S. Arora
e-mail: arorasanya40@gmail.com

S. Kaur
e-mail: skaur2@amity.edu

1 Introduction

Among the technological advancements taking place in the field of telecommunication, FSO, being an optical wireless communication, uses the concept of light propagation in free space for data transmission between two end points thereby becoming fair and efficient alternative to fiber optic links or radio frequency (RF) transmission systems [1–3]. Being a favorable optical wireless communication, radio over free-space optics (RoFSO) has been proved to be of great importance in various applications in telecommunication field for instance high speed, high bandwidth data link availability, and low consumption of energy. The RoFSO system basically comprises of radio over fiber and free-space optics system. The radio frequency along with the optical signal goes through modulation process in this RoFSO link. This optical wireless FSO technology gives numerous points of interest, which includes providing high data rate and bandwidth, low cost and power, immunization of radio frequency interference, and high security [4–6]. Be that as it may, atmospheric weather affects the transmission of data through the FSO wireless communication. Varying climatic depletion factors including cloudiness, downpour, and haze attenuate the transmitted signal quality. Rain is considered the significant cause of constriction causing degradation in the transmission of signal. Scintillation, absorption, and scattering are several other natural phenomena contributing to the degradation of FSO performance [7–9]. The variations in both the intensity and phase of the transmitted optical signal are a result of atmospheric turbulence which reduces the link performance [10]. RoFSO innovation expels RF spectrum overcrowding in wireless. RoFSO gives a wireless connection for the farthest mile association existing amid two end points in the region of troublesome physical connection for providing broadband services. This paper mainly focuses on the characterization of attenuation in a millimeter waves (MM waves) RoFSO link for 5G applications. Received signal quality has been evaluated in terms of maximum Q factor and eye diagram patterns under different downpour conditions.

2 System Design

The RoFSO framework is designed by using optisystem 17. This design comprises transmitter, FSO channel, and receiver. The target of the design is to generate modulated RF signal and the resulting RF signal and the optical carrier generated from CW laser is modulated and is then fed to FSO channel. The transmitter section comprising of user bit sequence generator, NRZ modulator, CW laser, Mach–Zehnder modulator, and DPSK sequence generator. The bit sequence generator generates the data rate at 10 Gb/s. Then, the duo-binary phase shift keying (DPSK) modulated scheme is utilized for extending the spectral efficiency of the framework. The signals coming from the DPSK modulator is then sent to M-ray pulse generator for converting into M-ray pulse and afterward is modulated using 60 GHz RF carriers [3]. The output RF

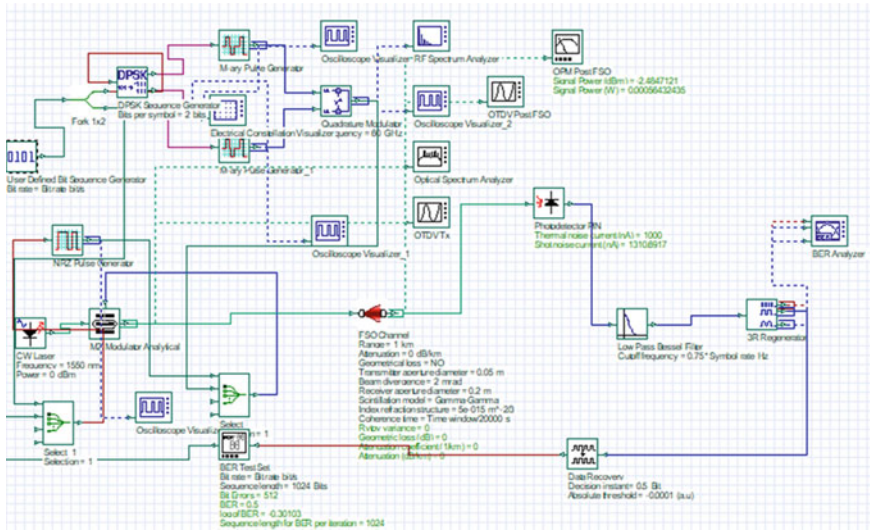


Fig. 1 A block diagram of RoFSO link

signal is modulated with optical carrier (laser), i.e., CW laser with frequency 193.1 THz by the Mach-Zehnder modulator (MZM) and is then fed to FSO channel under the downpour climate conditions. The laser diode has a wavelength of 1550 nm. For the FSO channel, the range is set up to 1 km as a default value. The signal transmits through FSO, transmitter power being 0 dBm and divergence angel being 2 mrad. The receiver comprises of a PIN photodetector with 1 A/W responsivity, and 10 nA dull is utilized for detection of the optical signal prior to transmission through the filter headed by the PIN photodetector. A low-pass Bessel filter cuts off the higher frequency components and thereby filters noises from the received signal. The bit error rate (BER) analyzer has been used to analyze the BER or Q factor of the received signal (Fig. 1; Table 1).

3 Rain Attenuation

The optical signal transmitted through atmosphere suffers from rain attenuation. Impact of rain is an essential factor that influences the optical communication link. Volume of rain drops being expansive causes refraction and reflection and thereby causes scattering. The precipitation expansion results in linearity expansion in attenuation. The normal size of raindrops increments with the precipitation, and this remains within the order of mm. Estimation of particular constrictions from the optical association with the rain rate R mm/hr by [11]

$$\alpha_{rain} = 1.076R^{0.67}, \text{ dB/km} \quad (1)$$

R being the rain rate.

Table 1 Simulations parameters

Parameter	Value
Bit rate (Gbps)	10
Range (km)	1
Wavelength (nm)	1550
Modulation Scheme	DPSK
Responsivity (A/W)	1
Input Power (dBm)	-10 to 0
Sequence length (bits)	512
Samples per bit	64
Samples rate (GHz)	3200
Beam divergence (mrad)	0.5-3.5
Dark current (nA)	10
Cut off frequency (GHz)	2.4

The attenuation for laser power in the weather is represented by the Beer's law as:

$$\tau_{(L)} = P_{\text{receive}}/P_{\text{total}} = e^{-\alpha L} \quad (2)$$

L is the link range provided in meters, $\tau_{(L)}$ stands for the transmittance at a distance L , P stands for the laser power (Watt) whereas α is the scattering coefficient (km^{-1}).

4 Simulation Results

The objective of the paper is to study characterization of MM waves RoFSO link under the effect of rain. The Q factor is calculated for different downpour or rain conditions (clear air, light rain, moderate rain, and heavy rain).

Figure 2 depicts the Q factor versus range plot in different downpour conditions. With the increase in range, the Q factor value reduces. Q factor is highest in clear air shown in brown curve. The Q factor for clear air condition is 1316, for light rain is 986, for moderate rain is 467, and for heavy rain is 241.7 at a range of 1000 m. Least Q factor is in heavy rain shown in blue curve. As seen in the figure, communication is not possible beyond the range of 2 km under heavy rain conditions.

Figure 3 depicts Q factor versus power plot in different downpour conditions. With the increase in power, there is increase in the value of Q factor. Q factor is highest in clear air condition at 0 dBm shown in brown curve. The Q factor for clear air condition is 1264, while 986 for light rain, 467 for moderate rain, and 241.7 for heavy rain at a power of 0 dBm. Least Q factor is for heavy rain shown in blue curve. In the case of heavy rain condition, transmitted signal power should be more than -3 dBm for acceptable value of BER at the receiver.

Fig. 2 Q factor versus range plot in different rain conditions

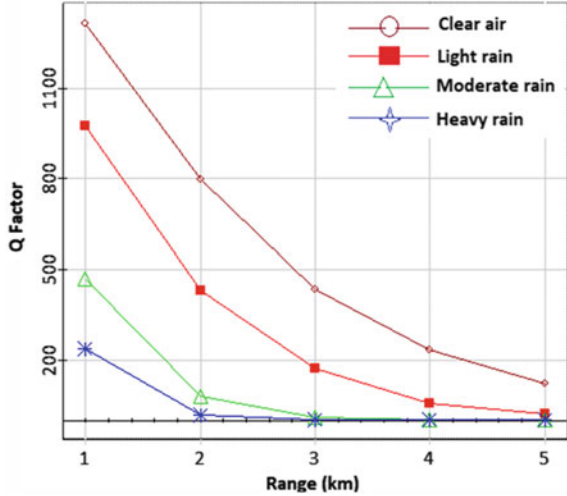


Fig. 3 Q factor versus power plot in different rain conditions

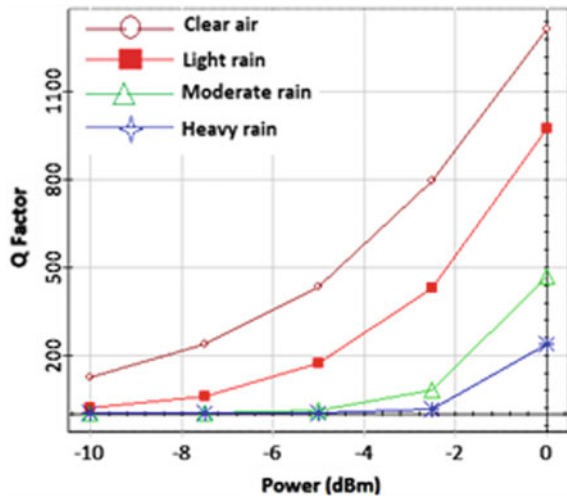


Figure 4 depicts Q factor versus bit rate plot in different downpour conditions. With the rise in data rate, Q factor reduces. Q factor is highest in clear air condition at 1 Gbps shown in brown curve. The Q factor for clear air condition is 1315, while 1015 for light rain, 464 for moderate rain, and 229 for heavy rain at data rate of 1 Gbps. Least Q factor is for heavy rain shown in blue curve.

Figure 5 depicts Q factor versus beam divergence plot in different downpour conditions. With the increase in beam divergence, Q factor reduces. Q factor is highest for clear air at 0.5 mrad shown in brown curve. The Q factor for clear air condition is 1329, while 1010 for light rain, 465 for moderate rain, and 232 for heavy rain at 0.5 mrad. Least Q factor is for heavy rain shown in blue curve.

Fig. 4 Q factor versus data rate plot in different rain conditions

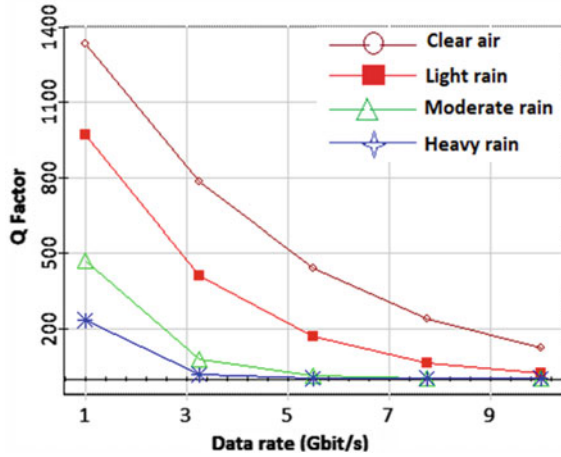


Fig. 5 Q factor versus beam divergence plot in different rain conditions

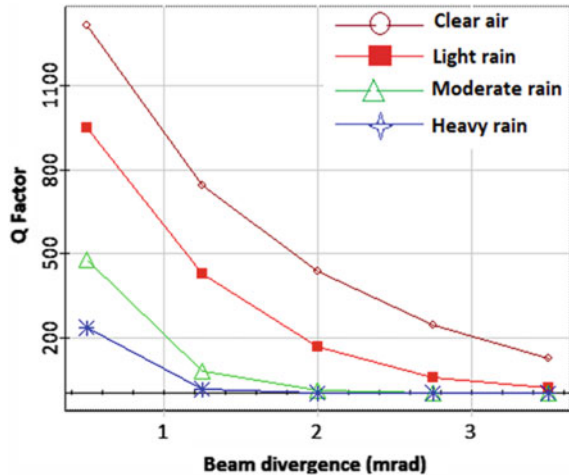


Figure 6 shows the eye diagram of received signal for a transmission of 10 Gbps data rate with RF of 60 GHz over 1Km of FSO length. The observed Q factor for heavy rain conditions at 1 km length is 25.05.

Figure 7 shows the eye diagram of received signal for a transmission of 10 Gbps data rate with RF of 60 GHz over 2.91 km of FSO length. The observed Q factor for heavy rain conditions at 2.91 km length is 3.68.

Table 2 represents the transmission of 10 Gbps signal by 60 GHz RF modulation through FSO. In clear air weather condition, maximum link length is 36.02 km with Q factor 6.29. In light rain weather condition, maximum link length is 7.61 km with Q factor 6.76. In moderate rain weather condition, maximum link length is 2.52 km with Q factor 6.36. In heavy rain weather condition, maximum link length is 1.66 km with Q factor 6.05.

Fig. 6 Eye diagram for heavy rain conditions at a range of 1 km and data rate of 10 Gbps

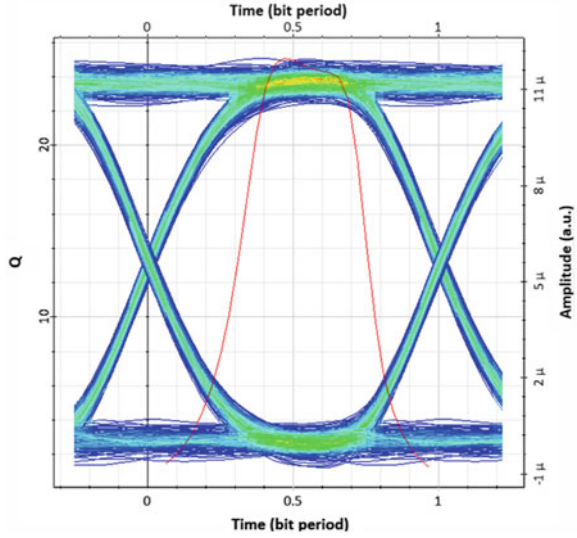
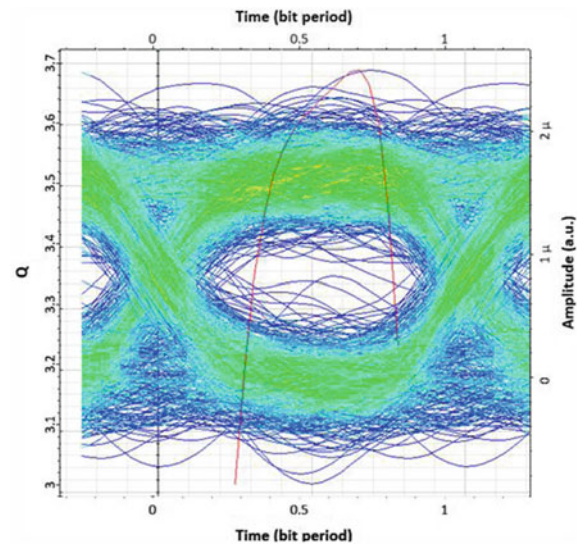


Fig. 7 Eye diagram for heavy rain conditions at a range of 2.91 km and data rate of 10 Gbps



5 Conclusion

In this paper, the simulation study of RoFSO link under the effect of rain has been performed. The MM waves RoFSO framework has been successfully designed and simulated for 5G applications. Received signal quality has been evaluated in terms of Q factor as a function of transmitted power, beam divergence, data rate, and transmission range under different downpour conditions. Our examinations demonstrate

Table 2 10 Gbps transmission at max link length

Weather condition	α db/Km	C_n^{-z}	RF (GHz)	Max link (km)	BER	Q factor
Clear air	0.43	5×10^{-14}	60	36.02	$1.5528e^{-010}$	6.29
Light rain	1.98	0.6×10^{-14}	60	7.61	$6.74797e^{-012}$	6.76
Moderate rain	5.34	0.5×10^{-14}	60	2.52	$7.82742e^{-015}$	6.36
Heavy rain	9.28	0.4×10^{-14}	60	1.66	$7.15249e^{139}$	6.05

that the signal experiences degradation with the increase in the value of attenuation in heavy rain as well as with FSO range. In the case of heavy rain condition, transmitted signal power should be more than -3 dBm for acceptable value of BER at the receiver. The maximum link length for heavy, moderate, and light rain conditions comes out to be 1.66, 2.52, and 7.61 km, respectively.

References

1. Shaker FK, Ali MAA (2018) Multi-beam free-space optical link to mitigation of rain attenuation. *J Opt Commun*
2. Kaur S (2019) Analysis of inter-satellite free-space optical link performance considering different system parameters. *Opto-Electron Rev* 27:10–13. <https://doi.org/10.1016/j.opelre.2018.11.002>
3. Kaur S, Kakati A (2018) Analysis of free space optics link performance considering the effect of different weather conditions and modulation formats for terrestrial communication. *J Opt Commun*. <https://doi.org/10.1515/joc-2018-0010>
4. Kim J, Sung M, Cho S-H, Won Y-J, Lim B-C, Pyun S-Y (2020) MIMO-supporting radio-over-fiber system and its application in mmwave based indoor 5G mobile network. *IEEE J Lightwave Technol* 38(1)
5. Grabner M, Kvicera V, Experimental study of atmospheric visibility and opticalwave attenuation for free space optics communication. Czech Science Foundation, ProjectNo. 102/08/0851
6. Grover M, Singh P, Kaur P, Madhu C (2017) Multibeam WDM-FSO system: an optimum solution for clear and hazy weather conditions. *Wireless Pers Commun* 97(4):5783–5795
7. Bohata J, Komanec M, Spáčil J, Ghassemlooy Z, Zvánovec S, Slavík R (2018) 24–26 GHz radio-over-fiber and free-space optics for fifth generation systems. *Opt Lett* 43:1035–1038
8. Anuranjana SK, Goyal R (2019) Analysis of terrestrial FSO link performance considering different fog conditions and internal parameters of the system. In: 2019 6th international conference on signal processing and integrated networks (SPIN), Noida, India, 2019, pp 552–557. <https://doi.org/10.1109/SPIN.2019.8711577>
9. Adnam S, Ali MAA (2018) Characterisation of RF signal in free space optics (RoFSO) considering rain effect. *J Eng Appl Sci* 13(7):1644–1648, ISSN: 1816-949X
10. Kaur H, Soni G (2015) Performance analysis of free space optical communication link using different modulation and wavelength. *J Sci Res Reports* 6(3)
11. Rahman AK, Julai N, Rashidi CBM, Zamhari N, Sahari S, Binti HM, Nur AA, Sharip MRM (2019) Impact of rain weather over free space optic communication transmission. *Indonesian J Electr Eng Comput Sci* 14(1):303–310. <https://doi.org/10.11591/ijeecs.v14.i1.pp303-310>

A Proficient Mathematical System (PMS) Design with Improvised Cross Layer Scheme for MANET QoS Support



Asha

Abstract A group of mobile nodes establishing a temporary network in a dynamic fashion with no infrastructural usage similar to access points in wireless mode or topology of network forms a MANET. Here, mobile nodes can work in an independent fashion causing recurrent change in configuration of the network. The provided service should be of high quality in the multimedia communications nowadays. But MANET faces many challenges in providing QoS when compared with wired communications because of the node's independent nature, restricted bandwidth, less battery power, impulsive mobility of nodes, no common authority to coordinate them, etc. Here, we discuss this issue of providing QoS in MANET with multimedia application support. We provide a collective design for cumulative performance of networking environment. Our suggested methodology contains three major categories which includes modeling of channels, queuing modeling method and transmission of data based on threshold in physical layer to bring QoS in MANET. The first approach is channel modeling where the provided data about channel like loss of path, channel fading's probability density function, path gain and path gain's probability density function are collected. The second phase will be selection model for threshold which is implemented to calculate the video packet's threshold value. The channel will be selected on threshold value basis through which transmission of video packet will happen. The final phase will be queuing model where MANET's communication delay will be greatly reduced. Here, we started our paper with proper introduction by providing the significant requirements of QoS improvisation (less overhead, fairness and robust architecture), we used the proficient mathematical system (PMS) pattern as well as hybrid cross layer methods for QoS conscious routing strategies, interference modeling, buffer modeling, cross layer architecture, etc., we have experimentally verified our concept based on throughput performance, efficiency of bandwidth, delivery rate of the packets, and a comparison is done with the existing approaches available in the market with mentioning of all possible graphical representations with readings and measurements, and the results show that our suggested model will be able to perform effectively in comparison with current approaches.

Asha (✉)

Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India
e-mail: asha.cse.ait@gmail.com

Keywords QoS · Cross-layer · Throughput · MANET · Modeling

1 Introduction

Communication in wireless mode has a rapid growth in recent years, and it is developing still in the communication field. There are two categories in the wireless networking environment like ad hoc networking methodology and cellular networking methodology. Networking methodology in cellular wireless mode network is mainly used in the cellular phones like GSM. In recent days, the usage of multimedia applications plays a vital role in MANET architecture. MANET is nothing but a wireless network with independent mobile nodes, and connection is established among mobile nodes with the aid of wireless link. MANET mobile nodes are independent in nature, and the communication between them is established with no infrastructural support or topology of network support. Every node in MANET can act as both router and a host. Due to demand in the multimedia communications, the QoS demand also increased. QoS in MANET is bit difficult to achieve due to certain factors like less efficiency, congestion problem of nodes, less bandwidth, etc., and all these problems are mainly caused due to the dynamic behavior of the mobile nodes.

The term QoS is brought up by means of data transfer assurance from source to destination within the allotted time slot. The significant requirements for improving QoS are less overhead, robustness in architecture and system fairness. There are certain measurement metrics utilized to measure QoS like delay of transmissions, loss of packets, etc. When these metrics are satisfied, then MANET can provide data transmission with high QoS support. Our methodology will satisfy all these metrics from source to destination for better QoS in MANET. Previously, MANET was mainly utilized for applications used in disaster management and military fields. There are certain significant features in MANET like architecture with autonomous support, mechanism with distributive behavior, topology of networks with re-configuration support, etc. In our paper, we present a unique approach for enhancing MANET performance to provide better QoS. The main components are channel modeling—in this phase, computation of loss of paths, gain of paths, probability density function of both path loss and path gain are computed, data transmission based on threshold value—video packet transmission is done based on threshold value for channel selection. A threshold value is set already for communication, and corresponding probability is calculated to proceed further either to discard the video packet or to transmit them, queuing model—this model is mainly developed for computing encoding rate of video on packet length basis therefore allowing the calculations of upper bound limits for transmission of video data. Here, the next section is followed by literature survey containing all possible research works related to the addressed issue, followed by system model with a detailed explanation of our approach will all modeling phases, followed by experimental analysis section

providing the simulation results of our suggested concept in comparison with the existing approach, and finally, our paper is concluded with a proper conclusion.

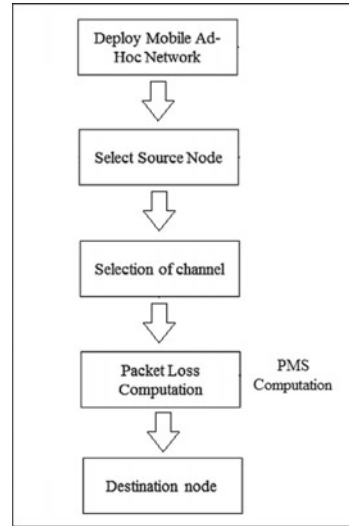
2 Literature Survey

We have undergone various research works regarding MANET's challenge in providing QoS, and the most recent technologies supporting QoS enhancement in MANET are mentioned further. The authors in [1] perform a cooperative scheme design employing superposition coding at D2D and cellular transmitters. This scheme can maximize the D2D's data rate with no deterioration in cellular links performance. They also derive a power payment contract optimally for all the cellular links at two scenarios like D2D link's private data is both continuous and in discrete form. Their concept is experimentally verified both analytically and numerically for confirming their efficiency. The authors demonstrate the interference impact on nodes' effective range of communication in [2]. Here, they suggest two alternative forwarding protocols, PCN and lifetime of network with PCN extension. With the help of MATLAB and NS2 simulation network, performance is verified statically via numerical results. This paper suggests a new solution to improve QAR performance and AC protocol's performance by means of shadowing factor, mobility factor and SINR [3]. The reliability of confirmed throughput services can be noticeably improved by proactive maintenance of backup routes by adjusting transmission rates, by means of active sessions and temporary routing on low SINR links.

3 Proposed Methodology

In this section, we discuss about the overall architecture of our suggested proposal for enhancing QoS in MANET. It normally contains N number of communication sessions mainly used for transmissions of video packets. Such architecture can be used for vehicular communication, device communication networks, etc. For every communication session, source and destination can be easily identified. Channels can be shared among various sessions causing interference from every source. In Fig. 1, our complete architecture of QoS model with cross layer design is depicted via architectural diagram. There are three different layers in our model such as PHY layer responsible for channel modeling, application layer responsible for distortion and MAC layer responsible for buffer modeling. At initial phase, a new network is deployed with all the essential channels and network parameters. A source node is assigned to it which is responsible for data packet transfer along with an essential time slot and channel parameters required for transmitting video packets. Next phase comes with a physical layer where channel modeling is utilized which helps in achieving required channels for transmitting video packets. The next phase will be MAC layer where interference modeling and buffer modeling are applied. The

Fig. 1 Overall architecture



final phase will be application layer where video distortion model is considered for computing packet loss and reconstruction of video packets.

Video transmission time is converted into numerous time slots, and during communication phase, a time slot is allocated for communication where appropriate channel is selected from the set of frequency channels. With the help of opted channel for making the transmission, threshold value and channel gain value are taken into account. When channel gain value is higher than the threshold value, then packet transmission happens through the opted channel, and when channel gain value is lesser compared with threshold value, then packets are placed in the buffer queue. When video packets are lost, then there will be an error in the transmission link, and deadline of packet delay might be expired as well. Thus, the overall performance of the network can be improvised by altering encoding rate and threshold during transmission in dynamic fashion for every layer. For improvising the performance, we utilize cross layer approach by means of below three modeling techniques.

Channel modeling in Wireless mode Here, probability during video packet transmission is computed by applying channel modeling at PHY layer in the network. For every session, path loss, channel gain, probability density function for path gain and channel fading are computed. This phase helps in data transmission on threshold basis.

Transmission of data on threshold value selection basis In this phase, channel modeling is considered for calculating probability of video packet transmission. For every user, an assumption is made that channel state data is known including channel gain and channel distance. For every session, threshold value is computed, and based on its value, the selected channel video packet transmission is proceeded further.

Interference modeling and buffer modeling Interference modeling is placed at the PHY layer with the help of ad hoc networks with random deployment. Accurate data is approximated of cumulative interference by means of applying ‘log normal distribution’. Buffer modeling is placed at MAC layer for improving network’s performance by increasing PSNR value and throughput value. Based on the threshold value computed earlier, certain packets are dropped and placed in the buffer model for successful transmission of video packets.

Proficient mathematical system (PMS) With this paper, we used the proficient mathematical system (PMS) pattern as well as hybrid cross layer methods for QoS conscious routing strategies. Functionality of suggested scheme is computed of terminology of throughput, bandwidth effectiveness as well as packet delivery rate (PDR) compared with existing models. Experimental analysis was satisfied with PDR, overhead and energy. Because of fast development in communication that is wireless, multimedia interaction of wireless community attracted scientists. MANET has demonstrated the significance of its. MANETs are the group of different communication nodes without using any kind of fixed network and infrastructure topology. Inside MANET, movable nodes may on their own stay in just about any course that causes repeated alteration of community setup. Utilization of multimedia interaction is growing day-by-day that calls for excessive quality-of-service (QoS) for conclusion operator knowledge. MANET is suffering from problems to offer QoS including powerful dynamics of mobile node, routing, limited bandwidth, as well as contention channel. With these efforts, we deal with the problem in QoS to come down with MANET for multimedia uses. QoS provisioning inside MANETs is an extremely demanding issue in comparison with wired IP networks. This is due to unforeseen node mobility, so the style of a reliable and efficient routing pattern offering QoS assistance for this kind of apps is a tough job. Using Eq. (1), we can use this mathematical formula for overcoming the overhead and energy, and a and b are the number of packets using binomial formula, and to sum the values, we can use the Eq. (2).

$$(a + b)^n = \sum_{k=0}^n a^k b^{n-k} \quad (1)$$

$$(a + b)^n = a + \frac{nb}{1!} + \frac{n(n-a)b^2}{2!} \quad (2)$$

4 Experimental Results

We have experimentally verified our suggested concept through simulation and analysis. We have utilized all the essential network simulation parameters. By means of calculating PSNR value, packet drop rate value, average throughput value, etc., our overall system performance is attained. We have utilized MATLAB tool with ad hoc

network of single hop. We have computed throughput per slot value in Fig. 2 where total count of slots is '50' with three channels and channel bandwidth value is 1. The graph is plotted between slots and throughput value per slot in terms of bits. Throughput value is calculated by means of simulation time and probability variation transition. Here, channel state remains constant at 0.8 probability value in case 1. Channel state equally changes in case 2, and in case 3, channel state has no change at probability value 0.5.

In Fig. 3, average throughput performance is depicted in the form of graph which is plotted between threshold selection probability value and average throughput value. From the graph, it is clear that proposed model has better results in comparison with existing model. Our system's average performance value of '46.91%' is consistent in all possible threshold values.

In Fig. 4, bandwidth efficiency performance is depicted in the form of graph which is plotted between threshold selection value starting from '0' to '0.55' and bandwidth efficiency value starting from '0' to '0.8'. It is very clear that the bandwidth efficiency performance of our proposed model is constant at 0.7 bandwidth efficiency and it

Fig. 2 Throughput per slot

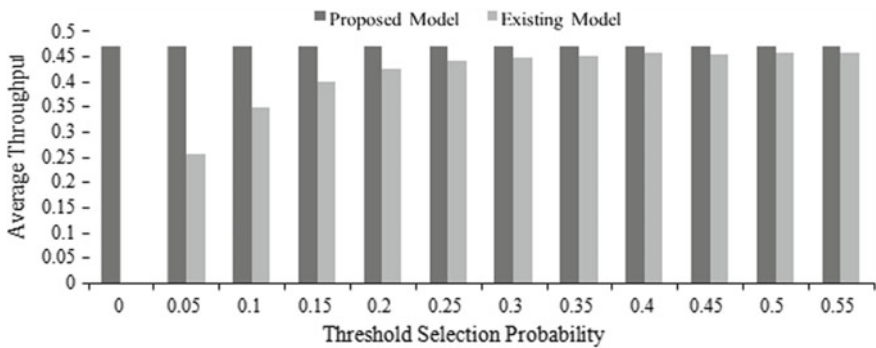
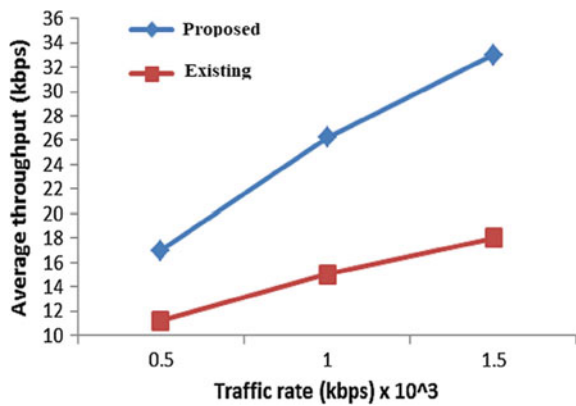


Fig. 3 Average throughput performance

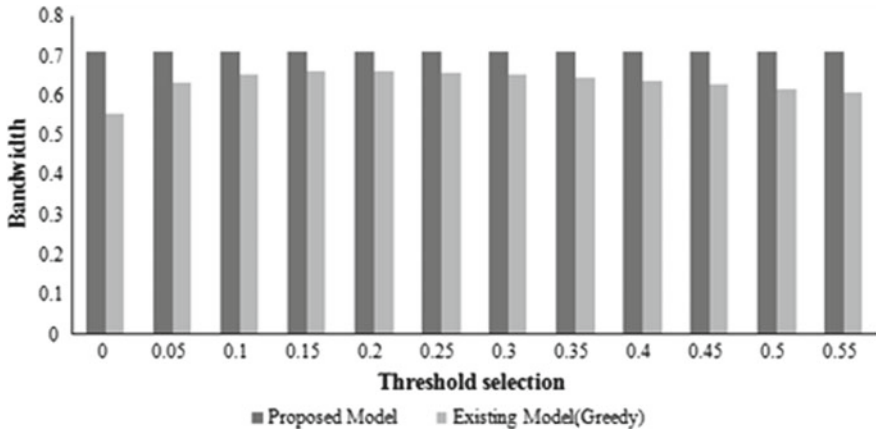


Fig. 4 Bandwidth Performance

is always higher compared to existing model. The average bandwidth percentage is ‘11.06%’ for our model.

In Fig. 5, [16] and [17] the graph is plotted between threshold value starting from ‘0’ to ‘0.25’ and delay parameter measured in milliseconds starting from ‘0.47’ to ‘0.56’. Total delay performance is depicted between existing and proposed model. By altering the threshold parameter, performance is analyzed. The average delay percentage is ‘2.10%’. From the graph, it is very clear that proposed model has less delay when compared with existing model. Therefore, proposed model can opt for efficient channel for transmission of video packets.

In Fig. 6, graph is depicted between throughput value and packet arrival date. Energy and packet loss rate value range from ‘0’ to ‘40’, and traffic rate value ranges from ‘0’ to ‘1.5’, respectively. From the graph, it is clear that when the arrival rate of the packet is low, then the throughput performance of the proposed system is also low. During the increase in simulation time, there is an increase in packet arrival rate as well and so the overall performance at its best since all the packets are received by

Fig. 5 Delay performance comparison

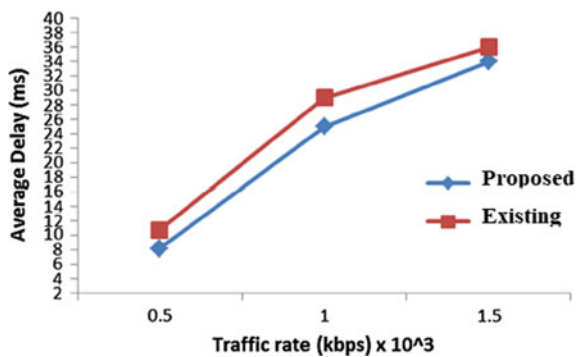
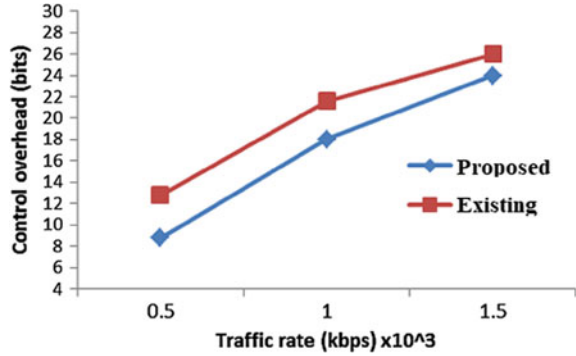


Fig. 6 Energy performance and packet loss rate



the receiver end. The system's performance is improved by '81.02%' in throughput basis.

5 Conclusion

One of the prominent research areas in wireless networking environment is multimedia communication. But there are few challenging issues here related to less utilization of resources, parameters related to delay, maintenance of data traffic and QoS. We have taken QoS into account in our paper for MANET in multimedia communication. Our suggested scheme contains three kinds of modeling named channel modeling, queuing modeling and video packet transmission based on threshold value. In channel modeling, path gain and path loss parameter are computed for selection of specific channel, in channel selection for transmission is done based on the computed threshold value in the second phase, and for addressing delay issues during video packet transmission, queuing model is considered. Our main aim is to increase the network's throughput. We have mentioned literature survey section will all our research works, system model section where our suggested model is briefly explained and experimental results section—here we have graphically represented the results for average throughput, bandwidth efficiency and delay parameters. Finally, paper is concluded with this conclusion section.

References

1. Ma C, Li Y, Yu H, Gan X, Wang X, Ren Y, Xu JJ (2016) Cooperative spectrum sharing in D2D-enabled cellular networks. *IEEE Trans Commun* 64(10):4394–4408
2. Panigrahi B, Sharma A, De S (2012) Interference aware power controlled forwarding for lifetime maximisation of wireless ad hoc networks. *IET Wireless Sens Syst* 2(1):22–30
3. Hanzo L II, Tafazolli R (2010) QoS-aware routing and admission control in shadow-fading environments for multirate MANETs. *IEEE Trans Mob Comput* 10(5):622–637

4. Li X, Liu T, Liu Y, Tang Y (2014) Optimized multicast routing algorithm based on tree structure in MANETs. *China Commun* 11(2):90–99
5. Zhang XM, Zhang Y, Yan F, Vasilakos AV (2014) Interference-based topology control algorithm for delay-constrained mobile ad hoc networks. *IEEE Trans Mob Comput* 14(4):742–754
6. Shijie J, Changqiao X, Muntean GM, Jianfeng G, Hongke Z (2013) Cross-layer and one-hop neighbour-assisted video sharing solution in mobile ad hoc networks. *China Commun* 10(6):111–126
7. Nivetha SK, Asokan R, Senthilkumaran N (2013, July) A swarm-based hybrid routing protocol to support multiple Quality of Service (QoS) metrics in mobile ad hoc networks. In: 2013 fourth international conference on computing, communications and networking technologies (ICCCNT). IEEE, New York, pp 1–8
8. Abrougui K, Boukerche A, Pazzi RW, Almulla M (2013) A scalable bandwidth-efficient hybrid adaptive service discovery protocol for vehicular networks with infrastructure support. *IEEE Trans Mob Comput* 13(7):1424–1442
9. Shin B, Han SY, Lee D, Yoon W (2013, October) Application-awareness support for QoS-aware routing protocols in wireless ad-hoc networks. In: 2013 fourth international conference on the network of the future (NoF). IEEE, New York, pp 1–5
10. Scutari G, Palomar DP, Facchinei F, Pang JS (2012) Monotone games for cognitive radio systems. In: *Distributed decision making and control*. Springer, London, pp 83–112
11. Baguda YS, Faisal N, Rashid RA, Yusof SK, Syed SH (2012, April) Threshold-based cross layer design for video streaming over lossy channels. In: *International conference on networked digital technologies*. Springer, Berlin, Heidelberg, pp 378–389
12. Wu Y, Kumar S, Hu F, Zhu Y, Matyjas JD (2014) Cross-layer forward error correction scheme using raptor and RCPC codes for prioritized video transmission over wireless channels. *IEEE Trans Circuits Syst Video Technol* 24(6):1047–1060
13. Hu D, Mao S (2012, March) Cooperative relay with interference alignment for video over cognitive radio networks. In: 2012 proceedings IEEE INFOCOM. IEEE, New York, pp 2014–2022
14. Kuo JL, Shih CH, Ho CY, Chen YC (2013) A cross-layer approach for real-time multimedia streaming on wireless peer-to-peer ad hoc network. *Ad Hoc Netw* 11(1):339–354
15. Guan Z, Melodia T, Yuan D (2011, June) Optimizing cooperative video streaming in wireless networks. In: 2011 8th annual IEEE communications society conference on sensor, mesh and Ad Hoc communications and networks. IEEE, New York, pp 503–511
16. Asha Mahadevan G (2017) An adaptive cross-layer architecture to optimize QoS provisioning in MANET. *Indonesian J Electr Eng Comput Sci* 5(1):16–25
17. Asha Mahadevan G (2018) A combined scheme of video packet transmission to improve cross layer to support QoS for MANET. *Alexandria Eng J* 57(3):1501–1508

VLSI Design and Its Application

Concurrent Architecture-Based Fast Fourier Transform on FPGA IP Cores



Varun Maheshwari, Mahendra Singh, Pooja Pandit, Reshu Aggarwal, Arun Navputra, Shubham Sharma, and Atul Kumar Shukla

Abstract Harmonics in signals play an important role in power system. Due to harmonics, power system suffers several problems such as false tripping, communication interference, voltage fluctuations, frequency fluctuation, and waveform fluctuation. The fundamental frequency component of the signal is to be extracted for the smooth functioning of the system. Fast Fourier transform (FFT) is one of the methods to remove harmonics and extraction of the fundamental frequency components. Initially, FFT was implemented on solid state devices, and later, microprocessors/microcontrollers comes into the picture due to programming flexibility. But this serial implementation enhances time delay. This paper describes the implementation of FFT on FPGA by configuring intellectual property (IP) cores with a parallel sense–process–communicate module resulting in reduced time delay. The design is demonstrated on GENESYS VIRTEX 5 FPGA board with help of VHDL. The fundamental frequency component rated at 50 Hz is communicated directly to the hyper-terminal with the help of the RS232

Keywords Communication · Concurrent architecture · Fast Fourier transform · Field programmable gate arrays · Intellectual property

1 Introduction

The fault analysis of current in power systems often involves different harmonics and transient direct current components that may result in false tripping of the relay, noise, skin effect, overheating, overloading of neutrals, etc. For the proper functioning of the system, the decision to trip should be based on the fundamental

V. Maheshwari (✉) · P. Pandit · R. Aggarwal · A. Navputra · S. Sharma · A. K. Shukla
Electronics & Communication Engineering Department, Faculty of Engineering & Technology,
Agra College, Agra, India
e-mail: varun_agr@yahoo.com

M. Singh
Vocational Training & Skill Development, Govt. Industrial Training Institute Balkeshwar, Agra,
UP, India

component of the current and voltage of power frequency. Discrete Fourier transform (DFT) is a well-known and proven technique for isolating the fundamental frequency component in a sampled data system and has been utilized in this paper. DFT of the voltage and current samples have been calculated with the FFT algorithm. This paper presents the implementation of the FFT algorithm on the FPGA chip. The FFT filter is implemented for taking out fundamental components and removing harmonics from the fault current to enhance the accuracy and reliability of the system. Initially, researchers implemented solid state devices-based FFT afterwards microprocessor/microcontroller-based FFT is implemented due to programming flexibility. Microprocessor based FFT implementation for Frequency Division Multiplexing (FDM) is done [1]. The proposed design is implemented on a DSP TMS320C30 processor. A 96-point FFT is used for 4 kHz FDM channel. Tabular-based results are presented in the paper. A low power 128-point FFT processor implementation is on VLSI-oriented FFT presented in [2]. In the paper, 64-bit point 128 point radix $2/4/8$ pipelined is presented. The proposed control logic is designed on VHDL. Mathematical-based results are presented in the paper. An FFT processor for rebuilding Fourier transform holograms with help of a defective area scanner is also presented in [3]. Waveform-based results are shown in the paper. FFT implementation in the advanced measurement system is presented in [4]. Three types of implementations are presented in the paper. Criteria for comparison of different alternatives are also given in the paper. A performance study of 16-bit microcomputer implemented FFT algorithm. FFT algorithm for Radix-2 and Radix-4 is done on 16-bit Intel SBC 86/12 [5]. A interface to Intel microcomputer development system MDS-225. FFT algorithm is programmed on Intel ASM 86 program. Test results in the tabular and mathematical form are presented in the paper. A 16-bit low energy asynchronous 128-point FFT/Inverse FFT (IFFT) processor implementing hearing aid application is presented in [6]. FFT/IFFT processor is comprised of a memory module, shifter, multiplier, adder and FFT/IFFT controller. Design is synthesized on Verilog. Graphical test results are reported in the paper. Implementation of an FFT-based simulation algorithm for a fully programmable cochlear prosthesis is presented in [7]. 16-point to 256-point FFT algorithms are available in system memory. The sound spectrum obtained by the FFT algorithms. In the proposed work, eight channel cochlear implementation is used. FFT implementations with fused floating-point operations are presented in [8]. FFT processor used butterfly operation consists of multiplier, adder and subtractor of valued data. Numerical results are presented in the paper. A sequential parallel architecture for Radix-2 FFT is presented in [9]. The design is implemented on VHDL. The design is implemented with dual port memories and lookup tables. Reference [10] presents extraction of the fundamental frequency component of the current signal with the help of FFT. The extracted signal is used for the implementation of overcurrent relay. All these reports demonstrate the implementation of FFT and its application in medical electronics, signal processing and power protection in power systems. Microprocessor/microcontroller-based FFT implementation is sequential in nature. In this paper, we propose a pipelined-based FFT implementation with help of IP cores.

Table 1 Comparison of recent publications

S. No.	Details of work reported in publications	Proposed work in our paper
1	This paper presents the implementation of all-phase FFT for phasor measurement unit (PMU) signal. The simulation-based results are reported on ARM9 LPC3250 microprocessor [11]	Microprocessor-based implementation is sequential in nature, results in slow in speed. In this paper, the concurrent architecture of FFT is reported thereby efficient and fast
2	In this paper, the FFT is implemented with a Radix-4 butterfly model, and test results are compared with MATLAB. The optimization of the IP cores is not presented [12]	In this paper, optimized FFT IP cores are used, and it can be optimized with speed and area. High speed optimization results are reported in paper

2 Fast Fourier Transform

The process of decomposition of large DFTs is known as FFT algorithms [13]. The fundamental frequency component of the signals can be extracted from FFT to avoid false tripping. In this work, FFT is implemented with a Radix-2 FFT algorithm [14]. Handling of floating-point numbers is difficult in VHDL, and therefore, the FFT is implemented with different floating-point cores and FFT IP cores.

2.1 Floating-Point Cores

Different floating-point cores viz. conversion of float to fix conversion and vice versa multiply and divide issued for mathematical calculations.

2.2 FFT IP Core

FFT IP core is used for the extraction of fundamental frequency component of signal. FFT core loads data in 2's complement form and in rectangular form. Different transform sizes can be used for implementing FFT. Scaling schedule is used for implementing FFT so that loading and unloading data widths remain the same [14].

3 Implementation Design

The design implementation of the proposed architecture is shown in Fig. 1. The input signal is given to rectifier circuit, the rectified signal is given to a 12-bit

Fig. 1 Design flow of the proposed work

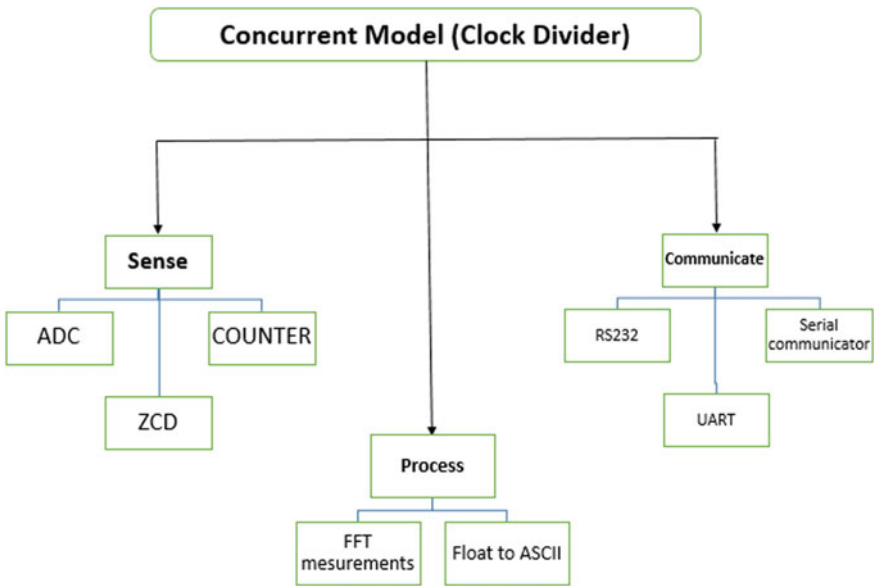
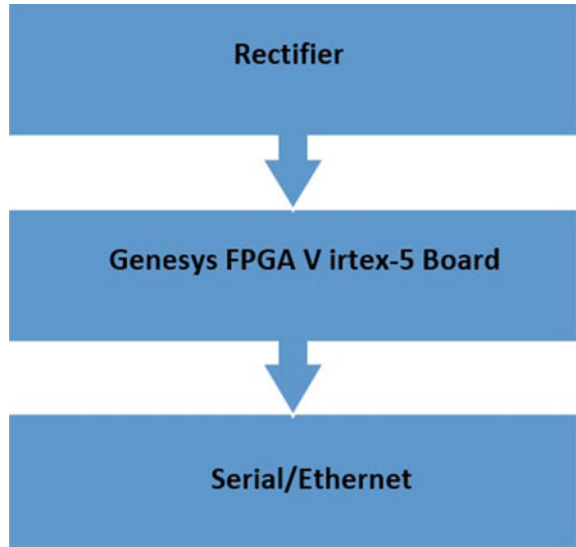


Fig. 2 Block diagram of the clock divider circuit in the proposed work

analog to digital converter (ADC), which is connected to Genesys Xilinx Virtex-5 board that works on concurrent architecture, and finally, results are displayed on serial or ethernet. The pipeline architecture is designed for computation of FFT as

Table 2 Different time delays for various transform size

S. No.	Delay	Transform size
1	186	128
2	93	256

Table 3 Different clocks of pipelined modules for various transform size

Transform size	Clock		
	Sense (kHz)	Process (MHz)	Communicate (kHz)
128	268.82	100	115.2
256	537.63	100	115.2

given in Fig. 2. Sense, process and communicate modules are designed in pipelined architecture.

3.1 Concurrent Model

Clock divider circuit is responsible for generating clocks for different modules. Different delays are shown in Table 1 for different transform size. This is also responsible for exchanging signals between the various modules. Different type of transform size can be taken depending on the system requirement. In this work, transform size of 128 and 256 is taken (Table 2).

Different clock frequencies are generated with help of clock divider circuit as shown in Table 3.

3.2 Sense Module

Voltage signal is provided to Xilinx Virtex-5 board with help of rectifier and ADC. Rectified voltage signal is given to one channel of 12 bit Analog 7476A ADC [15]. Analog 7476A is dual channel ADC. 128 samples are collected in half power cycle of voltage signal for transform size of 128 and 256. 128 or 256 samples are collected in half power of voltage signal. After collecting samples, it is given to process module for FFT. Samples are collected from zero crossing of the voltage signal.

3.3 Process Module

Process module receives signals from sense module and given in 2's complement form in FFT IP core. FFT is computed in Radix-2 Burst Mode after collecting 128 or

256 samples depending on transform size. After loading 128 samples if transform size is 128 or 256 samples if transform size is 256, FFT is computed. During unloading, magnitude of signal is computed, and index is computed where find the maximum magnitude. Fundamental frequency component is calculated as given in Eq. (1).

$$Ff = \frac{Sf}{2 * T} * I \quad (1)$$

$$Sf = \frac{\text{System Clock}}{2 * d * n} \quad (2)$$

where

Ff —fundamental frequency, Sf —sampling frequency, I —index

n —no. of clock cycles = 21.

d —186 for transform size 128.

93 for transform size 256.

T = Transform Size.

After calculating fundamental frequency, it is converted to ASCII and forwarded to communication module for displaying it to hyper-terminal.

3.4 Communicate Module

Fundamental frequency is computed as given in Eq. (1) and converted to ASCII with programming of float to ASCII conversion module. Fundamental frequency component rated at 50 Hz is communicated to hyper-terminal with UART interface, and further, it is converted to IP interface with help of MOXA serial to ethernet interface. UART is configured at 38,400 baud rates with the help of clock divider circuit. Different baud rates can be achieved with clock divider circuit.

4 Test Results

Fundamental frequency is computed after FFT as given in Eq. 1 and communicated directly to hyper-terminal for serial communication as shown in Fig. 3. All other harmonics are removed, and index for fundamental frequency rated at 50 Hz is computed, thereby avoiding noise signals. FFT can be computed for different transform size. In this work, FFT is computed for transform size 128 and 256. After collecting required samples from zero crossing in sense module, these samples are communicated to process module for FFT computation, while sense module starts collecting samples for next data window without any break in sampling. The time delay for various modules is given in Table 4.

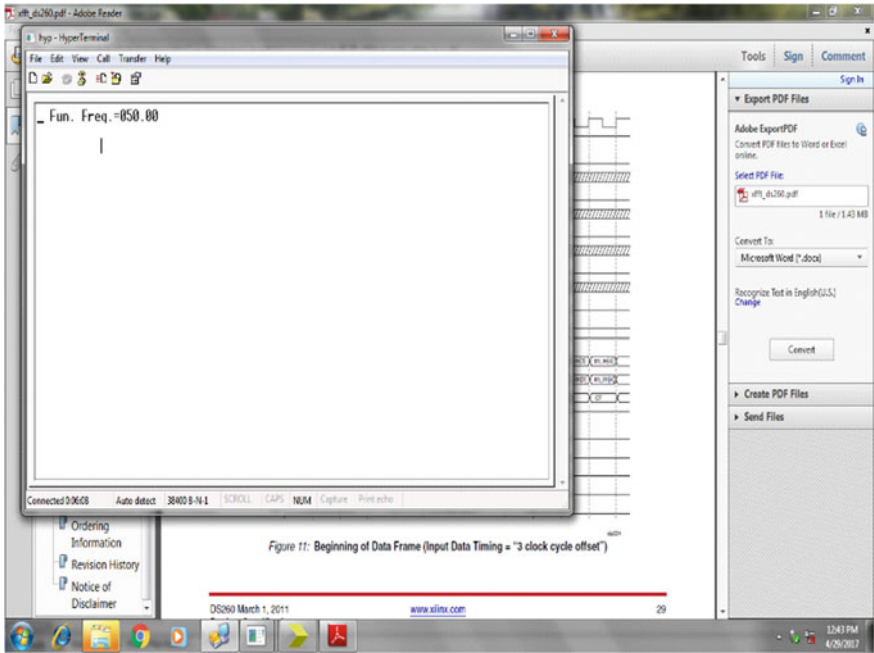


Fig. 3 Fundamental frequency rated at 50 Hz is communicated directly to hyper-terminal

Table 4 Time delay in different pipelines

	Pipelines		
	Sense (ms)	Process (μsec)	Communicate (ms)
Time delay (transform size 128)	9.99	102.2	7.8
Time delay (transform size 256)	9.99	110.7	7.8

Comparative time delay for sequential implementation viz. microprocessor/microcontroller and concurrent architecture viz. FPGA is shown in Table 5 and clearly shows total time reduced in FPGA-based implementation.

Table 5 Comparative time delay in sequential and concurrent architecture when transform size 128

Architecture	Time delay (ms)
Sequential (microprocessor/microcontroller)	17.8022
Concurrent (FPGA)	9.99
Total reduced time	7.8122

Table 4 shows sense and communicate modules consume delay in msec, whereas process module delay is in μ sec. Process module runs faster than sense and communicate module so we can add additional functionality in process module without sacrifice the system speed. In this implementation, FFT is calculated for voltage signals, but in other applications, FFT can be computed for current signals also, and different types of relays, PMU's, emergency conditions and other applications can be computed. Voltage and current signals are given to dual channel of ADC for FFT and other measurements.

FFT is implemented on GENESYS VIRTEX 5 FPGA board. Device utilization details can be computed for different optimization goal viz speed or area and optimization efforts high or normal. Device utilization for transform size 128 and 256 is shown in Tables 6 and 7 in high speed optimization.

Table 6 Device utilization summary for transform size 128 in high speed optimization

S. No.	FPGA (Genesys Virtex-5)	Used	Total available	Utilization (%)
1	Clock	3	32	9.37
2	Logic	2507	28,800	8.70
3	Slice registers	2600	28,800	9.03
4	Slice LUTs	2625	28,800	9.11
5	IOB's	14	480	2.92
6	Memory	118	7680	1.54
7	Registers	164	–	–
8	Flip flops	164		

Table 7 Device utilization summary for transform size 256 in high speed optimization

S. No.	FPGA (Genesys Virtex-5)	Used	Total available	Utilization (%)
1	Clock	3	32	9.37
2	Logic	2512	28,800	8.72
3	Slice Registers	2603	28,800	9.04
4	Slice LUTs	2630	28,800	9.13
5	IOB's	14	480	2.92
6	Memory	118	7680	1.54
7	Registers	252		
8	Flip flops	252		

5 Conclusion

Noise and harmonics affect the power system operation in terms of false triggering, noise, skin effect, overheating, overloading of neutrals, etc. In this paper, FFT is implemented for extracting fundamental frequency component of voltage in Radix-2 Burst Mode. Concurrent sense, process and communicate modules are implemented for FFT implementation and displaying fundamental frequency component on hyper-terminal and ethernet environment. Time delay for different modules in sequential implementation and concurrent architecture are presented. Total 44% time is condensed in FPGA-based implementation as contrast to microprocessor/microcontroller-based implementation. Additional required functionality and other applications can be added in process module without sacrificing the speed of the system. Device utilization summary for different transform size in high speed optimization is reported in the paper. Different types of IP cores are used for implementing FFT on GENESYS VIRTEX 5 FPGA board. Different modules of sense, process and communicate are implemented in VHDL programming.

References

1. Wyche MK (1991) An efficient, microprocessor-based FFT implementation for FDM demultiplexing. In: IEEE military communications conference, 1991. MILCOM '91, conference record, military communications in a changing world
2. Jia L, Li B, Gao Y, Tenhunen H (1998) Implementation of a low power 128 point FFT. In: IEEE 5th international conference on solid-state and integrated circuit technology
3. Stucki P (1975) Solid state array area scanning arrays as interface devices between optical and digital computing systems—a simulation study. IEEE Trans Comput C-24(4):370–380
4. Van der Auweraer H, Snoeys R (1987) FFT implementation alternatives in advanced measurement system. IEEE Micro 7(1):39–49
5. Stigall PD, Ziemer RE, Hudec L (1982) A performance study of 16 bit microcomputer-implemented FFT algorithms. IEEE Micro 2(4):61–66
6. Chong K-S, Gwee B-H, Chang JS (2006) A low energy asynchronous FFT/IFFT processor for hearing aid applications. In: IEEE conference on electron devices and solid-state circuits. <https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=108>
7. Mouine J, Hamida AB, Chtourou Z, Lakhoua N, Samet M (2002) Implementation of an FFT based simulation algorithm on a fully programmable cochlear prosthesis. In: IEEE Canadian conference on electrical and computer engineering
8. Swartzlander Jr EE, Saleh H (2012) FFT implementation with fused floating-point operations. IEEE Trans Comput 61(2):282–288
9. Phillipov Ph, Lazarov V, Zlatev Z, Ivanova M (2006) A parallel architecture for Radix-2 Fast Fourier Transform. In: IEEE symposium on modern computing, John Vincent Atanasoff 2006
10. Maheshwari V, Das DB, Saxena AK (2014) FPGA based digital overcurrent relay with concurrent sense-process-communicate cycles. Int J Electr Power Energy Syst 55
11. Luo P, Fan H, Zhang S, Hao X, Wang X (2019) A new measurement algorithm for PMU in power system based on all-phase Fourier transform. EURASIP J Wireless Commun Network 1 2019(165)
12. Mao C-A, Xie Y, Wei X, Xie Y-Z, Chen H (2019) FPGA-based fault injection design for 16K point FFT processor. J Eng 2019(21):7994–7997

13. Heideman MT, Johnson DH, Burrus CS (1984) Gauss and the history of the fast Fourier transform. IEEE ASSP Maga 1(4):14–21
14. Xilinx Homepage. <https://www.xilinx.com/ipcenter>. Last accessed 17 Dec 2019
15. Analog Homepage. https://www.analog.com/static/importedfiles/data_sheets/AD7476A_7477A_7478A.pdf. Last accessed 28 July 2019

Analysis of MTCMOS Cache Memory Architecture for Processor



Reeya Agrawal and Vishal Goyal

Abstract In this paper, power reduction technique is applied over SRAM and sense amplifier and then calculate the power consumption through cadence tool. Over the conclusion of analysis, A Architecture of cache memory has been done with lowest leakage power. In this paper from different sense amplifiers, we conclude that Charge-Transfer SA have lowest power dissipation, i.e., 11.06 μW . SRAM has 220.078 μW power dissipation after applying leakage power reduction technique such as MTCMOS_technique, Footer Stack Technique, Sleepy Keeper Technique and Sleep-Stack Technique, and there is 98–99% reduction and 75–76% reduction in Charge-Transfer SA. So, after all conclusion, architecture has been made with MTCMOS SRAM memory and MTCMOS Charge-Transfer SA with 98% reduction in power dissipation. This paper describes that MTCMOS_technique applied over different circuits reduces leakage power reduction as well as SRAM and CTSA with MTCMOS_technique in architecture gives low power consumption for a cache memory.

Keywords MTCMOS CTSA (charge-transfer sense amplifier with MTCMOS_technique) · MTCMOS SRAM (Static Random Access memory with MTCMOS_technique) · SA (Sense Amplifier)

1 Introduction

In today's world, technology needs low power dissipation devices. Due to portable handle devices as they have low number of power plugs at their surrounding, i.e., they need a high battery backup devices [1] which consume less amount of energy during non-working process and as well as in working process.

R. Agrawal (✉) · V. Goyal
GLA University, Mathura, India
e-mail: agrawalreeya98@gmail.com

V. Goyal
e-mail: vishal.goyal@gmail.com

Table 1 Power dissipation of different techniques applied on SRAM

S. No.	SRAM with techniques	Power consumption (μ W)
1	SRAM	220.078
2	MTCMOS SRAM	89.91
3	Sleep-stack SRAM	135.92
4	Sleepy keeper SRAM	133.39
5	Footer stacked SRAM	118.72

Table 2 Power dissipation of different sense amplifiers

S. No.	Sense amplifier	Power consumption (μ W)
1	Voltage mode sense amplifier	86.66
2	Current-mode sense amplifier	54.21
3	Charge-transfer sense amplifier	11.06
4	Voltage latch sense amplifier	419.502
5	Current latch sense amplifier	152.89

In proposed work, a design is made under convince of day to day life as it consumes less amount of energy. In this architecture [2], we apply MTCMOS technique over SRAM cell because after applying leakage reduction techniques, we compare all SRAM cell as shown in Table 1 conclusion arises that SRAM with MTCMOS technique consumes lowest power. Similarly, from Table 2, conclusion arises that VMSA, CMSA and CTSA consume lowest power among all sense amplifiers. Due to this reason, leakage reduction techniques [3] are applied over VMSA, CMSA and CTSA. And from Tables 1 and 2, architecture is form of MTCMOS SRAM, WDC, PCH and VMSA, CMSA and CTSA, respectively.

After this analysis, A “MTCMOS CACHE MEMORY ARCHITECTURE” is formed which is made up [4] of MTCMOS SRAM, PCH, WDC and MTCMOS CTSA which shows 98–99% less power consumption [5].

2 MTCMOS Cache Memory Architecture

2.1 Write Driver Circuit

When discharging bl (bit lines) voltage, then WDC reduces the write margin of “SRAM” from highest PCH value [6] as shown in Fig. 1.

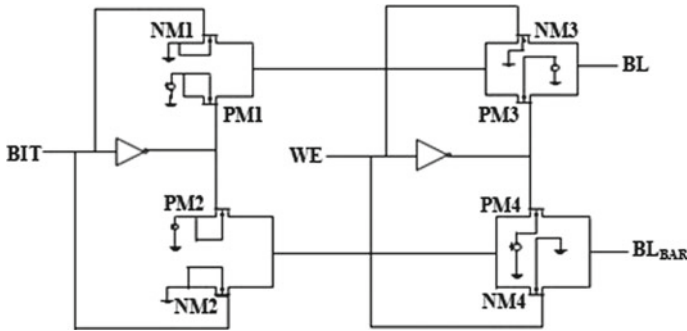


Fig. 1 Write driver circuit schematic

2.2 Conventional SRAM

“SRAM” is a 6T design. SRAM works as a cache memory in computers. It is used to store a data. It is made by connection of two inverters back to back with two complementary transistors [7]. The “read” and “write” operations can be done through bit lines in “SRAM” cell as shown in Fig. 2. It is popular due to its high stability property and lowest “static” power dissipation. “Access” transistors which are connected to the bl (bit_lines) are in working process when write line (wl) is enabled so “read” and “write” operation can be done [8]. Figure 3 shows the 8T design, i.e., 6T SRAM cell with 2T technique, i.e., MTCMOS technique [13].

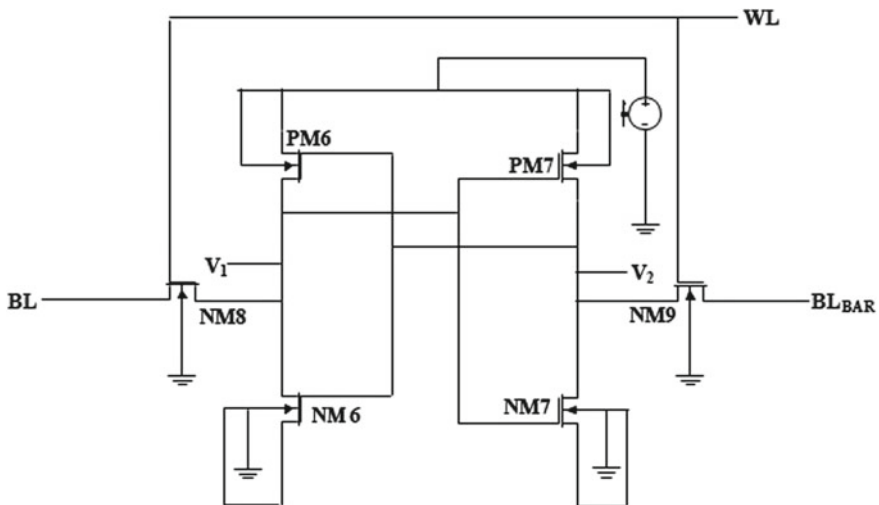


Fig. 2 Conventional SRAM circuit diagram

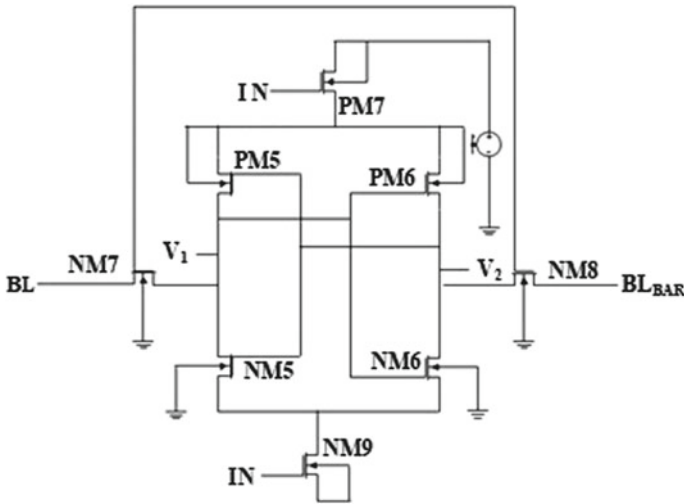


Fig. 3 MTCMOS SRAM cell

2.3 Sense Amplifier's

2.3.1 Voltage Mode Sense Amplifier

For sensing the potential difference at bit capacitance, VMSA requires differential discharging [9]. Some voltage differential is developing at bl when WL is enabling. Saen is enabled when sufficient voltage differential is developed in VMSA due to which interconnected inverters follow +ve feedback loop and convert the difference in o/p to full rail o/p [10]. Figure 4 presents the electrical picture of VMSA.

2.3.2 Current-Mode Sense Amplifier

The CMSA manage by sensing the bit_lines current. CMSA does not depend on variation in voltage value developing on bl (bit_lines). It is suitable in reducing the bit line (bl) voltage considering that low voltage at bit line can be clamped at high voltage. Figure 5 presents the electrical diagram of CMSA [11].

2.3.3 Charge-Transfer Sense Amplifier

CTSA works on a principle of charge rearrangement, i.e., from high capacitance bl to low capacitance SA o/p points. Due to this, CTSA consume low power [12]. Figure 6 presents the electrical diagram of CTSA. The circuit is divided into two parts. In first part, CG cascade made with P_1, P_3 and P_5 (and P_2, P_4 and P_6). The pmos

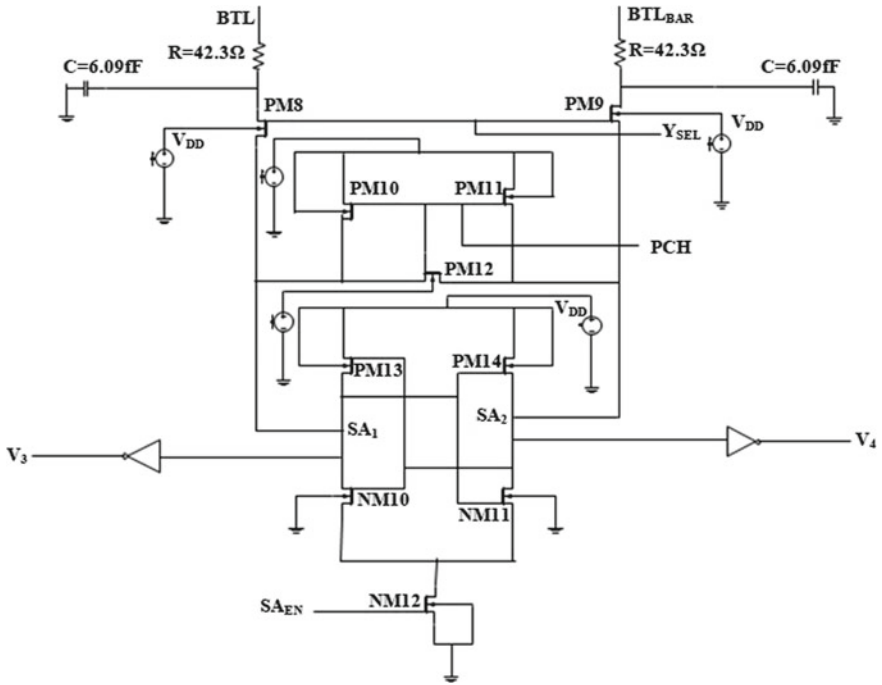


Fig. 4 VMSA schematic

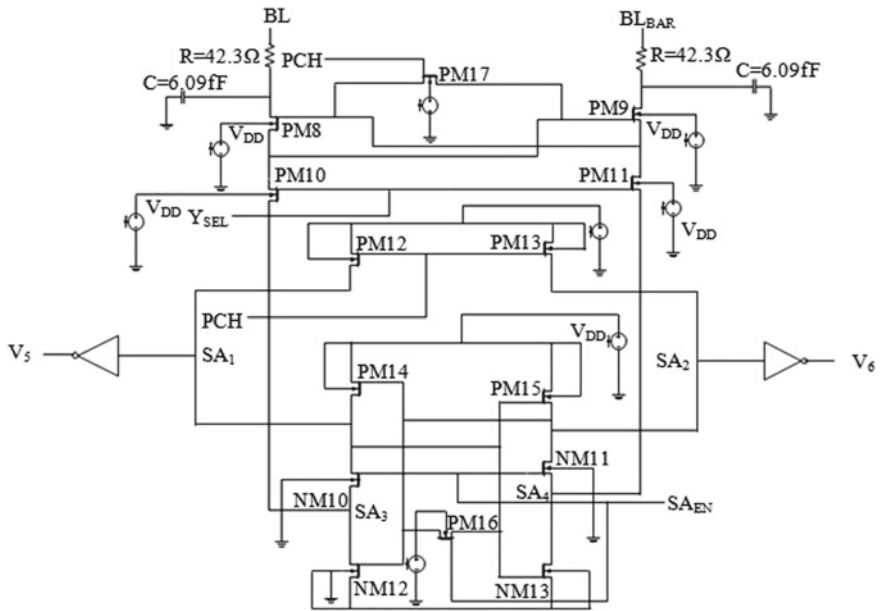


Fig. 5 CMSA schematic

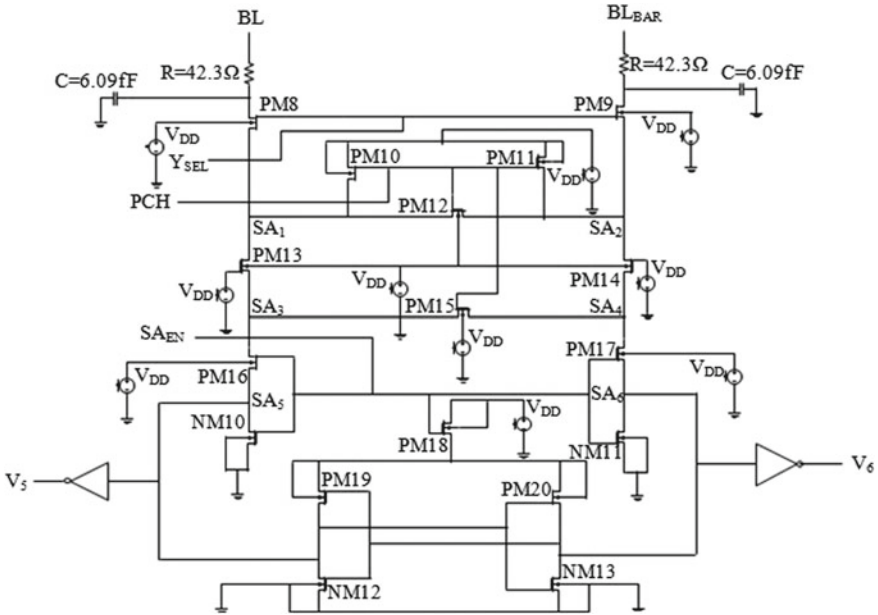


Fig. 6 CTSA Schematic

P₁ and P₂ are biased at potential V_b. In second part, P₇ through N₁₁ latches made cross-coupled inverters.

2.3.4 Voltage Latch Sense Amplifier

Figure 7 is a VLSA electrical diagram. P₁, P₂, N₁ and N₂ form the inverters. Differential voltage o/p bit line (bl) converted into full-swing o/p by inverter. The internal node of circuit is charged through bit lines (bl). Difference creates on internal nodes by input bit lines operates the electrical design [13].

2.3.5 Current-Latch Sense Amplifier

CLSA is popular due to low power consumption and an automatic power saving scheme as shown in Fig. 8. In read cycle, small difference on bit line (bl) is a data of cell [14]. The two gates n1 and n2 are connected to bl and blb. The serially connected latch circuit is controlled by current flow of two n-mos.

Fig. 7 VLSA schematic

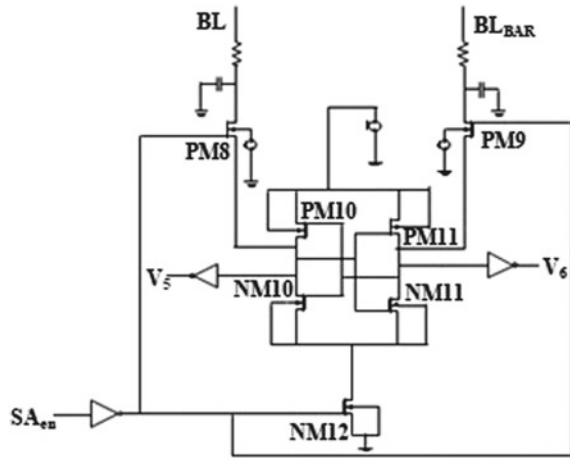
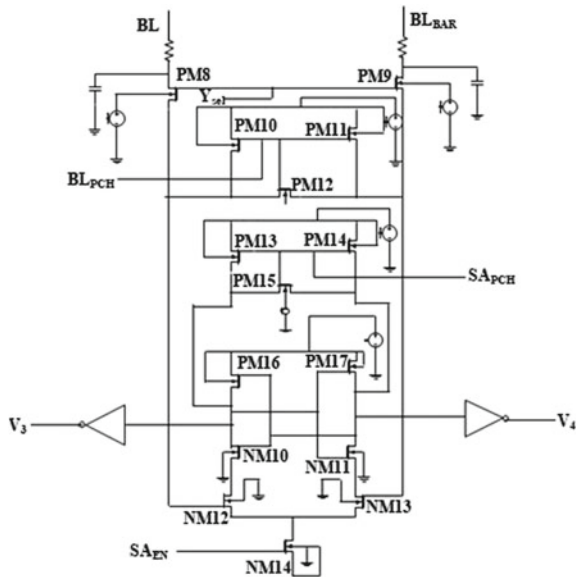


Fig. 8 CLSA schematic



3 Analysis of Result

Figure 10 represents the block structure of Single Bit MTCMOS CACHE MEMORY ARCHITECTURE [15] implemented with MTCMOS SRAM and MTCMOS CTSA (Fig. 9).

Cache memory architecture made of WDC, PCH Circuit, “SRAM” cell and sense amplifier, i.e., CTSA. Every block is evaluated and described below with their outputs [16].

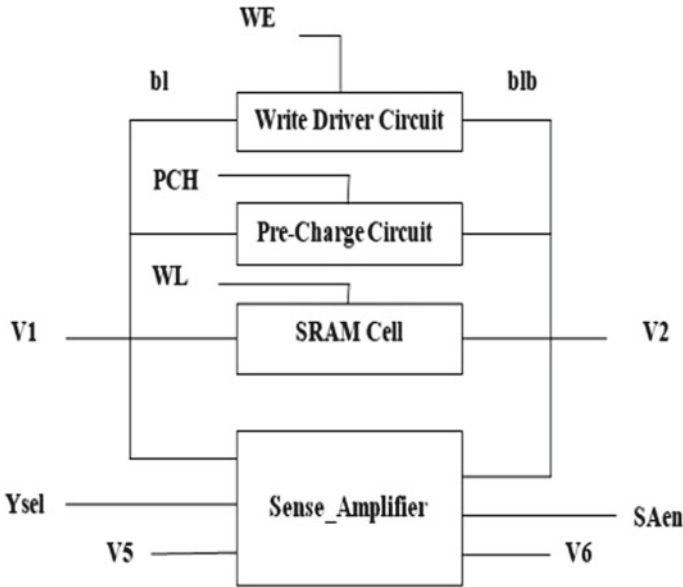


Fig. 9 Block structure of single-bit MTCMOS cache memory architecture

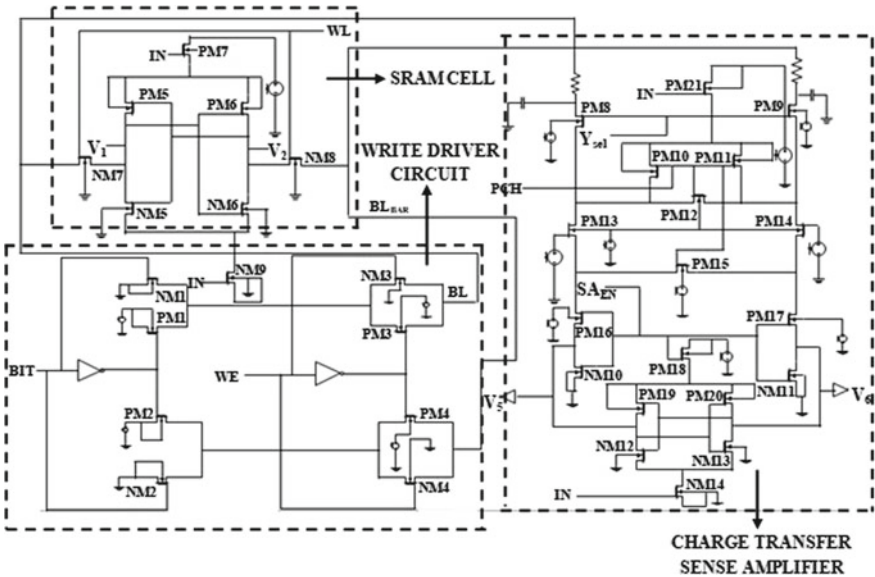


Fig. 10 Circuit Diagram of MTCMOS CACHE MEMORY ARCHITECTURE

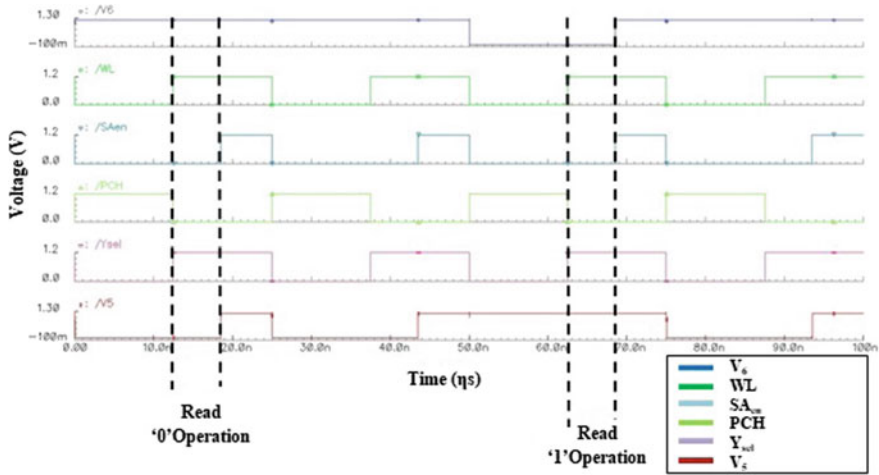


Fig. 11 Output waveform of MTCMOS CACHE MEMORY ARCHITECTURE

Figure 10 shows single-bit MTCMOS CACHE MEMORY ARCHITECTURE. The output from WDC is connected to the bit lines of SRAM Cell. O/p is stored in memory cell when write line (i.e., WL = 1) is enabled. Sense amplifier turns on when Saen = 1 [17]. Figure 13 represents the o/p wave form of “MTCMOS CACHE MEMORY ARCHITECTURE”. When bl is charged up to V_{dd} by PCH circuit, read and write operation can be done. When WE = 1, bit line stored the o/p data, and by charging and discharging the bit lines, write operation can be performed. Now, charge the bit lines up to V_{dd} (i.e., bl = V_{dd}), and the SA sensed the change in voltage at o/p points. The o/p matches the data saved in “SRAM” cell, which represent “read 0” and “read 1” operation (Fig. 11).

Table 1 displays the “power dissipation” of SRAM and SRAM with different leakage reduction techniques, whereas Table 2 demonstrates the power dissipation of different sense amplifiers. As today’s need is device with low power consumption, so different leakage reduction techniques are applied over different sense amplifiers as shown in Table 3.

From Table 2, conclusion arises that VMSA, CMSA and CTSA consume low power among all five types of sense amplifier, due to this reason, a Single-Bit Cache Memory Architecture is designed using VMSA, CMSA and CTSA, and their power consumption analysis has been seen as shown in Table 4.

From Table 1, conclusion arises that SRAM with MTCMOS technique consumes low power, and from Table 2, conclusion arises that Charge-Transfer sense amplifier consumes lowest power, and from Table 3, conclusion arises that CTSA with MTCMOS technique consume low power. So, now architecture has been designed having SRAM with MTCMOS technique, PCH, WDC and CTSA with MTCMOS technique, and the complete structure is known as “MTCMOS CACHE MEMORY

Table 3 Power dissipation of sense amplifiers using leakage power reduction techniques

S. No.	Technique	SA				
		VLISA (μW)	CLSA (μW)	CTSA (μW)	VLISA (μW)	CLSA (μW)
1	Forced-stack technique	52.94	5.54	2.587	257.102	109.838
2	MTCMOS technique	53.21	5.482	2.59	255.344	64.79
3	Sleep-stack technique	65.19	6.726	2.64	265.625	127.383
4	Sleepy keeper technique	53.18	6.738	2.66	261.103	108.154

Table 4 Power dissipation of CACHE MEMORY with different sense amplifiers

S. No.	Architecture	Power consumption (μW)
1	SRAM with VMSA	266.821
2	SRAM with CMSA	274.296
3	SRAM with CTSA	231.293

Table 5 Conclusion of all tables using different techniques over architecture

Power consumption of SRAM with CTSA in memory architecture	Power consumption of modified SRAM with CTSA in memory architecture	Power consumption of SRAM with modified CTSA in memory architecture	Power consumption of modified SRAM with modified CTSA in memory architecture
231.293 μW	11.30 μW	112.876 μW	2.00 μW

ARCHITECTURE”. There is 98–99% decrement in power consumption in new design as shown in Table 4 (Table 5).

4 Conclusion

In presented work, MTCMOS CACHE MEMORY ARCHITECTURE has been implemented with MTCMOS SRAM (i.e., SRAM with MTCMOS_technique) and MTCMOS CTSA (i.e., CTSA with MTCMOS_technique). Analysis concludes that MTCMOS CACHE MEMORY ARCHITECTURE with MTCMOS CTSA offers a much superior performance in terms of reduction in power dissipation. It was found that MTCMOS technique implemented in any circuit gives 75–76% reductions in power dissipation and MTCMOS CACHE MEMORY ARCHITECTURE consumes 98–99% less power as compared to conventional architecture. Furthermore, this

work has been implemented in array form of MTCMOS CACHE MEMORY ARCHITECTURE.

References

1. He Y, Zhang J, Wu X, Si X, Zhen S, Zhang B (2019) A half-select disturb-free 11T SRAM cell with built-in write/read-assist scheme for ultralow-voltage operations. *IEEE Trans Very Large Scale Integration (VLSI) Syst* 27(10):2344–2353. <https://doi.org/10.1109/TVLSI.2019.2919104>
2. Pandey S, Yadav S, Nigam K, Sharma D, Kondekar PN (2018) Realization of junctionless TFET-based power efficient 6T SRAM memory cell for internet of things applications. In: *Proceedings of first international conference on smart system, innovations and computing*. Springer, Singapore, pp 515–523
3. Jeong H, Oh TW, Song SC, Jung S-O (2018) Sense-amplifier-based flip-flop with transition completion detection for low-voltage operation. *IEEE Trans Very Large Scale Integrat (VLSI) Syst*
4. Tao Y, Hu W (2015) Design of sense amplifier in the high speed SRAM. In: *International conference on cyber-enabled distributed computing and knowledge discovery*, pp 384–387
5. Shalini AK (2013) Design of high speed and low power sense amplifier for SRAM applications. *Int J Sci Eng Res* 4(7)
6. Kaushik CSH, Vanjarlapati RR, Krishna VM, Gautam T, Elamaram V (2014) VLSI design of low power SRAM architectures for FPGAs. In: *2014 international conference on green computing communication and electrical engineering (ICGCCEE)*, pp 1–4
7. Nahid R, Singh BP (2013) Static-noise-margin analysis of conventional 6T SRAM cell at 45 nm technology. *Int J Comput Appl* 66(22)
8. Saun S, Kumar H (2019) Design and performance analysis of 6T SRAM cell on different CMOS technologies with stability characterization. In: *OP conference series: materials science and engineering*, vol 561, p 012093
9. Mohammad B, Dadabhoy P, Lin K, Bassett P (2013) Comparative study of current mode and voltage mode sense amplifier used for 28 nm SRAM. In: *24th international conference on microelectronic*, 07 March 2013
10. Sinha M, Hsu S, Alvandpour A, Burleson W, Krishnamurthy R, Borhr S (2003) High-performance and low-voltage sense-amplifier techniques for sub-90 nm SRAM. In: *SOC conference, 2003. Proceedings. IEEE international [Systems-on-Chip]*
11. Dutt R, Abhijeet (2012) High speed current mode sense amplifier for SRAM applications. *IOSR J Eng* 2:1124–1127
12. Heller L, Spampinato D, Yao Y, High-sensitivity charge-transfer sense amplifier. In: *Solid-state circuits conference. Digest of technical papers. 1975 IEEE international*.
13. Wei Z, Peng X, Wang J, Yin H, Gong N, Novel CMOS SRAM volatge latched sense amplifiers design based on 65 nm technology, pp 3281–3282
14. Chandankhede RD, Acharya DP, Patra PK (2014) Design of high speed sense amplifier for SRAM. In: *IEEE international conference on advanced communication control and computing technologies*, pp 340–343
15. Choudhary R, Padhy S, Rout NK (2011) Enhanced robust architecture of single bit SRAM cell using drowsy cache and super cut-off CMOS concept. *Int J Industr Electron Electr Eng* 3:63–68
16. Gajjar JP, Zala AS, Aggarwal SK (2016) Design and analysis of 32 bit SRAM architecture in 90 nm CMOS technology 03(04):2729–2733
17. Agrawal R, Tomar VK (2018) Analysis of cache (SRAM) memory for core i™ 7 processor. In: *9th international conference on computing, communication and networking technologies (ICCCNT)*, 2018, p 402

Optimization of LNA Consisting of CCG Stage and Mutually Coupled CS Stage Using PSO Algorithm for UWB Applications



Manish Kumar

Abstract This paper deals with optimal design of ultra-wide-band (UWB) low-noise amplifier (LNA) using particle swarm optimization (PSO) algorithm. The PSO algorithm not only helps to solve nonlinear analog circuit problem with an efficient manner but also remove the shortcoming of initial point in classical methods. The optimization of LNA also depends on topologies. Therefore, a two-stage LNA has been taken in which the first stage is common complimentary gate (CCG) and other is mutually coupled common source (MCCS). The CCG is good for low power, while MCCS is effective in frequency-dependent load. Both the stages are connected in cascade for higher gain and low noise figure(NF). In the proposed LNA, NF has been chosen as objective function, and Cadence Virtuoso tool is used for simulations with 90 nm CMOS technology. The proposed LNA shows a minimum NF of 1.92, power gain (S_{21})17.1 dB, input reflection coefficient(S_{11}) -16.4 dB at frequency 4.1 GHz. The power dissipation of the proposed LNA is less than 2 mW.

Keywords Low-noise amplifier (LNA) · Noise figure (NF) · Mutually coupled common source (MCCS) · Complimentary common gate (CCG) · Particle swarm optimization (PSO)

1 Introduction

Nowadays, a rapid growing demand of wireless sensor networks in different domains of industry applications required mixed analog technologies with high operating frequency (radio frequency). The scaling of sub-micrometer CMOS technology allows the researcher to achieve target of high bandwidth and low power to serve industry purpose.

The low-noise amplifier (LNA) is the first block of receiver chain in front-end circuit. It receives very weak signal from the antenna. Henceforth, in design criteria of this block, the designer mainly focused on to minimize the noise. Apart from

M. Kumar (✉)
GLA University, Mathura, India
e-mail: manish.kumar@gla.ac.in

the noise, LNA also required significant gain and input matching to minimize the reflection and transfer of maximum power to next block of receiver [1–4].

As noise figure, gain of LNA is highly nonlinear at high frequency. Therefore, researchers faced a complicated issue in the manually design of high-frequency analog circuit such as sizing of different components with topology. The optimizing of analog circuits in the band of design specifications is very complex and time consuming. A number of computer design software's have been used for the analysis of analog circuits and in the optimization in different parameters [5, 6]. In the past to solve the optimization problem, gradient search methods, Newton's method and nonlinear programming have been used. These all methods used the differential methods to solve the optimization, therefore always required the initial point. A pre-assume wrong initial point may deviate the scholar far away from the optimization [7–10]. Since analog optimization objective functions are complex and non-convex in nature, it is very difficult to solve these problems by classical methods as mentioned above. Proceeding for the possible solution for engineering problem, meta-heuristic optimization techniques play an important role, which are mathematically governed by nature-inspired laws. Working principles of these techniques are process of randomization and have more optimized result compared with classical method. Meta-heuristic popular methods are genetic algorithm [11], simulated annealing [12], ant colony algorithm [13], gravitational search algorithm [14], cuckoo search algorithm [15], etc. These algorithms are used to optimize different LNA parameters such as power gain (S_{21}) input reflection coefficient (S_{11}) and noise figure (NF) etc.

In this paper, particle swarm optimization (PSO) has been used for LNA optimization. Section 2 covers a design analysis of the proposed LNA, and Sect. 3 deals with PSO optimization of LNA which is followed by simulation results in Sect. 4 and conclusion of paper in Sect. 5.

2 CMOS LNA Design

The schematic of LNA for optimization is shown in Fig. 1, and its equivalent is shown in Fig. 2. Since LNA required a high gain and large bandwidth as design specification, optimization and topology both are required. The schematic has two stages, one complimentary common gate (CCG) and other one mutually coupled common source stage (MCCS). In the CCG stage, a NMOS transistor and PMOS transistor are connected in cascade. It has two advantages; one is that both are governed by common biasing circuit; henceforth, low power and other resultant gain are multiplication of gains of individual transistors [16].

The output of CCG stage is cascaded with the MCCS stage. The second stage provides the frequency-dependent load to the first stage; therefore, input impedance matching can be done at high frequency. The impedances at different points as shown in Fig. 2 are as follows.

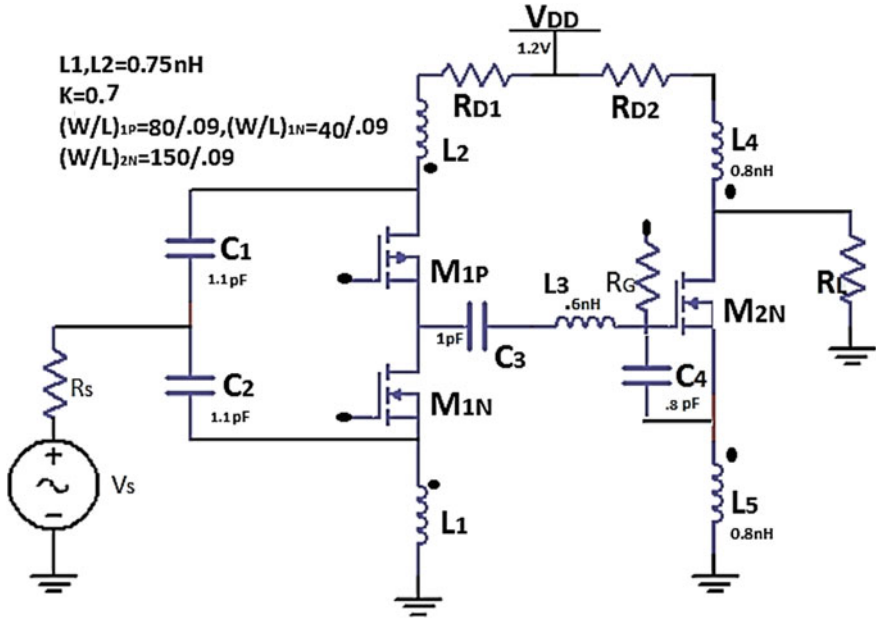


Fig. 1 Schematic of the proposed LNA

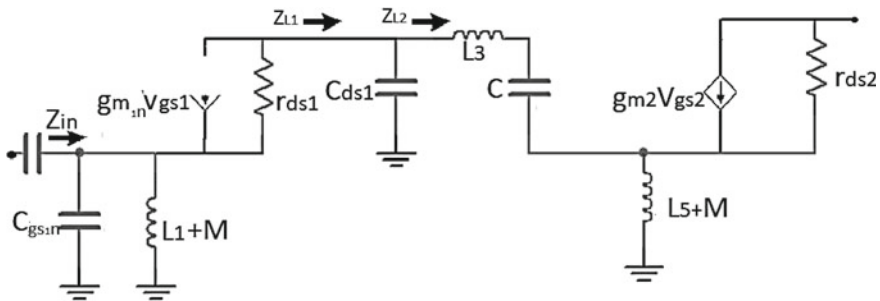


Fig. 2 Internal architecture of LNA

Input impedance $Z_{in} = \left(\frac{1}{g_{m1n}} + \frac{z_{L1}}{g_{m1n}r_{ds1n}} \right) \parallel \frac{1}{sC_{gs1n}} \parallel s(L_1 + M)$ and in the absence of load is purely resistive at frequency $f_0 = \frac{1}{2\pi\sqrt{(L_1+M).C_{gs1n}}}$. All the symbols have their as usual meaning shown in Fig. 2. Load impedance $Z_{L2} = \frac{1}{s((C_{gs2n}+C_4)\parallel C_3)} + s(L_3+L_5+M) + \frac{g_{m2n}L}{C}$ and $Z_{L1} = \left(Z_{L2} \parallel \frac{1}{sC_{d1}} \right)$ where $C = (C_{gs2n}+C_4)$, $L = L_5+M$ and $R_L = \frac{g_{m2n}L}{C}$. Therefore, sizing of all passive and active components is required to proper input impedance matching.

The noise factor of the proposed LNA for of CCG stage is gen by [16, 17]

$$\text{Noise Factor} \cong 1 + \frac{\left(4kT \frac{2}{3} \left(\frac{g_{m1p}}{g_{m1n}^2} + \frac{1}{g_{m1n}} \right) + \frac{4kT}{g_{m1n}^2 R_{D1}} + \left(\frac{K_{1p} g_{m1p}^2}{(WL)_{1p} g_{m1n}^2} + \frac{K_{1n}}{(WL)_{1n}} \right) \frac{1}{f C_{ox}} \right)}{4kT R_S}. \quad (1)$$

Equation (1) has a number of parameters; it is very difficult to optimize through conventional method; therefore, the author is used to optimize the Eq. (1) through PSO algorithm.

3 PSO and Mathematics to Optimized the LNA Circuit

Kennedy and Elbert proposed a model to optimize a single- and multi-objective function. They both study the behavior of birds and fish flocking for search of food. Based on the behavior, they proposed a model which includes mathematical equations for position best and global best as follows.

$$v_{ij}(t+1) = wv_{ij}(t) + r_2 c_2 (g_j(t) - x_{ij}(t)) + r_1 c_1 (p_{ij}(t) - x_{ij}(t)) \quad (2)$$

$$x_{ij}(t+1) = v_{ij}(t+1) + x_{ij}(t) \quad (3)$$

where i th position vector particle is given by $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]$ in the d directions; velocity and position of i th particle in j th direction are given by $v_{ij}(t+1)$ and $x_{ij}(t+1)$ after t time interval; w represents the inertia weight parameter; acceleration real-value coefficient are c_1 and c_2 , and random variable is represented by r_1 and r_2 , and its value is uniformly distributed between zero and one, while global best is represented by $g_j(t)$ and position best is denoted by $p_{ij}(t)$. Equations (2) and (3) are velocity vest and position best at time t and able to provide sufficient optimization for analog circuit. The PSO consumes a very little bit time compared with other heuristic algorithms [17, 18].

In the LNA, target specifications are biasing current 2 mA, supply voltage 0.8 V, power gain >15 dB and NF < 1.5 dB [4, 19, 20]. The proposed LNA has constrains of load resistance $R_L = \frac{g_{m2n} L}{C}$ and impedance $Z_{L1} = \left(Z_{L2} \parallel \frac{1}{sC_{d1}} \right)$ as shown in Fig. 2, and Eq. (1) has been chosen for objective function. Table 1 represents the range of different active and passive components, and the last coulombs represent the optimized value using the PSO during the optimization.

Table 1 Design parameter range with optimized value for NF

Parameters	Lower bound	Upper bound	Optimized value
C_1, C_2 (pF)	0.5	10	1.1
L_1, L_2 (nH)	0.25	10	0.75
L_3, L_4 (nH)	0	1	0.08
M (const)	0	1	0.7
C_3, C_4 (pF)	0.1	5	1, 0.8
$(W/L)_{1P}$	10/0.09	300/0.09	80/0.09
$(W/L)_{1N}$	20/0.09	200/0.09	40/0.09
$(W/L)_{2N}$	20/0.09	500/0.09	150/0.09

4 Simulated Results and Discussion

4.1 Noise Figure

Swarm size 100, number of iterations 1000, acceleration coefficient C_1 and C_2 $0.5 + \log(2)$ and inertial weight parameter (w) 0.15 has been chosen to optimize the NF and has been taken as objective function. Figure 3 represents the graph for function NF with number of iteration in PSO. NF is optimized just after few iterations and quote 1.92 dB.

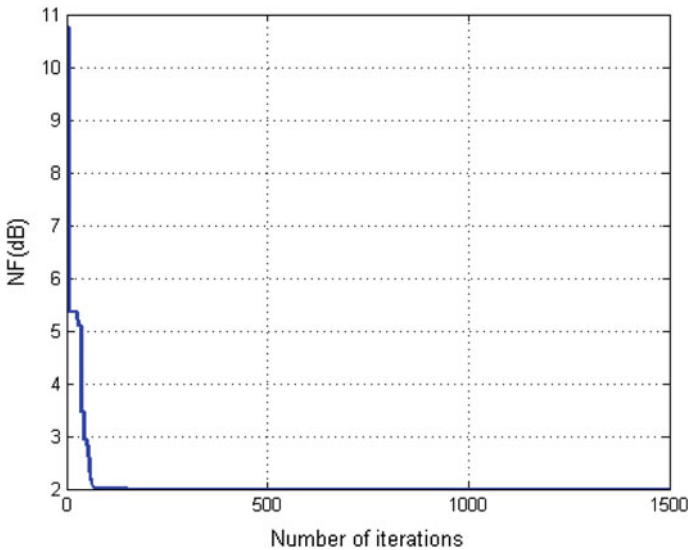


Fig. 3 NF plot using PSO

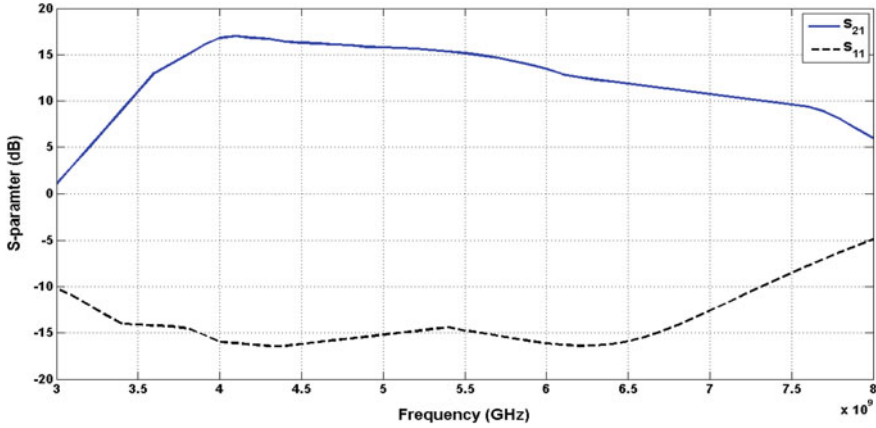


Fig. 4 S-parameters of the proposed LNA (a) S_{21} and S_{11}

4.2 S-Parameters

S-parameters always required the power gain and reflection at high frequency [15]. Power gain (S_{21}) and input reflection coefficient (S_{11}) are as shown in Fig. 4. All simulation has been done for 03–08 GHz frequency range. The power gain is positive, and input reflection is negative throughout the band. The desired range of S_{21} and S_{11} is 17.1 dB and -16.4 dB at frequency 4.1 GHz.

5 Conclusion

In this paper, the author proposed the optimal design of LNA in UWB frequency range using PSO algorithm. The PSO algorithm shows optimization for nonlinear analog circuit’s problem with an efficient manner. PSO also removes the shortcoming of initial point in classical methods. The LNA has two stages in which the first stage is common complimentary gate (CCG) and other is mutually coupled common source (MCCS). The CCG is good for low power, while MCCS is effective in frequency-dependent load. Both the stages are connected in cascade for higher gain and low noise figure (NF). In the proposed LNA, NF has been chosen as objective function and Cadence Virtuoso tool is used for simulations with 90 nm CMOS technology and depicted good results.

References

1. Chen HC, Wang T, Chiu HW, Kao TH, Lu SS (2009) 0.5-V 5.6-GHz CMOS receiver subsystem. *IEEE Trans Microwave Theory Tech* 57(2):329–35. <https://doi.org/10.1109/TMTT.2008.2011165>
2. Balankutty A, Yu SA, Feng Y, Kinget PR (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 Circuits* 45(3):538–53. <https://doi.org/10.1109/JSSC.2009.2039827>
3. 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 Circuits* 46(10):2268–2283. <https://doi.org/10.1109/JSSC.2011.2161425>
4. Parvizi M, Allidina K, El-Gamal MN (2015) A sub-mW, ultra-low-voltage, wideband low-noise amplifier design technique. *IEEE Trans Very Large Scale Integrat (VLSI) Syst* 1(6):1111–22. <https://doi.org/10.1109/TVLSI.2014.2334642>
5. Fakhfakh M et al (2010) Analog circuit design optimization through the particle swarm optimization technique. *Analog Integrated Circuits Sign Process* 63(1):71–82
6. Shams M, Rashedi E, Hakimi A (2015) Clustered-gravitational search algorithm and its application in parameter optimization of a low noise amplifier. *Appl Math Comput* 258:436–453
7. Kaveh A, Motie Share MA, Moslehi M (2013) Magnetic charged system search: a new meta-heuristic algorithm for optimization. *Acta Mech* 224(1):85–107
8. Khong SZ et al (2013) Unified frameworks for sampled-data extremum seeking control: global optimisation and multi-unit systems. *Automatica* 49(9):2720–2733
9. Kaveh A, Talatahari S (2010) A novel heuristic optimization method: charged system search. *Acta Mech* 213(3–4):267–289
10. Kennedy J (2011) Particle swarm optimization. In: *Encyclopedia of machine learning*. Springer, Boston, MA, pp 760–766. https://doi.org/10.1007/978-0-387-30164-8_630
11. Taherzadeh-Sani M et al (2003) Design optimization of analog integrated circuits using simulation-based genetic algorithm. In: *International symposium on signals, circuits and systems, 2003*. SCS 2003, vol 1. IEEE, New York
12. Kirkpatrick S, Daniel Gelatt C, Vecchi MP (1983) Optimization by simulated annealing. *science* 220(4598):671–680
13. Gonzalez E et al (2005) BSA: a complete coverage algorithm. In: *Proceedings of the 2005 IEEE international conference on robotics and automation*. IEEE, New York
14. Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. *Inf Sci* 179(13):2232–2248
15. Yang X-S, Deb S (2009) Cuckoo search via Lévy flights. In: *2009 World Congress on nature & biologically inspired computing (NaBIC)*. IEEE, New York
16. Kumar M, Deolia VK (2018) A wideband design analysis of LNA utilizing complimentary common gate stage with mutually coupled common source stage. In: *Analog integrated circuits and signal processing*, pp 1–1. <https://doi.org/10.1007/s10470-018-135>
17. Khurram M, Hasan SR (2012) A 3–5 GHz current-reuse g_m -boosted CG LNA for ultrawideband in 130 nm CMOS. *IEEE Trans Very Large Scale Integrat (VLSI) Syst* 20(3):400–409. <https://doi.org/10.1109/TVLSI.2011.2106229>
18. Toofan S, Rahmati AR, Abrishamifar A, Lahiji GR (2008) Low power and high gain current reuse LNA with modified input matching and inter-stage inductors. *Microelectron J* 39(12):1534–1537. <https://doi.org/10.1016/j.mejo.2008.07.073>
19. Sahoozadeh H, Jannesari A, Dousti M (2018) Noise suppression in a common-gate UWB LNA with an inductor resonating at the source node. *AEU-Int J Electron Commun* 1(96):144–153. <https://doi.org/10.1016/j.aeue.2018.09.007>
20. 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

Review: Parametric Variations in Analog-to-Digital Converters Using Different CMOS Technologies



Rohini Sharma, Aditya Goswami, and V. K. Tomar

Abstract This paper presents a brief review on analog-to-digital converters based on different CMOS technologies. Various parameters such as signal-to-noise and distortion ratio (SNDR), frequency sample, input frequency, spurious-free dynamic range (SFDR), power and figure of merit (FOM) of analog-to-digital converter are compared on different technology nodes. Different architectures of ADCs and their performance are also compared. It is observed that a 10-bit pipeline ADC operated at 1.8 V supply voltage has low power consumption and serves best among all the ADCs discussed at 0.18 μm technology. A two-way interleaved pipeline ADC with 12-bit 3 GS/s in 40 nm technology is acceptable because of high sampling speed. However, it has high power dissipation as compared to 12-bit successive approximation register ADC. In 28 nm, single-core L P flash ADC has highest sampling speed of 24 GS/s as compared to 12-bit interleaved pipeline ADC and SAR ADC using CDEC.

Keywords Pipelined · SAR · Flash type ADC · MDAC · CMOS

1 Introduction

The world is moving toward more digitization, to ease the complexities of physical activities. As nature only inherent the analog signals, so we require fast interface between digital machine and real world. In order to achieve this, we require faster ADCs with low power consumption, better FOM and capability of handling more and more bits in a sample [1]. With the increasing speed of data conversion, reducing the supply voltage, increasing the converter's accuracy [2] and also power reduction as well as speed are important issues while designing ADCs [3]. As the IC manufacturing technology process is moving toward advancement, reducing power consumption is easier which in turn reduces the temperature of ICs and helps in boosting battery life [4]. ADCs are made using pipeline structure in order to reduce the power consumption and conversion time [5]. In this paper, we have made a

R. Sharma · A. Goswami (✉) · V. K. Tomar
Department of Electronics and Communication, Engineering, GLA University, Mathura, India
e-mail: aditya.goswami@gla.ac.in

comparison between various analog-to-digital converters based on different CMOS technology nodes (0.065, 0.5, 0.04, 0.18, 0.09 and 0.028 μm), and we will compare the ADCs' important parameters and their architectures.

2 Important Parameters of ADC

There are different types of DC errors such as (i) offset error (ii) gain error (iii) differential nonlinearity (DNL) (iv) integral nonlinearity (INL) which define the performance of the ADCs. Let us assume a 12-bit ADCs with INL of 4 LSB. It effectively provides the accuracy of 10 bit only. In Fig. 1, the straight line is the ideal transfer function having infinite resolution. This means that for any change in the input voltage, it has corresponding output code. However, the actual transfer function will look like a staircase which is the transfer function of a 3-bit ADC with a 5 V of reference voltage. This transfer function has higher accuracy without any kind of error. In this transfer function, each step corresponds to 1 LSB.

2.1 Offset Error

Whenever the perfect transfer function or actual transfer functions shift either left or right on the horizontal axis, it results in an offset error in the ADCs.

$$\text{Offset error (V)} = \frac{\text{error in LSB} \times V_{\text{ref}}}{2^n} \tag{1}$$

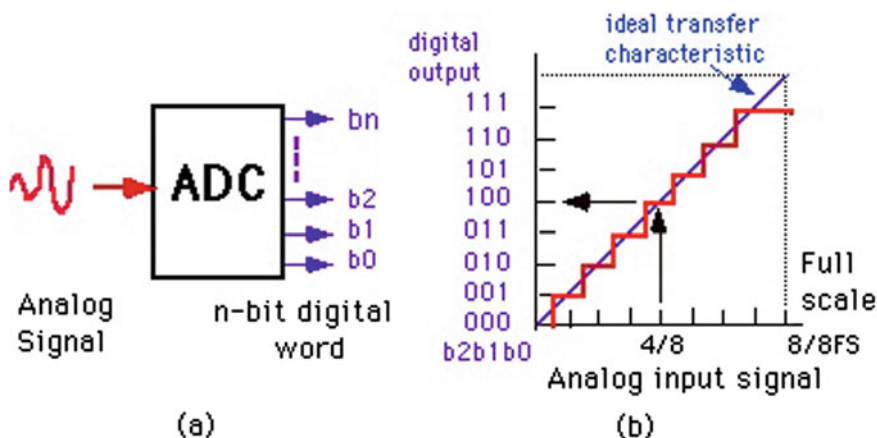


Fig. 1 Transfer function of 3-bit ADC with a 5 V of reference voltage

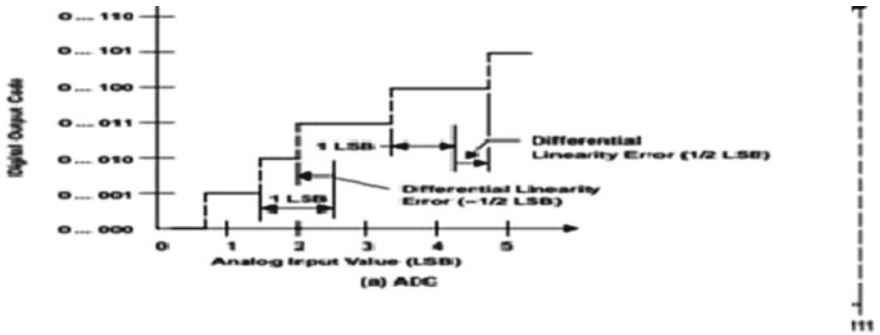


Fig. 2 Differential linearity error of a linear ADC

2.2 Gain Error

The difference between the ideal voltage at which converter provides the full-scale output voltage versus the actual voltage at which the converter provide the full-scale output code.

$$\text{Full scale voltage(FSV)} = \frac{(\text{offset error in volt})}{\text{full scale volt.}} \times 100 \tag{2}$$

2.3 Differential Nonlinearity (DNL)

It is the difference between the actual step width and the ideal step width. Due to this DNL, the actual step width could be either more or less than the ideal step size [6] (Fig. 2).

2.4 Integral Nonlinearity (INL)

It indicates how the actual transfer curve deviates from the straight line. The INL is the accumulation of all the previous DNLs up to that point. While selecting the ADC, one should look all the parameters in the data sheet.

2.5 *Signal-to-Noise and Distortion Ratio (SNDR)*

It indicates the signal quality from a signal to a communication device which is defined as [7].

$$\text{SNDR} = \frac{(\text{P}_{\text{signal}} + \text{P}_{\text{noise}} + \text{P}_{\text{distortion}})}{(\text{P}_{\text{noise}} + \text{P}_{\text{distortion}})}$$

2.6 *Spurious-Free Dynamic Range (SFDR)*

It is the ratio of fundamental frequency amplitudes to the largest harmonic amplitudes or spurious component of signal which is observed throughout full bandwidth [8].

2.7 *Effective Number of Bits (ENOB)*

Effective number of bits is a measure of dynamic range of an ADC, DAC or their associated circuitry. It describes effective resolution of the system in bits. For example, it may have 12-bit resolution, but effective number of bits can be 9.5 [9].

2.8 *Aperture Error*

With the reference of Fig. 3, the time required moving from sample mode to hold mode gives the uncertainty which in turn causes finite error in the form of delay signal [10]. The difference between the input edge of sampling clock and the real zero-crossing of input sine wave is the aperture delay time as shown in Fig. 4.

3 **Different CMOS Technologies for Implementing ADCs**

Roy et al. [11] reported a design of a pipeline 9-bit ADC for superheterodyne receiver which is implemented. This ADC works on 0.18 μm CMOS process with 1.8 V supply voltage. This implementation is tested on frequencies of 50 and 75 MS/s which provides 6.86 and 6.11 as effective number of bits. This ADC has two 2.5 bits' stages followed by two 1.5 bits' stages with three-bit flash stage. This ADC only consumes 65 mW of power. By removing SHA, aperture error will be completely removed. For getting low power design, we take the help of interpolation, capacitor splitting and

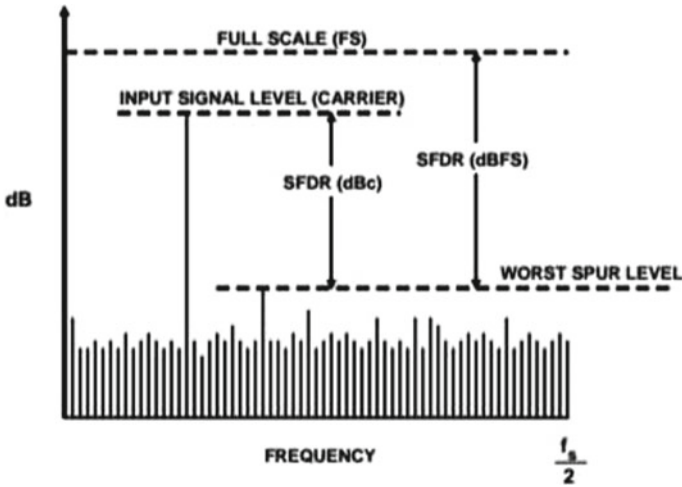


Fig. 3 Spurious-free dynamic range (SFDR)

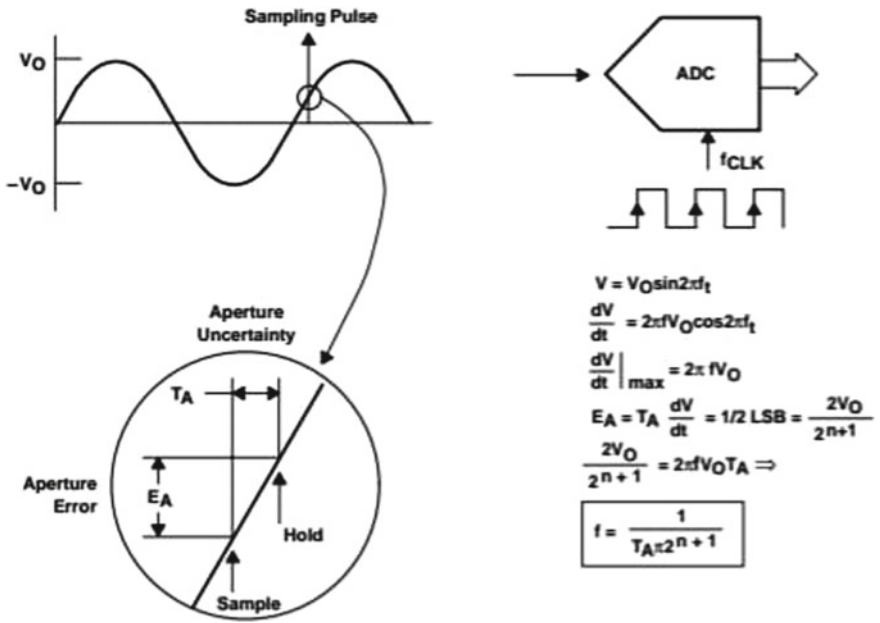


Fig. 4 Effective aperture delay time measured with respect to ADC input

offset calibration. The idea of SHA removal is to decrease the power consumption and makes it more immune to noise. [12] This 14-bit ADC with 250 MS/s speed with $0.18 \mu\text{m}$ is proposed having good speed, but a high power consumption about 120 mW. Rezapour et al. [13] reported a 10-bit pipeline ADC with high speed, better figure of merit and low power consumption. In this architecture, one SHA circuit is attached to every stage which samples residual values of predecessor stage and allows it to process a brand new sample. Sehgal et al. [14] implemented an ADC with digital CMOS and gives FOM of 157.5 dB at 1 V supply with sampling 195 MS/s at SNDR/SFDR of 64.8/82 dB. In this architecture, they are dealing with background calibration to correct nonlinearity in the first two stages and gain error in stages 1–6. Here, split ADC calibration is used for higher conversion speed. It requires 70,000 samples to attain conversions which make it a fast speed technique for distortion calibration. A 12-bit SAR ADC [15] at 40 nm CMOS technology node presents 0.9 V, 150 MS/s with adaptive radix DAC. The proposed ADC has two input comparators with two input DAC arrays. With reference to resolution of 12 bit, the power consumption is essential for meeting the requirements of speed and noise. Comparator with adaptive noise reduction (ANR) related to the redundancy is implemented for speed and power improvement. With the proposed alternating – reference – switching (ARS), DAC and noise reduction comparator demonstrates a good performance at a 0.9 V supply with excellent power efficiency. A two-way time-interleaved pipeline ADC [16] which is presented with 12-bit 3 GS/s and 40 nm CMOS technology. This architecture eliminates the headroom limitation due to the scaled power supply of nm CMOS technologies, while maintaining the intrinsic speed of MOSFETs with minimum channel length. Also, this architecture provides extrapolation scheme for multiplying DAC converter stages and the last flash stage. Using the above two techniques SNR of 61 db and DNL of $-0.5/ +0.5$ LSB, with 5000 mW, consumption is achieved in ADC. This architecture consists of a sampler, digital-to-analog converter, a coarse ADC and an amplifier. Here, we are feeding the pipeline with difference of sample voltage and with its quantized version, given by coarse ADC and DAC to get the residual voltage. This residual voltage gets amplified by precision amplifier, and it is quantized by different stages. The DAC and amplifier are combined known as multiplying DAC (MDAC). Lv et al. [17] presented a 14-bit ADC fabricated using $0.5 \mu\text{m}$ double-poly triple-metal CMOS process with a sampling frequency of 30 MS/s. This technique improves DNL and INL to 0.2 LSB and 0.29 LSB, respectively. For 1.25 MHz sinusoidal signal, SNDR and SFDR improve by value of 75.51 dB and 83.61 dB from 41.3 dB to 52.1 dB, respectively. This architecture will consume the total power of 136 mW. In [18], the upgrade of Pierre Auger Observatory is done by decreasing the power consumption using low-power ADC of 10-bit 50 MS/s 5-stage pipeline implemented in 65 nm CMOS having SNDR and SNR of 50.3 and 52.3 dB, respectively, with 59.2 dB SFDR. The main feature of this ADC is power consumption which is 12 mW for a 1.2 V power supply while taking 0.8 mm^2 area. This observatory requires large number of detectors which in turn increases the power consumption. As ADC is a most important block in the circuit, different design techniques are employed to bring down the power consumption. In order to save power, input sampling is completed on the first stage

without using extra S&H circuit. 7-bit 160 MS/s pipeline ADC [19] with 0.55 V using dynamic amplifiers is presented. To improve the ADC speed, increase supply voltage scaling robustness, and to realize scalable power consumption, common mode detection techniques are employed for amplifying residues. It is fabricated at 90 nm CMOS technology and ENOB of 6.0-bit. This ADC has FOM of 240 fJ/conversion step.

Interleaved pipeline ADC [20] with 12-bit in 28 nm CMOS technology is introduced with SNDR and SFDR of 55 and 66 dB, respectively, and dissipates the power of 2.9 W. In this architecture, eight pipeline ADCs are interleaved to get 10 GS/s sample rate. It has different techniques to decrease the power, like eliminating a SHA, residue scaling and inter-stage gain error. It employs a push-pull input buffer which drives sub-ADCs to give >7 GHz bandwidth. It also has fast bootstrap switch which gives 10 GS/s sampling operation. By using residue scaling technique, MDAC amplifiers work besides core power supply to give low power dissipation. 3-bit single-core flash ADC [21] with sampling rate of 24 GS/s without time interleaving. This ADC is sufficiently delivering full sampling rate which in turn make this ADC fast in CMOS 28 nm technology. Alioto et al. [22] proposed an ADC which is designed using source-coupled logic (SCL). These SCL circuits are comfortable for high operating frequency and also prevent switching noise which CMOS creates. This contains track-and-hold stage (T/H) with buffer at the input, followed by comparator, amplifiers and latches. It allows communication with high data output and low latency. SAR ADC [23] is proposed with 8-bit resolution and implemented in 28 nm FDSOI technology. It dissipates 5.77 mW power in total from which 0.67 mW go to CDAC, 1.27 mW to the comparators and logic blocks dissipate 3.84 mW. In this architecture by adding extra comparator pair increase the output of ADC. By using alternate comparators and using asynchronous logic, sampling speed is maximized and reaches up to 1.8 GS/s. The settling issues are overcome by implementing storing DACs in the place of static capacitors.

4 Discussion

The performance of all considered ADC is summarized in Table 1. It can be seen from Table 1 that 14-bit low-power pipeline ADC has good sampling speed, SNDR and better SFDR as compared to 9-bit quadrature parallel pipeline ADC, 14-bit low-power pipeline ADC and 12-bit pipeline implemented on 0.18 μm technology. It is also able to process more bit at a time. However, it has higher power dissipation. A 12-bit pipeline split-ADC, 12-bit SAR with adaptive radix DAC and 12-bit pipeline ADCs are implemented on 40 nm technology. It is observed that 12-bit/SAR with adaptive radix DAC has minimum power dissipation among all other ADCs simulated on 40 nm technology. 12-bit pipeline ADC has large power dissipation with higher sampling speed. A 12-bit interleaved pipeline ADC, 3-bit LP flash ADC and 8-bit SAR CDAC are based on 28 nm technology. The 3-bit LP flash ADC has the highest

Table 1 Comparison chart

Bit/type	F_{in}	Tech (μm)	Power (mW)	V_{ss}	F_s	SNDR (dB)	SFDR (dB)	FOM (dB)
9-bit quadrature pipeline ADC [11]	70 MHz	0.18	65	1.8	50 MS/s	43.05	58.09	-437
14-bit pipeline ADC [12]	45 MHz	0.18	120	1.8	250 MS/s	74.4	87.1	-433.2
10-bit pipeline ADC [13]	10 MHz	0.18	19.7	1.8	100 MS/s	54.4	-	-469.6
12-bit pipeline split ADC [14]	32.5 MHz	0.04	58	1	195 MS/s	64.4	82	-428.3
12-bit SAR [15]	70 MHz	0.04	1.5	0.9	150 MS/s	61.7	74.4	10.3FJ/cs
12-bit pipeline ADC [16]	1.5 GHz	0.04	500	2.5	3 GS/s	58	-	-
14-bit self-calibrated pipeline ADC [17]	1.25 MHz	0.5	136	5/3.3	30 MS/s	75.51	83.61	0.931 pJ/cs
10-bit /low-power pipeline ADC [18]	7.5/22 M	0.065	12	1.2	50 MS/s	50.3	52.9	897 FJ/cs
7-bit interpolated pipeline ADC [19]	80 MHz	0.09	2.43	0.55/0.5	160 MS/s	38	53	240 FJ/cs
12-bit interleaved pipeline ADC [20]	4 GHz	0.028	2900	-	10 GS/s	55	66	-147
3-bit LP flash ADC [21]	10 GHz	0.028	460	1.4/1.75	24 GS/s	-	-	3.6 pJ/cs
8-bit SAR CDAC [23]	-	0.028	5.77	1	1.8 GS/s	-	-	-

sampling rate and moderate power dissipation, while 8-bit SAR CDAC has lower sampling rate and lowest power dissipation.

5 Conclusion

This paper reviews the performance of various analog-to-digital converters based on their functional properties such as flash, time-interleaved, pipeline and successive approximation ADCs. It has been found that single-core L P flash ADC has highest sampling speed as compared to 12-bit interleaved pipeline ADC. Further, the proposed comparative analysis shall be helpful in selection of ADC for particular applications.

References

1. Yoshioka K, Itakura T, Furuta M (2017) A/D converter circuit, pipeline A/D converter, and wireless communication device. U.S. Patent No. 9,608,657
2. Matsuzawa A, Miyahara M (2015) Pipeline A/D converter and A/D converting method. U.S. Patent No. 8,947,287
3. Donno A, D'Amico S, De Matteis M, Baschiroto A (2014) A 10-b 50-MSPS low power pipeline ADC for ultra-high energy cosmic rays detection. *IEEE Trans NuclSci* 61(1):568–573
4. Chandrashekar K, Bakkaloglu B (2011) A 10 b 50 MS/s opamp-sharing pipeline A/D with current-reuse OTAs. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 19(9):1610–1616
5. Kitamura K et al (2012) A 33-megapixel 120-frames-per-second 2.5-watt CMOS image sensor with column-parallel two-stage cyclic analog-to-digital converters. *IEEE Trans Electron Devices* 59(12):3426–3433. <https://doi.org/10.1109/TED.2012.2220364>
6. <https://microchipdeveloper.com/adc:adc-differential-nonlinearity>
7. ADC parameters by Silicon Laboratory
8. SNDR|IEEE Standard for Terminology and Test Methods of Digital-to-Analog Converter Devices. *IEEEESTD1658-2011:1–126*. 1 Feb 2012. doi:<https://doi.org/10.1109/IEEESTD.2012.6152113>. ISBN 978-0-7381-7147-0
9. Understand SINAD, ENOB, SNR, THD, THD + N, and SFDR so You Don't Get Lost in the Noise Floor (PDF), Analog Devices, Inc., 2009, Retrieved 17 Aug 2012
10. EMaxim (December 17, 2001), Glossary of frequently used high-speed data converter terms, Application Note, Maxim, 740
11. Roy S, Banerjee S (2018) A 9-Bit 50 MSPS quadrature parallel pipeline ADC for communication receiver application. *J Inst Eng (India) (Ser B):1–1*
12. Wang C, Wang X, Ding Y, Li F, Wang Z (2018) A 14-bit 250MS/s low-power pipeline ADC with aperture error eliminating technique. *IEEE*, pp 1–5
13. Rezapour A, Setoudeh F, Tavakoli MB (2019) Design an improved structure for 10-bit pipeline analog to digital converter based on 0.18 μ m CMOS technology. *J Appl Eng Sci* 9:169–176. <https://doi.org/10.2478/jaes-2019-0023>
14. Sehgal R, van der Goes F, Bult K (2015) A 12 b 53 mW 195 MS/s pipeline ADC with 82dB SFDR using split-ADC calibration. *IEEE J Solid-State Circ* 50(7):1592–1603
15. Chang K, Hsieh C (2016) A 12 bit 150 MS/s 1.5 mW SAR ADC with adaptive radix DAC in 40 nm CMOS. In: 2016 IEEE Asian solid-state circuits conference (A-SSCC), Toyama, pp 157–160
16. Chen C, Wu J, Hung J, Li T, Liu W, Shih W (2012) A 12-bit 3 GS/s pipeline ADC with 0.4 mm² and 500 mW in 40 nm digital CMOS. *IEEE J Solid-State Circ* 47(4):1013–1021
17. Lv J, Que L, Wei L, Meng Z, Zhou Y (2018) A low power and small area digital self-calibration technique for pipeline. *ADC AEU-Int J Electron Commun* 83:52–57
18. Donno A, D'Amico S, De Matteis M, Baschiroto A (2014) A 10-b 50-MSPS low power pipeline ADC for ultra-high energy cosmic rays detection. *IEEE Trans Nucl Sci* 61(1):568–573

19. Lin J, Paik D, Lee S, Miyahara M, Matsuzawa A (2015) An ultra-low-voltage 160 MS/s 7bit interpolated pipeline ADC using dynamic amplifiers. *IEEE J Solid-State Circ* 50(6):1399–1411
20. Devarajan S et al (2017) A 12-b 10-GS/s interleaved pipeline ADC in 28-nm CMOS technology. *IEEE J Solid-State Circ* 52(12):3204–3218
21. Tretter G, Khafaji M, Fritsche D, Carta C, Ellinger F (2015) A 24 GS/s single-core flash ADC with 3 bit resolution in 28 nm low-power digital CMOS. In: 2015 IEEE radio frequency integrated circuits symposium (RFIC), phoenix, AZ, pp 347–350
22. Alioto M, Palumbo G (2003) Design strategies for source coupled logic gates. *IEEE Trans Circ Syst I, Fundam Theory Appl* 50(5):640–654
23. Kilic M, Leblebici Y (2018) A pipelined speed enhancement technique for CDAC-threshold configuring SAR ADC. In: 2018 16th IEEE international new circuits and systems conference (NEWCAS), Montreal, QC, pp 273–276

Power-Efficient Code Converters Using Sub-Threshold Adiabatic Logic Ultra-Low-Power Applications



Prashant Gaurav, Prashant, and Sangeeta Singh

Abstract Power dissipation becomes a decisive parameter in VLSI design in modern-day ultra-low-power applications. Sub-threshold has shown its potential as more efficient logic for ultra-low energy-consuming circuits. Circuits using sub-threshold logic have more timing delay comparable to conventional CMOS logic. Here, code converter circuits are realized using sub-threshold adiabatic logic (SAL) by deploying Cadence 45 nm technology. An extensive simulation study has been carried out, and our study validates the improved circuit performance using sub-threshold adiabatic logic. The present work will facilitate researchers for circuit realization for energy-efficient code converter circuit applications.

Keywords Sub-threshold adiabatic logic · Binary code · Gray code · Excess-3 code · Code converters

1 Introduction

Adiabatic logic is a concept which reduces the power dissipation excessively as compared to conventional CMOS logic. Low power consumption will be achieved by providing a supply that has a gradually varying voltage. Sub-threshold adiabatic logic has very less total power dissipation compared to conventional CMOS logic. The word adiabatic is derived from Greek which is referred to as a thermodynamic activity in which there is no exchange of energy with the surroundings and therefore concluded no power dissipation loss. The transistors count will be nearly half in adiabatic logic as compared to conventional CMOS logic design. Area required and delays are also comparatively lower in adiabatic logic as compared to conventional CMOS logic. Energy recovery loss is another name for adiabatic logic. Low power dissipation can be achieved by the adiabatic technique through charging and discharging the nodes using adiabatic nature. Energy stored in the load capacitor is reused in adiabatic logic circuits [1–3].

P. Gaurav · Prashant · S. Singh (✉)
Microelectronics and VLSI Design, National Institute of Technology Patna, Patna, Bihar, India
e-mail: sangeeta.singh@nitp.ac.in

1.1 Adiabatic Logic

Time-varying voltage source or constant current source as shown in Fig. 1a is used for charging the capacitor in adiabatic switching. ON resistance of the PMOS network is represented by R . Some fraction of the total energy gets stored in the capacitor which can later be claimed back by reversing the current source direction, and thus, the charge gets switch from the capacitance back into the supply.

Adiabatic switching during discharging phase is shown in Fig. 1b. Therefore, adiabatic logic circuit requires time-varying voltage source as a supply voltage. For charging the load capacitance, a constant voltage source is used in conventional CMOS logic, whereas it gets charged by constant current source in adiabatic logic.

Supply voltage applied to the adiabatic logic changes gradually (e.g., ramped waveform). The potential drop across the resistor becomes very less due to the ramp voltage. As a result, the energy dissipation across the resistance during the charge–discharge operation decreases. Power is represented as $P = E_{\text{Total}}/T = C_L V_{\text{DD}}^2 f$ if the supplied voltage is a ramped waveform with period T (i.e., frequency $f = 1/T$). Total energy consumption for both operations in the above case is given by

$$E_{\text{adiabatic}} = k(PT) = kI^2RT = k(C_L V_{\text{DD}}/T)^2 RT \quad (1)$$

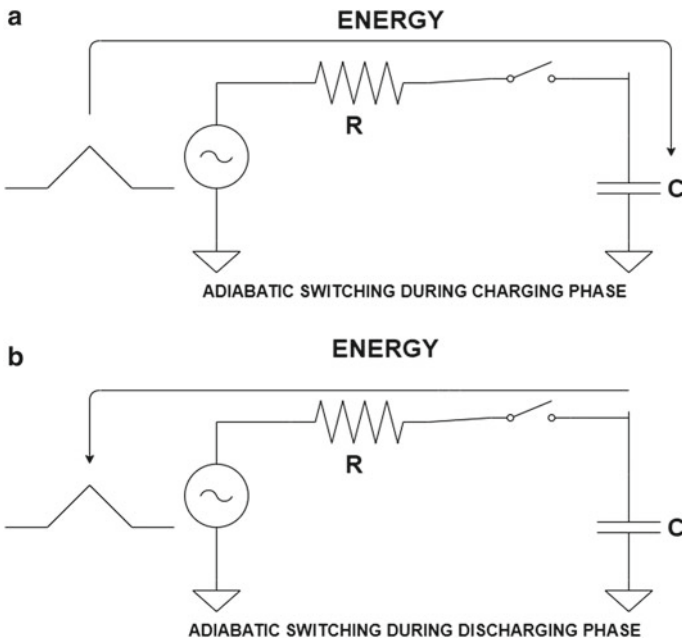
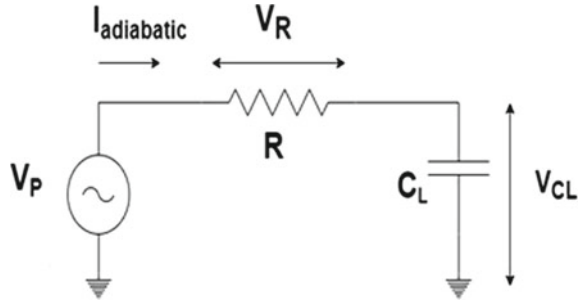


Fig. 1 Adiabatic switching **a** during charging and **b** during discharging

Fig. 2 Circuit diagram with ramp voltage source (used in adiabatic logic)



where k is the shape factor, and shape of the clock edges decide the value of k . It can be concluded here that when the period of signal T is adequately large, the energy consumption in adiabatic logic is considerably less as compared to CMOS logic. Both of these transistors can be modeled as ideal switches in series with a resistor and load capacitance C_L , as shown in Fig. 2. The channel resistance of each transistor is equal to the resistor.

Sub-threshold Adiabatic Logic The analysis of $I_D - V_{GS}$ characteristics of an NMOS transistor (W/L ratio:45 nm/45 nm) suggests that at $V_{GS} = V_T$ (where V_T is the threshold voltage of MOSFET), drain current is not equal to zero because MOS already conducts at $V_{GS} < V_T$. This region is known as the “sub-threshold” or “weak-inversion” conduction region. Leakage current in sub-threshold logic flows between drain and source regions in MOSFET and is expressed as

$$I_o = \mu C_{OX}(W/L)(n - 1)V_T^2 \tag{2}$$

$$I_{ds} = I_{oe}(V_{GS} - V_{TH})/nV_T \tag{3}$$

where μ is the mobility, C_{ox} is gate oxide film capacity, V_T is the thermal voltage which is equal to 26 mV at 300 K, W&L is width and length of channel, respectively, n is sub-threshold slope parameter [4–9]. The main component of leakage in sub-threshold devices is sub-threshold leakage current, and many other leakage components which are nearly equal in magnitudes are dependent on the device design parameters. The delay of the circuit increases quickly since the driving current reduces exponentially. Therefore, SAL logic can only be applied to confined areas where performance is not of primary importance. Sub-threshold conduction is very small for long-channel devices in OFF state [10–16]. It is a considerable factor when transistor size, as well as supply voltage, is scaled down. Delay in SAL logic is comparable to conventional CMOS logic. SAL logic can be used where performance is not the key.

2 Code and Code Converter

2.1 Gray Code

Gray code is an ordering of the binary number system in such a manner that each incremental value can only differ from the previous value by only one bit. It is also known as cyclic code as each successive code word differs from the preceding one in only one bit position. It is also a popular example of reflective codes. It is widely used in digital communication for error correction. Gray codes are used in linear and rotary position encoders instead of weighted binary encoding. This means that while using gray code in rotator shaft encoder, only a single bit differs in successive bits; so if multiple bit differs, it will be easy to detect errors.

2.1.1 Binary-to-Gray Code Converter and Gray-to-Binary Code Converter

This converter is a combinational circuit that converts binary code to equivalent gray code. The leftmost bit of gray code is equivalent to the leftmost bit of the given binary code. The second leftmost bit of the gray code is the EX-OR of the leftmost and the second leftmost bit of given binary number. The third leftmost bit of the gray code is the EX-OR of the second leftmost and third leftmost bit of given binary number. And in this manner, binary code to gray code conversion goes on. The gate-level circuit implementation for binary-to-gray code converter is shown in Fig. 3.

This converter is a logical circuit that converts gray code to equivalent binary code. The leftmost bit of binary code is equivalent to the leftmost bit of given gray code. The second leftmost bit of the binary code is the XOR of the leftmost and second leftmost bit of the given gray code. The third leftmost bit of the binary code is the XOR of the second leftmost bit of gray code and third leftmost bit of given binary code. Hence, in this manner, gray code to binary code conversion goes on.

Fig. 3 Gate-level circuit of binary-to-gray code converter

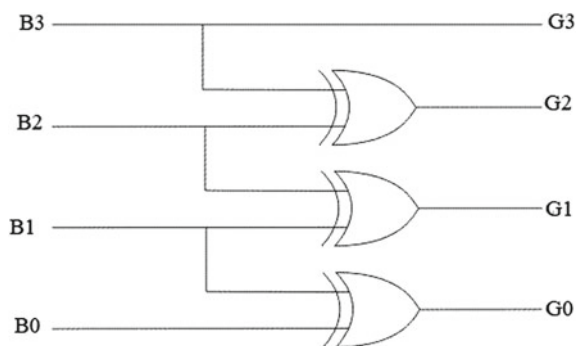
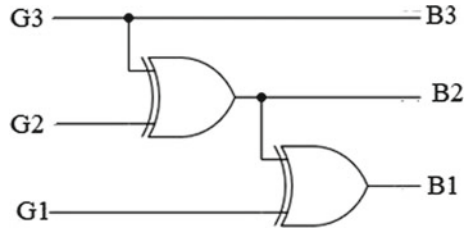


Fig. 4 Gate-level circuit of gray-to-binary code converter



The gate-level circuit implementation for gray-to-binary code converter is shown in Fig. 4.

2.1.2 Excess-3 Code and Binary-to-Excess-3 Code Converter

Excess-3 code is a non-weighted code, where each digit binary code word is the combination of corresponding 8421 code word and 0011. Non-weighted are codes that are not assigned fixed values. It is a biased representation. It is also a self-complementary code. It overcomes the difficulties faced during arithmetic operation in 8421 BCD code. Another major advantage of this representation is that the 0000 and 1111 codes are not used for representation of any digit. It is a logical circuit which converts binary-coded decimal to excess-3 code converter. Binary-coded decimal can be converted to excess-3 code by adding 0011 to the given code. Logical expression for this conversion is given below.

$$Y_3 = B_3 + B_2 \cdot B_1 + B_2 \cdot B_0 \tag{4}$$

$$Y_2 = B_2' B_1 + B_2' B_0 + B_2 B_1' B_0' \tag{5}$$

$$Y_1 = B_1 B_0 + B_1' B_0' \tag{6}$$

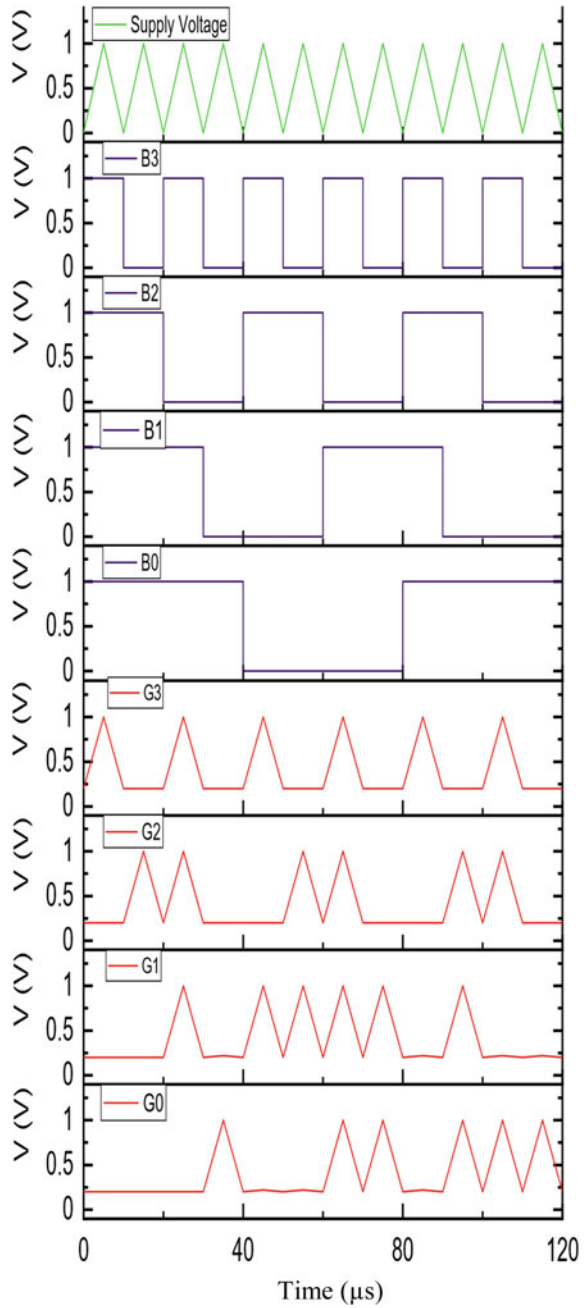
$$Y_0 = B_0' \tag{7}$$

where binary and excess-3 codes are given by $B_3 B_2 B_1 B_0$ and $Y_3 Y_2 Y_1 Y_0$, respectively.

3 Simulation Result and Discussion

The simulation analysis results presented are obtained on 45-nm Cadence Virtuoso using SAL logic [17, 18]. Figure 5 demonstrates the output waveform of SAL binary-to-gray code converter. In this figure, the first plot is of supply voltage; the second, third, fourth and fifth plots are of B_3 , B_2 , B_1 and B_0 , respectively, where $B_3 B_2 B_1 B_0$ is

Fig. 5 Output waveform of SAL binary-to-gray code converter



the binary code data. The fifth, sixth, seventh and eighth plot are of G_3 , G_2 , G_1 and G_0 , respectively, where $G_3G_2G_1G_0$ is the gray code data. Figure 6 shows the output waveform of SAL gray-to-binary code converter. Here, in this figure, the first plot is of supply voltage; the second, third, fourth and fifth plots are of G_3 , G_2 , G_1 and G_0 , respectively, where $G_3G_2G_1G_0$ is the gray code data. The fifth, sixth, seventh and eighth plot are of B_3 , B_2 , B_1 and B_0 , respectively, where $B_3B_2B_1B_0$ is the binary code data. Figure 7 shows the output waveform of SAL binary-to-excess-3 code converter. Here, in this figure, the first plot is of supply voltage; the second, third, fourth and fifth plots are of B_3 , B_2 , B_1 and B_0 , respectively, where $B_3B_2B_1B_0$ is the binary code data. The fifth, sixth, seventh and eighth plot are of Y_3 , Y_2 , Y_1 and Y_0 , respectively, where $Y_3Y_2Y_1Y_0$ is the excess-3 code data.

From Tables 1, 2 and 3, it can easily be observed that power dissipation in SAL logic is approximately 10^{-3} times lower than that of power dissipation in CMOS logic.

For binary-to-gray code converter bit G_3 , the adiabatic circuit has a considerably low power dissipation of 2.71 pW against CMOS logic, 3.495 nW. For gray-to-binary code converter bit B_0 , the adiabatic circuit has a considerably low power dissipation of 27.52 pW as compared to conventional CMOS logic, 8.11 nW. For binary-to-excess-3-code converter bit Y_3 , the adiabatic circuit has a considerably low power dissipation of 129.06 pW against conventional CMOS logic, 3.86 nW. It can easily be observed that power dissipation in SAL logic is approximately 10^{-3} times lower than that of power dissipation in CMOS logic. The detailed performance comparison of SAL and CMOS-based converter circuit is shown in Tables 1, 2 and 3. For CMOS circuit and SAL circuit, the peak DC voltage connected and peak ramp voltage applied are 1 V.

4 Conclusion

Binary code to gray code converter, gray code to binary code converter and binary code to excess-3 code converter are realized using sub-threshold adiabatic logic (SAL) and compared with conventional CMOS logic in this paper. From the above simulation results, we conclude that SAL reduces appreciable amount of energy as in parallel with conventional CMOS logic. Scaling down of power dissipation in adiabatic circuits is mainly because of recycling of energy stored in the capacitive loads. SAL is preferred for application which requires low frequency. This proposed sub-threshold adiabatic logic can be used in energy-efficient converter circuit.

Fig. 6 Output waveform of SAL gray-to-binary code converter

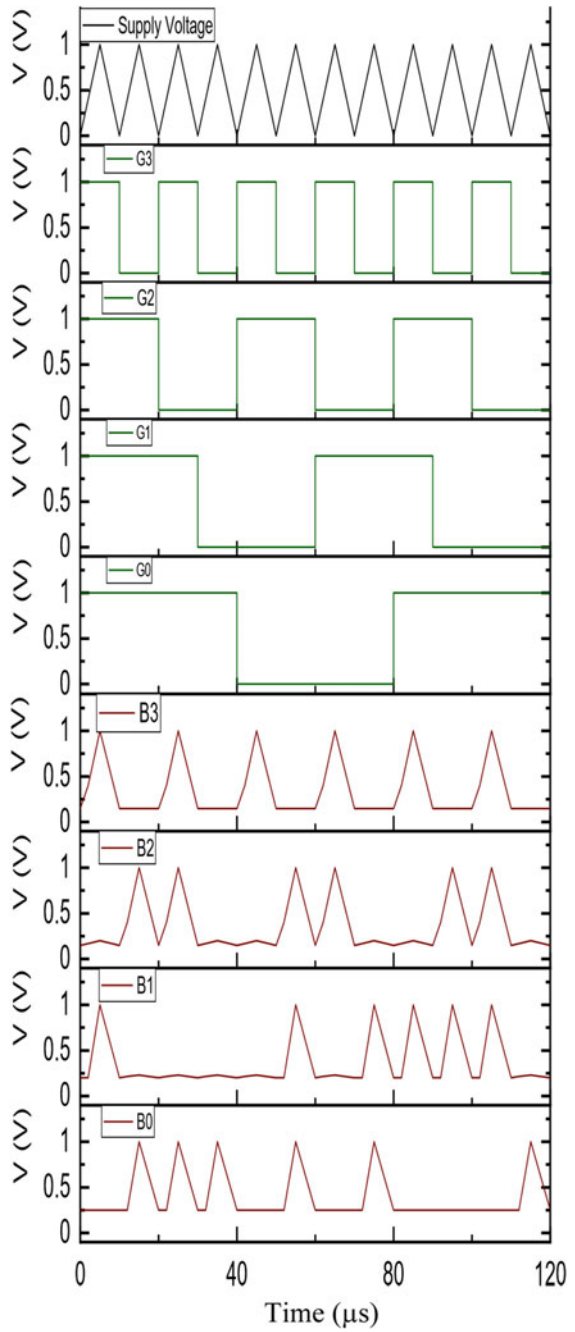


Fig. 7 Output waveform of SAL binary-to-excess-3 code converter

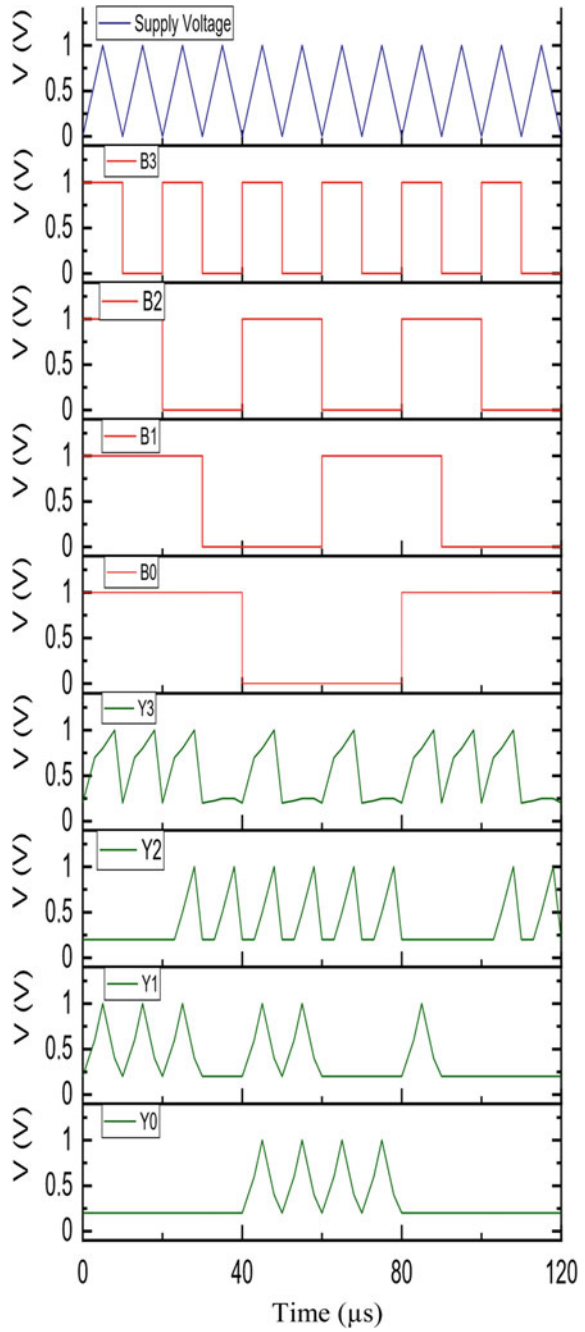


Table 1 Comparison of total power dissipation of binary-to-gray code converter using SAL logic and CMOS logic

BITS	SAL logic (pw)	CMOS logic (nw)
G_3	2.71	3.495
G_2	2.714	1.84
G_1	2.452	2.08
G_0	2.306	2.07

Table 2 Comparison of total power dissipation of gray-to-binary code converter using SAL logic and CMOS logic

BITS	SAL logic (pw)	CMOS logic (nw)
B_3	2.716	3.495
B_2	2.714	1.86
B_1	13.52	5.117
B_0	27.52	8.11

Table 3 Comparison of total power dissipation of binary-to-excess-3 code converter using SAL logic and CMOS logic

BITS	SAL logic (pw)	CMOS logic (nw)
Y_3	129.06	3.86
Y_2	87.12	2.418
Y_1	2.316	2.96
Y_0	1.68	2.89

References

1. Chaudhuri A, Saha M, Bhowmik M, Pradhan SN, Das S (2015) Implementation of circuit in different adiabatic logic. In: 2015 2nd IEEE international conference on electronics and communication systems (ICECS), pp 353–359
2. Chanda M, Jain S, De S, Sarkar CK (2015) Implementation of subthreshold adiabatic logic for ultralow-power application. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 23(12):2782–2790
3. Yadav RK, Rana AK, Chauhan S, Ranka D, Yadav K (2011) Adiabatic technique for energy efficient logic circuits design. In: 2011 IEEE international conference on emerging trends in electrical and computer technology, pp 776–780
4. Pindoo IA, Dhariwal S, Sharma R, Lata S (2018) Speed enhancement in the performance of two phase clocked adiabatic static CMOS logic circuits. In: International conference on intelligent circuits and systems (ICICS), Phagwara, India, pp 149–154
5. Grover V, Gosain V, Pandey N, Gupta K (2018) Arithmetic logic unit using diode free adiabatic logic and selection unit for adiabatic logic family. In: 5th international conference on signal processing and integrated networks (SPIN), Noida, India, pp 777–781
6. Wang A, Calhoun BH, Chandrakasan AP (2006) Sub-threshold design for ultra low-power systems, vol 95. Springer, New York
7. Takahashi Y, Sekine T, Nayan NA, Yokoyama M (2012) Power saving analysis of adiabatic logic in subthreshold region. In: 2012 IEEE international symposium on intelligent signal processing and communications systems, pp 590–594
8. Khatir M, Mohammadi HG, Ejlali A (2010) Sub-threshold charge recovery circuits. In: IEEE international conference on computer design, pp 138–144
9. Moon Y, Jeong DK (1996) An efficient charge recovery logic circuit. *IEICE Trans Electr* 79(7):925–933

10. Goyal S, Singh G, Sharma P (2015) Variation of power dissipation for adiabatic CMOS and conventional CMOS digital circuits. In: 2015 2nd international conference on electronics and communication systems (ICECS), pp 162–166
11. Kim S, Papaefthymiou MC (2001) True single-phase adiabatic circuitry. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 9(1):52–63
12. Yadav RK, Rana AK, Chauhan S, Ranka D, Yadav K (2011) Four phase clocking rule for energy efficient digital circuits—an adiabatic concept. In: 2011 2nd IEEE international conference on computer and communication technology (ICCCT-2011), pp 209–214
13. Samanta S (2009) Adiabatic computing: a contemporary review. In: 2009 4th international conference on computers and devices for communication (CODEC), pp 1–4
14. Chanda M, Dutta R, Rahaman A, Sarkar CK (2016) Analysis of NAND/NOR gates using subthreshold adiabatic logic (SAL) for ultra low power applications. In: 2016 international conference on microelectronics, computing and communications (MicroCom), pp 1–5
15. Maheshwari S, Bartlett VA, Kale I (2018) VHDL-based modelling approach for the digital simulation of 4-phase adiabatic logic design. In: 2018 28th IEEE international symposium on power and timing modeling, optimization and simulation (PATMOS), pp 111–117
16. Maheshwari S, Bartlett VA, Kale I (2019) Modelling, simulation and verification of 4-phase adiabatic logic design: a VHDL-Based approach. *Integration* 67:144–154
17. Chanda M, Mal S, Mondal A, Sarkar CK (2018) Design and analysis of a logic model for ultra-low power near threshold adiabatic computing. *IET Circ Dev Syst* 12(4):439–446
18. Chanda M, Ganguli T, Mal S, Podder A, Sarkar CK (2017) Energy efficient adiabatic logic styles in sub-threshold region for ultra low power application. *J Low Power Electr* 13(3):472481; Eason G, Noble B, Sneddon IN (1955) On certain integrals of Lipschitz-Hankel type involving products of Bessel functions. *Phil. Trans. Roy. Soc. Lond.* A247:529–551

Design Optimization of Doping-less InGaAs TFET and GaAs/Si-Heterojunction Doping-less TFET for Potential Breast Cancer Sensing Applications



Shradhya Singh, Navaneet Kumar Singh, Sangeeta Singh, Alok Naughariya, and Neha Niharika

Abstract This work demonstrates in-depth comparative analysis for the performance optimization of doping-less InGaAs TFET and GaAs/Si-heterojunction doping-less TFET (GaAs/Si-HJDL-TFET). As doping-less TFET is a prominent device in terms of lower sub-threshold slope (SS) (<60 mV/decade) and it realizes steep switching speed. Lower bandgap materials such as InGaAs are deployed, which results in the narrowing of tunneling width which causes a large amount of carriers that can tunnel across the source-channel junction and thereby increases the drive current significantly. Moreover, to suppress the ambipolar current, heterojunction structure is very helpful. Furthermore, device structure optimization has been achieved with the mole fraction (x) composition variation of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ in both the devices. Study reveals that GaAs/Si-HJDL-TFET structure represents enhanced performance in terms of lower SS, minimum threshold voltage and greater $I_{\text{ON}}/I_{\text{OFF}}$ ratio. Interestingly, the charge sensitive characteristics of the reported device can be deployed for detection of C-erbB-2 protein, the breast cancer bio-marker sensing applications.

Keywords Hetero-junction · Doping-less · Bio-marker

S. Singh · S. Singh (✉) · N. Niharika
Microelectronics and VLSI Lab, National Institute of Technology, Patna, India
e-mail: sangeeta.singh@nitp.ac.in

S. Singh · N. Niharika
Lok Nayak Jai Prakash Institute of Technology Chapra, Chapra, India

N. K. Singh
Department of ECE, University College of Engineering and Technology (UCET), VBU,
Hazaribag, Jharkhand, India

A. Naughariya
National Institute of Technology, Raipur, India

1 Introduction

As the semiconductor industry is stepping toward the sub-10 nm regime, it imposes more challenges on the conventional MOSFET, when different scaling techniques are applied in terms of increasing the complexity in the fabrication of device and the short channel effects (SCEs) [1]. Various emerging device structures have been investigated to overcome these problems such as DG-MOSFET, SOI MOSFET, junction-less nanowire transistor, gate-all around MOSFET, etc. All the devices listed above employ the thermionic emission as a current gating mechanism, which restricts its sub-threshold slope (SS) to Boltzmann limit, i.e., 60 mV/decade. Therefore, for reducing the SS below 60 mV/decade, the device conduction mechanism has to be different. For the future switching transistor technology, the TFETs have drawn huge attention of researchers because of its ability to achieve SS below Boltzmann limit (60 mV/decade), which works on band-to-band-tunneling phenomenon [2–5]. Further, due to its exceptionally low leakage current, lower V_{th} and immunity toward various short channel effects (SCEs), it is a predominant device to be used for the low power application. Moreover, TFET performance is restricted due to its limited ON-state current because of less transmission probability of band-to-band tunneling, random dopant fluctuations (RDFs), ambipolar current, etc. Many device engineering techniques have been employed with TFET to overcome these issues such as dual material TFET, double gate TFET, n^+ pocket doped TFET, hetero-junction TFET, doping-less TFET, [6–15]. It is quite challenging to fabricate TFET at nanoscale dimensions, thus doping-less TFET has been reported to reduce this process. The work function engineering has been done on source and drain side for implementing doping-less TFET [16] by choosing appropriate metal with required work function for source as well as drain electrodes, and electrons and holes carriers are induced in the semiconductor layer beneath these electrodes. n^+ region can be formed toward drain side by incorporating hafnium (Hf) (w.f. = 3.9 eV). On the other hand, side p^+ region can be formed toward source by considering platinum (Pt) (w.f. = 5.93 eV) [17]. Moreover, to reduce the problem of lower I_{ON} current, the hetero-junction DL-TFETs, line charge III-V group compound-based DL-TFET are of great interest, because of the inherently high carrier mobility, composition-dependent bandgap variation and lower effective masses of their carriers [18]. The band tuning of low bandgap (III-V) material-based compounds in the source region has been explored to improve the tunneling current and to reduction of the tunneling width [19–21]. The major disadvantage of using III-V group hetero-junction TFETs is that they realize significantly increased OFF-state current in comparison to conventional Si-TFETs [22]. Hence, to suppress the OFF-state current and to increase the ON-state current, device structural optimization along with source material engineering must be done. In this regard, different device structures have been explored such as L -shaped/ U -shaped TFET [23], vertical TFET [24, 25] and source pocket engineered TFET [5]. These concurrent research findings inspired the authors for analysis and optimization of the performance of doping-less InGaAs TFET and GaAs/Si-hetero-junction doping-less TFET (GaAs/Si-HJDL-TFET). Here, the device performance

has been improvised by deploying the parametric sweep optimization. Firstly, the basic Si-TFET structure is replaced with the material $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ for all regions. Further, another structural variant $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ only in the source region and GaAs as pocket doped material at source/channel junction and rest regions is of Si material, which shows the improved performance for the later structure. The key focus of this research is to analyze the device DC performance of both the structures and also to analyze the impact of mole fraction (x) variation on different parameters. It is worth mentioning here that the reported charge sensitive characteristics of both the device structure can be used for the detection of C-erbB-2 protein, the breast cancer bio-marker sensing application.

The rest of the paper is ordered as stated below:

Device structure and models of simulation are defined in Sect. 2. Simulation result and discussion are mentioned in Sect. 3. Further, Sect. 4 concludes the research work with some critical findings.

2 Device Structure and Simulation

The two different 2D structures considered for the InGaAs-DL-TFET are shown in Fig. 1a, b. In the first structure, $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ material, with $x = 0.75$ composition, is selected for designing the InGaAs-DL-TFET structure, which is having a similar bandgap as of Si as depicted in Fig. 1a. Second structure has been implemented by choosing $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ material, with $x = 0.75$ composition in the source region and GaAs/Si as hetero-channel junction for GaAs/Si-HJDL-TFET realization as shown in Fig. 1b. Doping-less technique has been utilized for both the devices [16], by selecting the desired metal work function above the source as well as drain regions. Due to this work function engineering, the electrons and holes are induced thereby realizing virtual doping. Toward the drain side, n^+ region is formed by selecting hafnium electrode with work function of 3.9 eV and toward source side p^+ region can be formed by choosing platinum (Pt) electrode with work function as 5.93 eV. To avoid the formation of silicide, there must be a space between the drain electrode and source electrode with the semiconductor film. This gap is optimized between

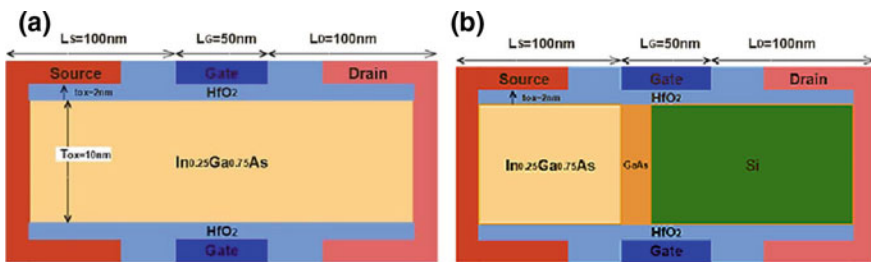


Fig. 1 2D structure of a InGaAs-DL-TFET b GaAs/Si-HJDL-TFET

Table 1 Simulation parameters for InGaAs-DL-TFET and GaAs/Si-HJD-L-TFET

Parameters	InGaAs-DL-TFET(D_1)	GaAs/Si-HJD-L-TFET(D_2)
Silicon thickness (T_{Si})	10 nm	10 nm
Gate oxide thickness (T_{ox})	2 nm	2 nm
Gate length (L_G)	50 nm	50 nm
Source length (L_S)	100 nm	100 nm
Drain length (L_D)	100 nm	100 nm
Background doping (N_{in})	$1 \times 10^{16} \text{ cm}^{-3}$	$1 \times 10^{16} \text{ cm}^{-3}$
Source work function (ϕ_S (Pt))	5.93 eV	5.93 eV
Drain work function (ϕ_D (Hf))	3.9 eV	3.9 eV
Gate dielectric constant (ϵ_{ox})	HfO ₂ (31)	HfO ₂ (31)
Source material	In _(1-x) Ga _x As	In _(1-x) Ga _x As
Channel material	In _(1-x) Ga _x As	GaAs/Si (10–40 nm) Hetero-junction
Drain material	In _(1-x) Ga _x As	Si

source and channel at 2 nm, whereas at 15 nm between the channel region and drain region. It leads to the enhancement in the ON-state current and reduces the ambipolar effect [27]. The gate oxide thickness value is optimized at 2 nm. TCAD Silvaco ATLAS simulator has been used to accomplish the study of the considered device structures. The models used in the simulation are listed as Shockley–Read–Hall (SRH), concentration and field-dependent mobility, bandgap narrowing model, Fermi model. Non-local BTBT model is used to justify the special profile and to more accurately model the process of tunneling [28]. All the device parameters considered for the device simulation are listed in Table 1.

3 Results and Discussion

This section investigates the comparative analysis of both the structures with the help of exhaustive calibrated 2D-TCAD simulation. Figure 2 shows the comparative analysis of the concentration of carriers of both the structures for OFF state for $V_{GS} = 0 \text{ V}$, $V_{DS} = 1 \text{ V}$ and ON state for $V_{GS} = 1 \text{ V}$, $V_{DS} = 1 \text{ V}$ below the gate oxide along the horizontal cut-line of 1 nm. From this figure, it is evident that both the devices have achieved the required concentration profile similar to the conventional doped TFET by electrostatic doping to realize $p^+ - i - n^+$ charge carrier profile even without metallurgical doping. Figure 3, represents the energy band diagrams of both device structures in OFF state for $V_{GS} = 0 \text{ V}$, $V_{DS} = 1 \text{ V}$ and ON state for $V_{GS} = 1 \text{ V}$, $V_{DS} = 1 \text{ V}$ along with the horizontal cut-line 1 nm below the gate oxide. Here, toward source side, energy barrier gap is lower as compared to drain in both the structures. It can be noted that staggered hetero-junctions [26] are formed at the

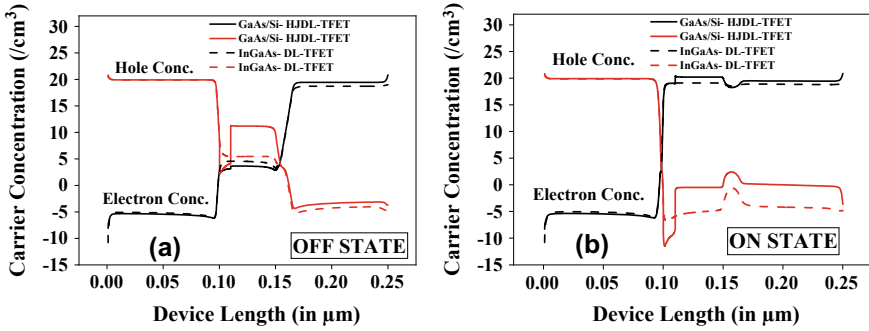


Fig. 2 Carrier concentration of InGaAs-DL-TFET and GaAs/Si-HJDL-TFET a OFF state ($V_{GS} = 0\text{ V}$, $V_{DS} = 1\text{ V}$) b ON state ($V_{GS} = 1\text{ V}$, $V_{DS} = 1\text{ V}$)

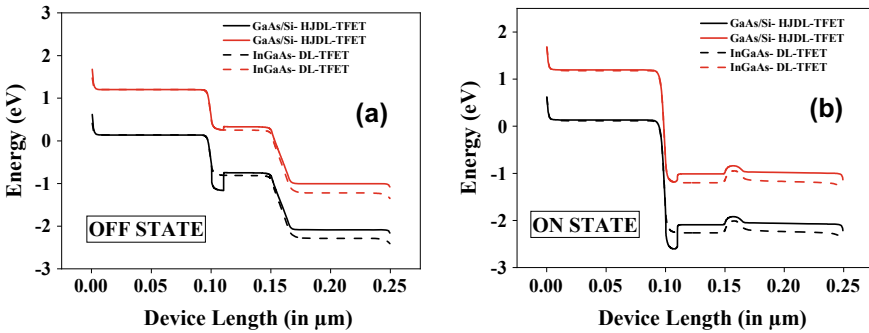


Fig. 3 Energy band diagram of InGaAs-DL-TFET and GaAs/Si-HJDL-TFET a OFF state ($V_{GS} = 0\text{ V}$, $V_{DS} = 1\text{ V}$) b ON state ($V_{GS} = 1\text{ V}$, $V_{DS} = 1\text{ V}$)

source-channel junction of the HJDL-TFET structure, which improves the OFF-state performance of the device.

The transfer characteristic of both the device structures is shown in Fig. 4. The figure depicts that the I_{ON} current of HJDL-TFET and DL-TFET is almost similar around $0.3\text{ mA}/\mu\text{m}$ but the OFF-state current decreases significantly in HJDL-TFET than other structure. It is because of the fact that in HJDL-TFET is having wider energy bandgap toward drain side as shown in Fig. 3a. The simulated performance parameters like SS, I_{ON} and I_{ON}/I_{OFF} ratio of both the structures under study are mentioned in Table 2. It is observed that I_{ON}/I_{OFF} of the proposed structure HJDL-TFET is approximately four times improved than the DL-TFET. The comparative analysis of the proposed structure with the previous works has also been listed in this table.

In Fig. 5a, the energy band diagrams of InGaAs-DL-TFET are shown in.

which as the mole fraction of material $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ increases from $x = 0.55$ to $x = 0.7$ in all the regions of structure results in an enhancement in tunneling width leads to the reduction in the tunneling current but OFF-state current also reduces

Fig. 4 Transfer characteristics of InGaAs-DL-TFET and GaAs/Si-HJDL-TFET

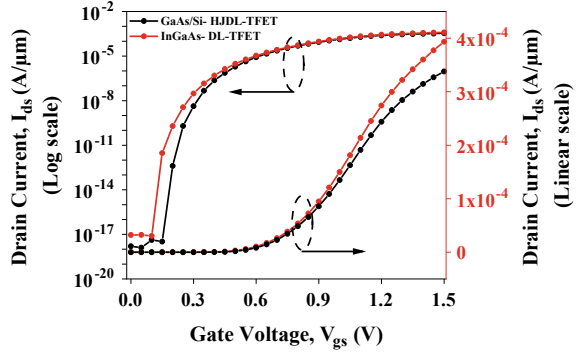


Table 2 Device performance characteristics for InGaAs-DL-TFET and GaAs/Si-HJDL-TFET at $x = 0.75$ and at $V_{GS} = 1 \text{ V}$, $V_{DS} = 1 \text{ V}$

Parameters	D_1	D_2	Sharma et al. [29]	Tripathy et al. [30]
SS (mV/decade)	9.05	9.82	9.78	26
V_{LH} (V)	0.338	0.374	0.519	0.31
I_{ON}/I_{OFF}	2.21×10^{13}	8.48×10^{13}	2.06×10^{11}	7.5×10^{11}

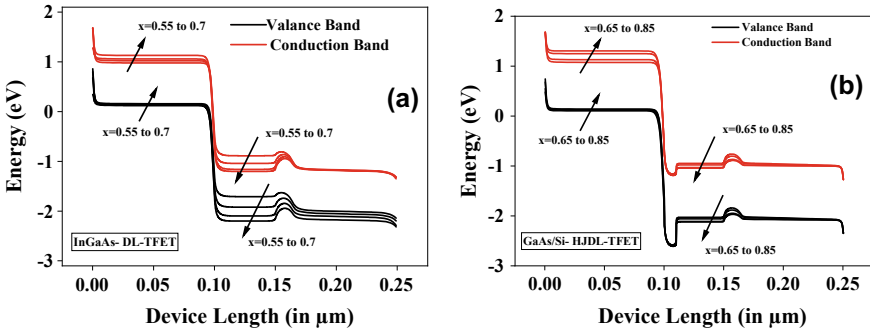


Fig. 5 Energy band diagram of **a** InGaAs-DL-TFET and **b** GaAs/Si-HJDL-TFET with the variation in mole fraction of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$

with increase in the mole fraction. Figure 5b shows the impact on the energy band diagram of GaAs/Si-HJDL-TFET with the variation in mole fraction of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ toward source region from $x = 0.65$ to $x = 0.85$. It can be noticed from the figure that with an increase in the mole fraction x , the tunneling width increases that leads to the reduction in tunneling current and I_{OFF} increases with the increase in mole fraction. In Fig. 6a, b, transfer characteristic of InGaAs-DL-TFET and GaAs/Si-HJDL-TFET has been shown. Here, the variation in the mole fraction of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ for all the regions of former structure and source region for the later structure are considered. The graph shifts toward the right with the increasing value of x . It shows

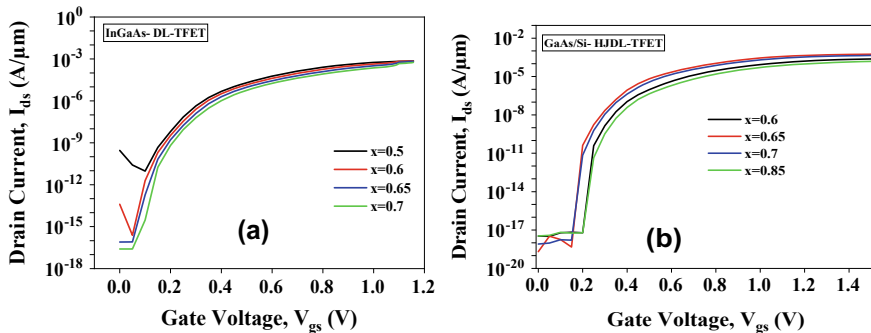


Fig. 6 Transfer characteristics of **a** InGaAs-DL-TFET and **b** GaAs/Si-HJDL-TFET with the variation in mole fraction of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$

the faster switching operation for a higher value of x . Figure 7 shows the impact on the performance characteristics of InGaAs-DL-TFET with the variation in mole fraction of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ in all regions. Figure 7a shows the variation of SS with x and maximum SS as 29 mV/decade has been found for the minimum mole fraction value as $x = 0.5$ and minimum threshold voltage as 0.25 V for the same structure as shown in Fig. 7b and $I_{\text{ON}}/I_{\text{OFF}}$ ratio (1.96×10^6) is lower for $x = 0.5$ as illustrated in Fig. 7e. Figure 8 shows the impact of mole fraction variation of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$ toward the source side on the performance characteristics of GaAs/Si-HJDL-TFET. Figure 8a shows the variation of SS with x and minimum SS as 6.3 mV/decade has been found for $x = 0.65$ and minimum threshold voltage as 0.34 V for the same structure as shown in Fig. 8b. Maximum I_{ON} (2.8 mA/ μm) and minimum I_{OFF} (2.04×10^{19}) have been achieved at $x = 0.65$ as shown in Fig. 8c, d, respectively. A maximum of $I_{\text{ON}}/I_{\text{OFF}}$ ratio is for hetero-junction DL-TFET, i.e., for $x = 0.65$ as 1.38×10^{15} which makes this structure as a significant structure among both structures. Moreover, hetero-junction DL-TFET has also been investigated for its charge sensitivity. Figure 9 shows the variation in the device performance characteristics such as SS, threshold voltage and $I_{\text{ON}}/I_{\text{OFF}}$ ratio. It is to be noted from the figure that this device shows sensitivity toward charged bio-molecules (for positive as well as negative charge), which leads to enhance the suitability of this device structure for the designing of sensor device for charged bio-molecules such as breast cancer cells.

4 Conclusion

Here, two novel TFET device structures, doping-less InGaAs TFET and GaAs/Si-hetero-junction doping-less TFET (GaAs/Si-HJDL-TFET), are analyzed using an exhaustive calibrated TCAD simulation study. These structures are investigated for improving the performance parameters with the variation in mole fraction x of $\text{In}_{(1-x)}\text{Ga}_x\text{As}$. Among both the structures, GaAs/Si-HJDL-TFET with $x = 0.6$

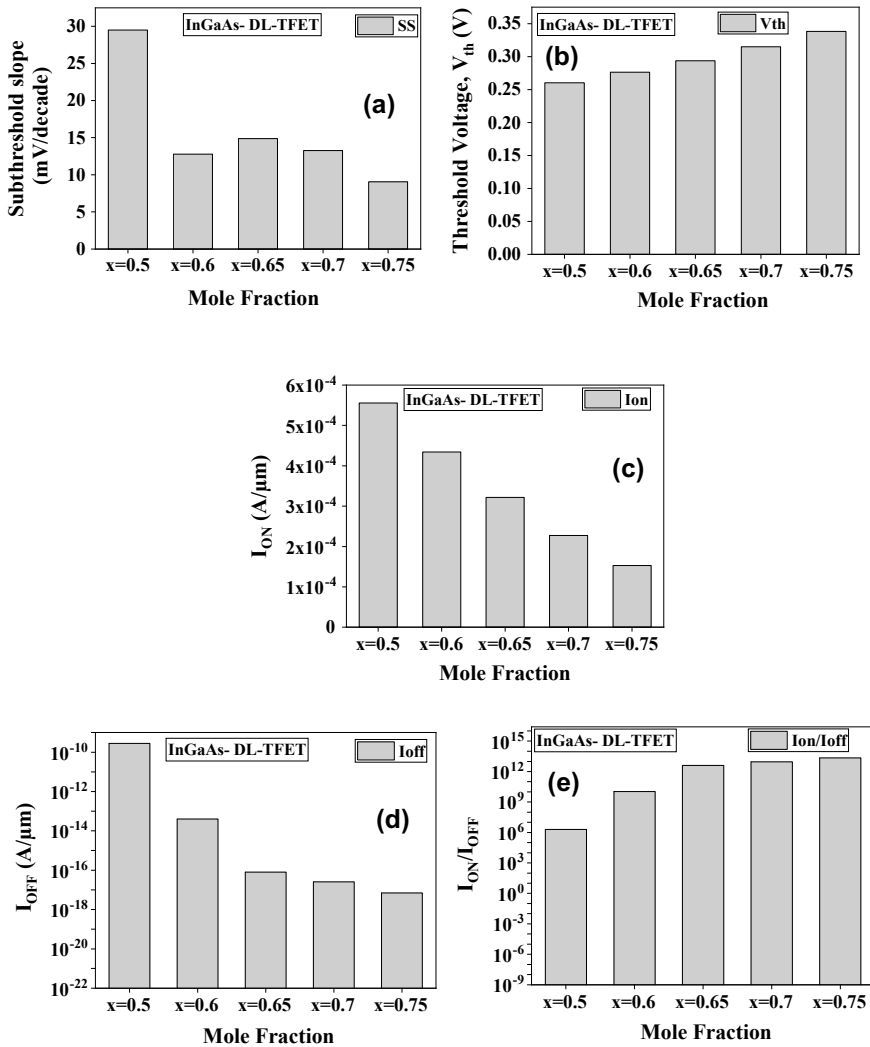


Fig. 7 Performance characteristics of InGaAs-DL-TFET **a** SS **b** V_{th} **c** I_{ON} **d** I_{OFF} **e** I_{ON}/I_{OFF}

shows better performance characteristics such as low threshold voltage (0.33 V), higher I_{ON} (0.2 mA/ μm), lower SS (6.3 mV/decade) and higher I_{ON}/I_{OFF} ratio is of (1.38×10^{15}) Furthermore, hetero-junction DL-TFET structures have also been investigated for charge sensitivity. Therefore, GaAs/Si-based hetero-junction DL-TFET with $x = 0.6$ can be used as an application of the fast switching circuits and charge-based bio-molecules such as C-erbB-2 protein, the breast cancer bio-marker detection device.

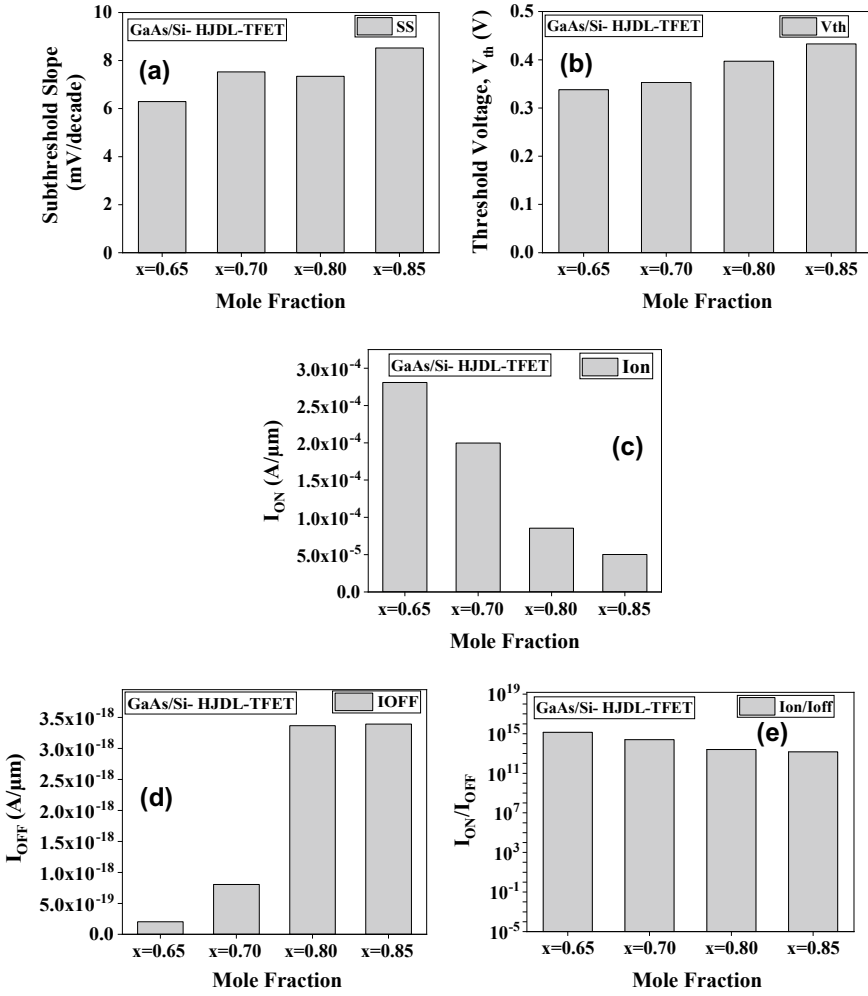


Fig. 8 Performance characteristics of GaAs/Si-HJDL-TFET a SS b V_{th} c I_{ON} d I_{OFF} e I_{ON}/I_{OFF}

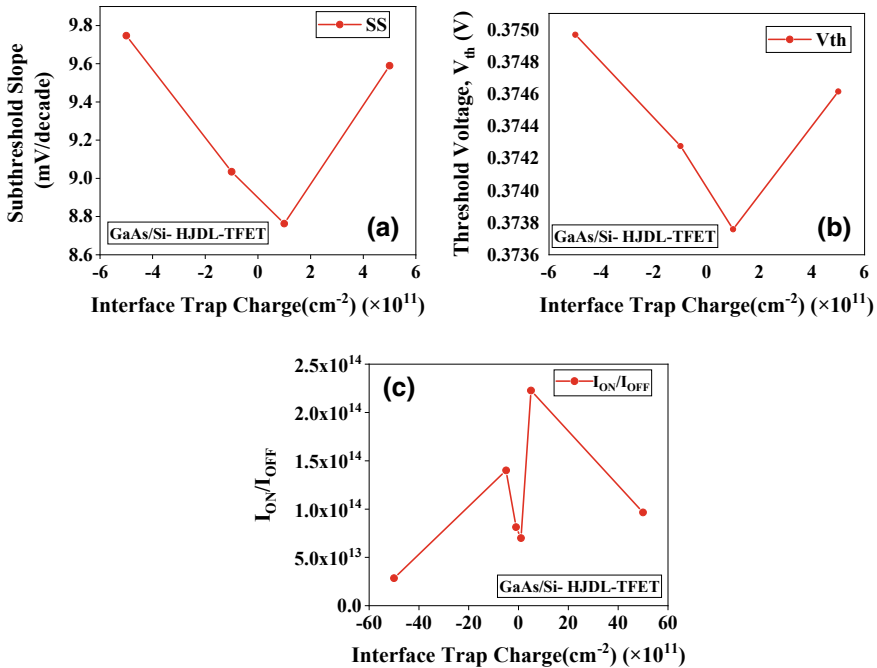


Fig. 9 Performance characteristics of GaAs/Si-HJDL-TFET **a** SS **b** V_{th} **(c)** I_{ON}/I_{OFF} with the variation of interface trap charges

References

1. Arora N (1993) Mosfet modeling for VLSI simulation. World Scientific, Cadence Design System
2. Boucart K, Ionescu AM (2008) A new definition of threshold voltage in tunnel FETs. Solid State Electron 52:1318–1323
3. Seabaugh AC, Zhang Q (2010) Low-voltage tunnel transistors for beyond CMOS logic. Proc IEEE 98(12):2095–2110
4. Tripathy MR, Singh AK, Samad A, Chander S, Baral K, Singh PK, Jit S (2020) Device and circuit-level assesment of Gasb/Si heterojunction vertical tunnel-FET for low-power application. IEEE Trans Electron Dev 67(3):1285–1292
5. Singh AK, Tripathy MR, Chander S, Baral K, Singh PK, Jit S (2019) Simulation study and comparative analysis of some TFET structures with a novel partial-ground-plane (PGP) based TFET on SELBOX Structure, Silicon
6. Ionescu AM, Riel H (2011) Tunnel field-effect transistors as energy-efficient electronic switches. Nature 479:329–337
7. Boucart K, Ionescu AM (2007) Double gate tunnel FET with high-k gate dielectric. IEEE Trans Electron Devices 54(7):1725–1733
8. Saurabh S, Kumar MJ (2011) Investigation of novel attributes of a dual material gate nanoscale tunnel field effect transistor. IEEE Trans Electron Devices 58(2):404–410
9. Taur Y, Wu J, Min J (2015) An analytic model for heterojunction tunnel FETs with exponential barrier. IEEE Trans Electron Devices 62(5):1399–1404
10. Turkane1 SM, Kharate1 GK, Kureshi AK (2017) Ge/Si hetero-junction hetero-gate PNPn TFET with heterodielectric box to improve I_{ON}/I_{OFF} . Indian J Sci Technol 10(14):1–7

11. Kumar MJ, Janardhanan S (2013) Dopingless tunnel field effect high performance design and investigation. *IEEE Trans. Electron Devices* 60(10):3285–3290
12. Zhou J, Han G, Li Q, Peng Y, Lu X, Zhang C, Zhang J, Sun Q-Q, Zhang DW, Hao Y (2016) FerroelectricHfZrOx Ge and GeSn PMOSFETs with Sub-60 mV/decade subthreshold swing, negligible hysteresis, and improved IDS. *IEEE Int Electron Devices Meeting (IEDM)*. doi:<https://doi.org/10.1109/IEDM.2016.7838401>
13. Zhou J, Wu J, Han G, Kanyang R, Peng Y, Li J, Wang H et al (2017) Frequency dependence of performance in Ge negative capacitance PTFETs achieving sub-30 mV/decade swing and 110 mV hysteresis at MHz. *IEEE Int. Electron Devices Meeting (IEDM)*. doi:<https://doi.org/10.1109/IEDM.2017.8268397>
14. Zhou J, Han E, Xu N, Li J, Peng Y, Liu Y, Zhang J, Sun Q-Q (2019) Experimental validation of depolarization field produced voltage gains in negative capacitance field-effect transistors. *IEEE Trans Electron Device*. <https://doi.org/10.1109/TED.2019.2931402>
15. Zhou J, Han G, Xu N, Li J, Peng Y, Liu Y, Zhang J, Sun Q-Q, Zhang DW, Hao Y (2019) Incomplete dipoles flipping produced near hysteresis-free negative capacitance transistors. *IEEE Electron Device Lett* 40(2):329–332
16. Singh S, Pal P, Kondekar PN (2014) Charge-plasma-based super-steep negative capacitance junctionless tunnel field effect transistor: design and performance. *Electron Lett* 50(25):1963–1965
17. Damrongplait N, Kim SH, Liu TJK (2013) Study of random dopant fluctuation induced variability in the raised-Ge-source TFET. *IEEE Electron Device Letter* 34(2):184–186
18. Duan X, Zhang J, Wang S, Li Y, Xu S, Hao Y (2018) A high-performance gate engineered InGaN dopinless tunnel FET. *IEEE Trans Electron Devices* 65(3):1223–1229
19. Chander S, Baishya S (2015) A two-dimensional gate threshold voltage model for a heterojunction SOI-tunnel FET with oxide/source overlap. *IEEE Electron Device Lett* 36(7):714–716
20. Neves FS et al (2016) Low-Frequency noise analysis and modeling in vertical tunnel FETs with Ge. *Source* 63:1658–1665
21. Kumar S, Singh K, Chander S, Goel E, Singh PK, Baral K, Singh B, Jit S (2018) 2-D analytical drain current model of double-gate heterojunction TFETs with a SiO₂/HfO₂ stacked gate-oxide structure. *IEEE Trans Electron Dev* 65(1):331–338
22. Ionescu AM, Riel H (2011) Tunnel field effect transistors as energy efficient electronic switches. *Nature* 479:329–337
23. Wang Q, Wang S, Liu H, Li W, Chen S (2017) Analog/RF performance of L- and U-shaped channel tunneling field-effect transistors and their application as digital inverters. *Jpn J Appl Phys* 56:064102
24. Bhuwalka KK, Schulze J (2005) Scaling the vertical tunnel FET with tunnel bandgap modulation and gate workfunction engineering, *IEEE Trans Electron Dev* 52(5):909–918
25. Mookerjee S, Mohata D, Mayer T, Narayanan V, Datta S (2010) Temperature-dependent I–V characteristics of a vertical In_{0.53}Ga_{0.47}As tunnel FET. *IEEE Electron Device Lett* 31(6):564–566
26. Wang L, Yu E, Taur Y, Asbeck P (2010) Design of tunneling field-effect transistors based on staggered heterojunctions for ultralow-power applications. *IEEE Electron Device Lett* 31(5):431–433
27. Jagadesh Kumar M (2013) Janardhanan, Sindhu: doping-less tunnel field effect transistor: design and investigation. *IEEE Trans Electron Dev* 60(10):3285–3290
28. ATLAS Device Simulation Software, Santa Clara, CA, USA (2014)
29. Sharma M, Narang R, Saxena M, Gupta M (2020) Optimized DL-TFET design for enhancing its performance parameters by using different engineering methods. *IETE Tech Rev* 1–9
30. Tripathy MR, Singh AK, Baral K, Singh PK, Jit S (2020) III-V/Si staggered heterojunction based source-pocket engineered vertical TFETs for low power applications. *Superlattices Microstruct* 142:106494

Impact of Temperature on DC and Analog/RF Performance for DM-DG-Ge Pocket TFET



Kumari Nibha Priyadarshani, Sangeeta Singh, and Alok Naugarhiya

Abstract This work illustrates the effect of temperature on the performance of dual-metal double-gate Ge pocket TFET with hetero-dielectric. The study reveals that the performance of DM-DG-Ge pocket TFET improves with an increase in temperature. Both the DC and analog/RF parameters improve with increase in temperature, whereas I_{ON}/I_{OFF} decreases with an increase in temperature due to higher increase in OFF current than ON current. The improvement in DC and analog/RF parameters makes the device suitable for operation at the higher temperatures.

Keywords TFET · Temperature · Dual gate · Hetero-dielectric

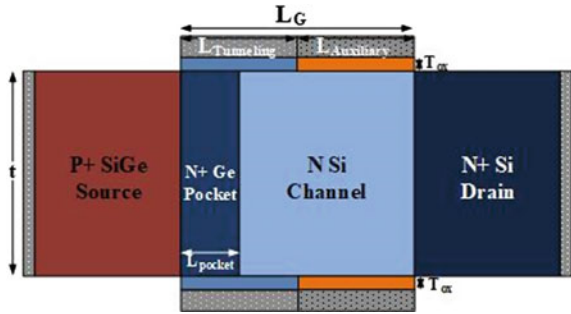
1 Introduction

TFET has gained huge interest of researchers and industry due to its low-power circuit application, low off-state leakage current, low SS and thus fast switching [1–8]. The main agenda of TFET is to achieve low SS as 60 mV/decade is the thermionic transport limitation of MOSFET. The transport mechanism of TFET is completely different from MOSFET, and it works on band-to-band tunneling mechanism removing SS limitation. TFET is also immune to short-channel effects. The limitation of TFET is low I_{ON} and ambipolar current [9]. Researchers have found performance improvement with low band gap material, n^+ pocket region, double gate, double metal gate, etc. [5, 10–12]. Further, with low k dielectric ambipolar current suppression can be achieved. Dual-metal double-gate Ge pocket TFET (DM-DG-Ge pocket TFET) with hetero-dielectric consists of the advantage of both dual-metal double-gate and hetero-dielectric. This work reports the investigation on performance of DM-DG-Ge pocket TFET at higher temperatures. The study has been done for a wide range of temperatures, viz 250–500 K.

K. N. Priyadarshani · S. Singh (✉)
National Institute of Technology Patna, Patna, India
e-mail: sangeeta.singh@nitp.ac.in

A. Naugarhiya
National Institute of Technology Raipur, Raipur, India

Fig. 1 2D device structure for DM-DG-Ge pocket TFET



In this work, DC and analog/RF performance parameter such as V_{TH} , I_{ON} , I_{ON}/I_{OFF} , SS , g_m , C_{gd} , C_{gg} , f_T , GBW, TGF and TFP analysis has been done with variation in temperature for DM-DG-Ge pocket TFET.

2 Device Structure and Simulation models

2D cross-sectional structure of DM-DG-Ge pocket TFET with hetero-dielectric is shown in Fig. 1. The source region is of $Si_{0.5}Ge_{0.5}$, the pocket region is of Ge, and the source and the channel regions are of silicon. With pocket width (L_{pocket}) 3 nm, t_{ox} as 1.2 nm, $L_G = 20$ nm and body thickness (t) is 8 nm. The length of auxiliary gate ($L_{auxiliary}$) and tunneling gate ($L_{tunneling}$) is $L_G/2$. The oxide material used is HfO_2 under tunneling gate and SiO_2 under auxiliary gate. Further, the doping in source region is $N_{source} = 10^{20} \text{ cm}^{-3}$, in pocket region is $N_{pocket} = 4 \times 10^{19} \text{ cm}^{-3}$, in channel region is $N_{channel} = 10^{16} \text{ cm}^{-3}$, and in drain region is $N_{drain} = 10^{20} \text{ cm}^{-3}$. The work function of gate metal for tunneling gate is ($\Phi_{tunn} = 4.6$ eV) and for auxiliary gate is ($\Phi_{aux} = 4.2$ eV). Mo, Ta and W metals can be used for tunneling gate metal, and Ti-Ni and IrO_2 can be used for auxiliary gate metal.

All simulation has been carried out on Atlas Silvaco device simulator [13]. Silvaco solves Poisson's equation with the current continuity equation self-consistently. Non-local band-to-band tunneling has been considered for precise modeling of tunneling. Kane model has also been considered with default value of material parameters. Band gap narrowing, Shockley-Read-Hall recombination and Fermi-Dirac model have also been considered for precise modeling of device. The body thickness of device is 8 nm; hence, quantum confinement effect has not been considered.

3 Result and Discussion

The property of semiconductor material depends on temperature. Thus, operating temperature indirectly influences the performance of device. The dependence of

energy band gap on temperature is as depicted in the following equation:

$$E_g(T) = E_g(0) - \left(\frac{\alpha T^2}{T + \beta} \right) \quad (1)$$

where $E_g(T)$ is the energy band gap at temperature T K. $E_g(0)$ is the energy band gap at 0 K, and α and β are fitting parameters of semiconductor material.

Equation 1 shows that the energy band decreases with increase in operating temperature. The carrier transport of device depends on band-to-band tunneling (BTBT) of electron from valence band of source region to conduction band of channel region. Thus, the drain current can be evaluated by Kane model, given by

$$I_{DS} = D^2 A_{\text{kane}} V_{\text{gs}} E_g^{-1.5} \times \exp\left(\frac{-B_{\text{kane}} E_g^{-1.5}}{V_{\text{gs}} \times D}\right) \quad (2)$$

Here, D , A_{kane} and B_{kane} are constant parameters and E_g is energy band gap of material. So, drain current is indirectly dependent on temperature. With increase in temperature, energy band gap of semiconductor decreases, and with decrease in energy band gap, the drain current increases as depicted in Eq. (2).

3.1 DC Performance Analysis

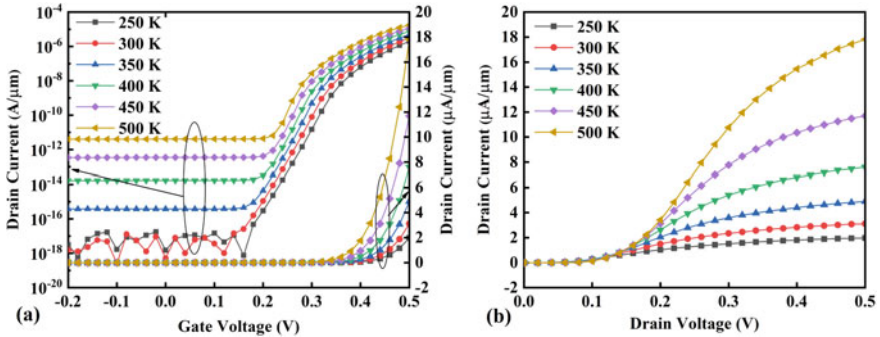
With increase in temperature, the drain current increases. Both I_{ON} and I_{OFF} increase, but the increase in ON current is less in comparison with the increase in OFF current. Thus, the $I_{\text{ON}}/I_{\text{OFF}}$ decreases with increase in temperature. V_{TH} , g_m and SS also improve with increase in temperature. The variation in device parameter with increase in temperature is shown in Table 1. With increase in temperature from 250 to 500 K, the decrease in V_{TH} is 4%, increase in I_{ON} is 799.5%, increase in I_{OFF} is $\sim 10^7$, and decrease in $I_{\text{ON}}/I_{\text{OFF}}$ is $\sim 10^7$. Further, the increase in transconductance (g_m) is 552.7% and improvement in SS is 10.8%.

3.2 Analog/RF Performance Parameter Analysis

Further, the analog and RF parameter analysis with variation in temperature has been done. Figure 2a shows the $I_D - V_G$ at $V_D = 0.5$ V for DM-DG-Ge pocket TFET, and it shows the increase in drain current with V_{GS} variation at different temperatures. The ambipolar current is fully suppressed for the DM-DG-Ge pocket TFET as shown in figure. The increase in $I_D - V_G$ with increase in temperature from 250 to 500 K is 799.5%. Figure 2b shows the $I_D - V_D$ at $V_G = 0.5$ V for DM-DG-Ge pocket TFET. The figure shows increase in drain current with the drain voltage variation at

Table 1 Device characteristic with variation in temperature

Temp (K)	V_{TH} (V)	I_{ON} ($\mu\text{A}/\mu\text{m}$)	I_{OFF} ($\text{pA}/\mu\text{m}$)	I_{ON}/I_{OFF}	g_m ($\text{mS}/\mu\text{m}$)	SS (mV/decade)
250	0.4522258	1.98	5.58×10^7	3.54×10^{12}	41.47	11.07
300	0.449525	3.11	2.56×10^7	1.21×10^{13}	61.62	15.63
350	0.446228	4.89	3.76×10^4	1.30×10^{10}	91.08	17.41
400	0.442189	7.60	1.68×10^2	4.52×10^8	131.49	17.14
450	0.438105	11.69	3.44×10^1	3.39×10^7	188.97	16.06
500	0.434166	17.81	4.02	4.43×10^6	270.61	15.53

**Fig. 2** **a** $I_D - V_G$ at $V_D = 0.5$ V and **b** $I_D - V_D$ at $V_G = 0.5$ V for DM-DG-Ge pocket TFET

higher temperatures. The analog analysis is done at 1 GHz frequency. The C_{gd} and C_{gg} are shown in Fig. 3a, b, respectively. Capacitance values are also increasing with increase in the temperature.

Further, Fig. 4a, b shows g_m and g_d with the variation in temperature, respectively. For analog applications of device in OPAMP, differential amplifier, etc., circuit designer needs device to have higher g_m in order to get higher amplification ($A_V = g_m/g_d = g_m R_o$). The g_m of the device increases with increase in temperature and thus suitable for higher gain. Lesser g_d value is needed for higher gain. However, g_d increases with increase in temperature. This shows that R_o decreases with increase in temperature, giving rise to drain current. Figure 5a, b shows f_T and GBW of device with variation in temperature, respectively. The circuit designer needs the device to work as amplifier for higher frequency. Cut-off frequency f_T is the maximum frequency at which the gain of device becomes unity. After this frequency, the device

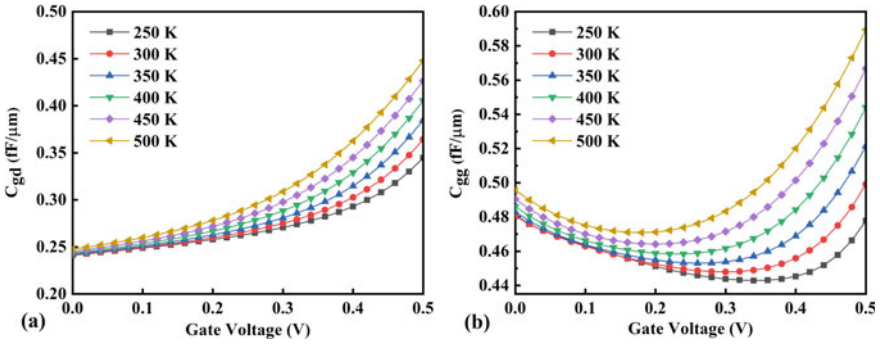


Fig. 3 a C_{gd} and b C_{gg} at $V_D = 0.5$ V for DM-DG-Ge pocket TFET

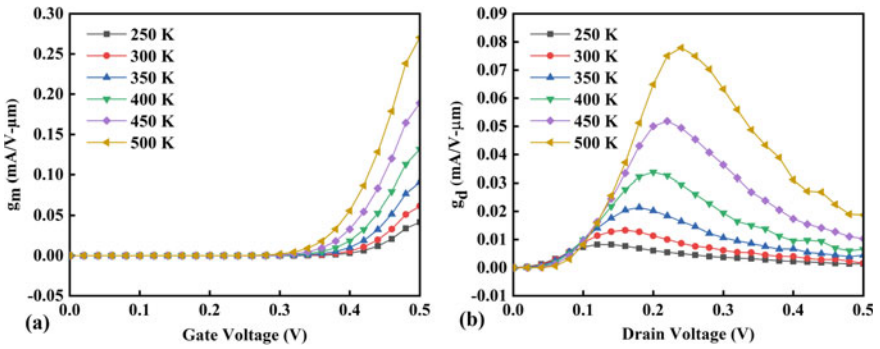


Fig. 4 a g_m at $V_D = 0.5$ V and b g_d at $V_G = 0.5$ V for DM-DG-Ge pocket TFET

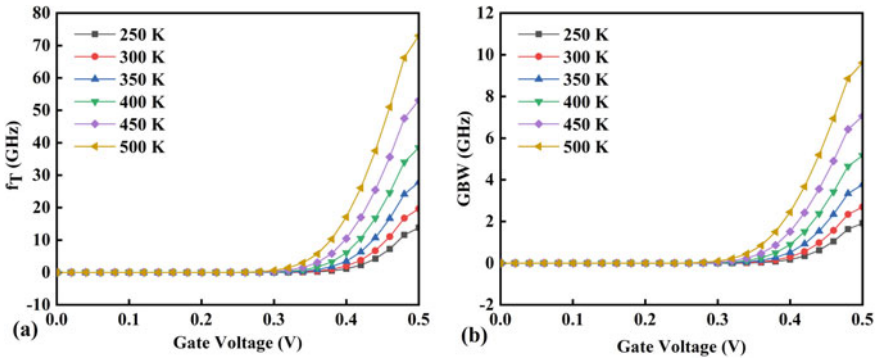


Fig. 5 a f_T and b GBW at $V_D = 0.5$ V for DM-DG-Ge pocket TFET

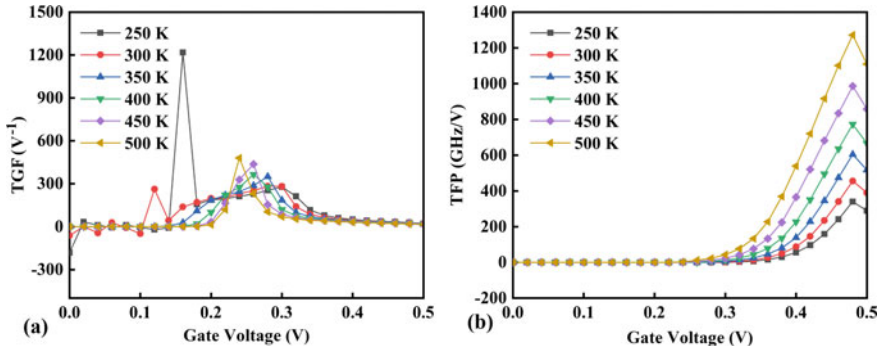


Fig. 6 **a** TGF and **b** TFP at $V_D = 0.5$ V for DM-DG-Ge pocket TFET

could not work as amplifier. f_T is defined as $f_T = g_m/2\pi C_{gg}$. The f_T of DM-DG-Ge pocket TFET increases with increase in temperature. The GBW is the operating frequency and voltage gain product. GBW is defined as $GBW = g_m/2\pi 10C_{gd}$. It is a constant value, for a DC voltage gain of 10; the GBW is shown in the figure. It increases with increase in temperature. Figure 6a, b shows TGF and TFP of DM-DG-Ge pocket TFET, respectively. It is defined as $TGF = g_m/I_D = \ln(10)/SS$. SS value is being low in transition region from ON to OFF. The TGF value achieves its maximum and then again decreases. TFP describes power bandwidth trade-off.

and is important parameter for design of moderate to high-speed circuit. It is defined as $TFP = (g_m/I_D) \times f_T = TGF \times f_T$. TFP also increases with increase in temperature.

4 Conclusion

An exhaustive analysis of temperature sensitivity of DM-DG-Ge pocket TFET has been presented in this work. The study reports excellent improvement in DC parameter as well as analog/RF performance parameters with increase in temperature. The study shows improvement in g_m , f_T , GBW, TGF and TFP with operating temperature increase, thus making it suitable for application in higher temperature ranges.

Acknowledgements The authors are grateful to the National Institute of Technology, Raipur, India, for providing the computational resources.

References

1. Seabaugh AC, Zhang Q (2010) Low voltage tunnel transistors for beyond CMOS logic. Proc IEEE 98(12):2095–2110

2. Ionescu AM, Riel H (2011) Tunnel field-effect transistors as energy efficient electronic switches. *Nature* 479(7373):329–337
3. Ionescu AM, De Michielis L, Dagtekin N, Salvatore G, Cao J, Rusu A et al (2011) Ultra low power: emerging devices and their benefits for integrated circuits. *IEDM Tech Dig*: 1611–1614
4. Nigam K, Kondekar P, Sharma D (2016) DC characteristics and analog/RF performance of novel polarity control GaAs-Ge based tunnel field effect transistor. *Superlattices Microstruct* 92:224–231. ISSN 0749–6036
5. Boucart K, Ionescu AM (2007) Double gate tunnel FET with high-k gate dielectric. *IEEE Trans Electron Devices* 54(7):1725–1733
6. Kumar MJ, Janardhanan S (2013) Dopingless tunnel field effect transistor: design and investigation. *IEEE Trans Electron Devices* 60(10):3285–3290
7. Leung G, Chui CO (2013) Stochastic variability in silicon double-gate lateral tunnel field-effect transistors. *IEEE Trans Electron Devices* 60(1):89–91
8. Shrivastava V, Kumar A, Sahu C, Singh J (2016) Temperature sensitivity analysis of dopingless charge-plasma transistor. *Solid-State Electron* 117:94–99
9. Avci UE, Rios R, Kuhn KJ, Young IA (2011) Comparison of power and performance for the TFET and MOSFET and considerations for P-TFET. In: 11th IEEE international conference on nanotechnology, IEEE, pp 869–872
10. Saurabh S, Kumar MJ (2009) Impact of strain on drain current and threshold voltage of nanoscale double gate tunnel field effect transistor: Theoretical investigation and analysis. *Jpn J Appl Phys* 48(6R):064503
11. Krishnamohan T, Kim D, Raghunathan S, Saraswat K (2008) Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) with record high drive currents and $\ll 60$ mV/dec subthreshold slope. In: IEEE international electron devices meeting. IEEE, pp 1–3
12. Li W, Woo JC (2018) Optimization and scaling of Ge-pocket TFET. *IEEE Trans Electron Devices* 65(12):5289–5294
13. Atlas AUM (2015) Silvaco international. Santa Clara, CA

Control System and IoT

Multidisciplinary Real-Time Model for Smart City Using Internet of Things



Ajitesh Kumar and Mona Kumari

Abstract The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. Nowadays, IoT plays a very important role in the field of the corporate sector, to make smart devices like smart watches, smart houses, and smart city and security purposes. One of the application of smart city which have smart light, smart parking system, smart garbage management system the author will propose in this research paper. Here, IoT works with technology like big data, cloud, and wireless sensors. The combination of these techniques leads to making a smart city. In this proposed model, the smart light will be used for energy consumption, smart parking to monitor the parking traffic, smart garbage management where we reduce the contaminated substances. The author proposes a real-time model that will enhance efficiency.

Keywords IoT · Arduino · NodeMCU · Sensor

1 Introduction

IoT is by far the most important development of the twenty-first century, and it will continue to lead to great things in future. IoT involves the extension of Internet connectivity beyond personal computers and mobile devices. The author explained how IoT is related to big data and cloud computing. All these technologies play a very important role to build the IoT model. Cloud is used to store the data and the useful information generated by the street light and parking system. Big data is used in IoT to improve data storage efficiency. As we know the IoT devices increases day by day, a large amount of data consumption so the big data plays a very important

A. Kumar (✉) · M. Kumari
GLA University, Mathura, UP, India
e-mail: ajitesh.kumar@gla.ac.in

M. Kumari
e-mail: mona.kumari@gla.ac

role to store the data efficiently. One cannot imagine the single automated system without the help of IoT. The application of IoT is very broad that means they are in almost every field. Some of its applications in different sectors are the medical and healthcare sector, transportation sector, education sector; agriculture sector shows its expansion in nature. In this paper, the author has shown how IoT and its applications are used to make the smart city.

In this model, the authors have proposed a multi real-time model. Through this model, with the help of street lights, system power consumption became less because street light opens only when some vehicle is passed from the streets; with the help of garbage management system, a person gets to know when the garbage is full; with the help of smart parking system, person knows which parking slot is empty so they can park their vehicle in that parking slot. A smart city project is useful to the day-to-day life.

All the elements of a smart city model are useful in daily life. The smart street light system is very useful because with the help of this model power consumption is being less and as we all know power is one of the limited sources that we have. In the module smart garbage management system, we use an ultrasonic sensor so it can detect when the garbage is about to fill it sends the data to the cloud and with the help of the Blynk app in our smart phone we got the message that the garbage is about to fill please empty it. In a smart parking system, we use the IR sensor, and it also sends the data to the cloud, and with the help of Blynk app, we get the location where we park the vehicle.

The use of smart materials to demonstrate several functions such as street lighting, smart parking, smart garbage management system, these all, is collectively known as smart cities. To replace the way we look at things is the main purpose of smart cities. Concerning several aspects where IoT is set to rule, we can say that IoT will affect all from the extremely reliable daily movements to the extremely complicated human gesture. Generally, the citizens will avail firstly from the smart city experiments and the fundamental atmosphere. A smart city is based on three components—firstly, street light where there is energy consumption, second, smart parking to monitor the parking traffic and smart garbage management, where we reduce the contaminated substances, and finally, the scope of the smart city is to provide a needful and healthy life for every citizen.

2 Related Work

The author explains that the primary objects of smart cities are based on the value of living, e.g., smart street light, sustainable, e.g., parking management, and quality of life, e.g., garbage management system [1].

The author provides a sketch of smart city prevalence throughout the globe. This model deploys a network of sensors to maintain a healthy environment at optimal level, e.g., to monitor garbage management or to monitor parking space [2]. To maintain this, connect the cloud management platform to the network of Sensors.

This helps to retrieve all the data from the secure platform. But, it takes more time and memory [3]. For better performance in the project in terms of time and memory, the author considered the concept of fog computing in smart city applications and services [4]. We know that fog is the middle level between ground and cloud, and fog computing deals with data. For better dealing with data, instead of sending data to the cloud for processing, it can be processed locally in smart devices [5].

In one survey, the paper presents the structure of exclusive smart city application which is known as logistic mobile application [6]. This application can create a smart location management system which is a future goal point in our project that will help us to easily track vehicles or activities of drivers in smart cities [7, 8]. Smart city must have wireless sensors for better performance. In smart city, the utilization of wireless sensors is described by Rehmani and Rashid [9]. Due to the data-central representation of smart city, this survey is genuinely diverse. These sensors help a lot in the application; we further try these sensors for better achievement in the project [10]. The author works on the concept of data lifecycle in the context of parking systems in smart cities. This data lifecycle is based on product and services. Product is mainly the sensors that are being used with their high capacity and services provided by these products to maintain the parking system bug-free. This data lifecycle helps the parking system to manage large data in an effective custom [11].

3 Proposed Model

A smart city model using IoT provides various functionality like monitoring, controlling, and managing devices in real time. The Internet of Things is all about installing different types of sensors (IR, ultrasonic, etc.) in the different devices for getting real-time data, applying some algorithms and connecting them to the Internet for communications between devices, to monitoring, track, and manage the real-time data.

This model generally focuses on saving time and energy. In a smart city project, real-time data is captured through different sensors and uses various algorithms to provide an optimal solution that helps in saving time and energy. In a smart parking model, IR sensors are used to check whether a vehicle is parked in a parking slot or not. IR sensor collects data and transfers it to the Blynk cloud. So finally, it saves time that wastes in finding empty slots or checking parking slots. Proposed algorithm suggest that, use data that sense or capture by different sensors and show the data in the Blynk app for further analysis.

3.1 Proposed Algorithm

Smart Parking System

1. System initialization
2. Set up Blynk cloud to get data
 - (i) Initialize auth = “ ”
 - (ii) Initialize SSID and password
3. Let widgetLED led1(V1), widgetLED led2(V2)
4. Call Blynk.run() and Blynk.begin(auth, ssid, pass) to set up Blynk.
5. Read the value from the sensor through NodeMCU
6. Call digitalRead(*d1*) to get value
7. Store in *a* and *b* variable.
8. Send results of sensors to the Blynk cloud.
9. If the sensor senses some vehicle, then LED ON otherwise off.

Smart Garbage Management System

1. System Initialization
2. Setup Blynk cloud to get data
 - (i) Initialize auth = “ ”
 - (ii) Initialize SSID and password
3. Let widgetLED led1(V1), widgetLED led2(V2), widgetLED led3(V3) which shows how much the garage is filled.
4. Call Blynk.run() and Blynk.begin(auth,SSID,pass) to set up Blynk.
5. Read the value from the sensor through NodeMCU
6. Makes trigPin low after a delay of 2 s makes triPin high and again after 10-s delay makes trigPin Low.
7. Read echo pin and time in microseconds then calculate distance
8. Send results of sensors to the Blynk cloud.
9. Led in the Blynk app shows how much garbage is filled.
 - I. If red led blink, it means the garbage is filled.
 - II. If green led blinks, it means the garbage is empty.

4 Simulation and Result Analysis

Many nations predict that the population of the World is likely to double. This rise in population generates an adverse effect on the local peoples of the cities, and they have to face different opportunities and challenges. The impacts are like accidents, parking problems, and also a dirty environment. As more population requires more power or energy supply, so we have to consume energy from street light. To overcome these

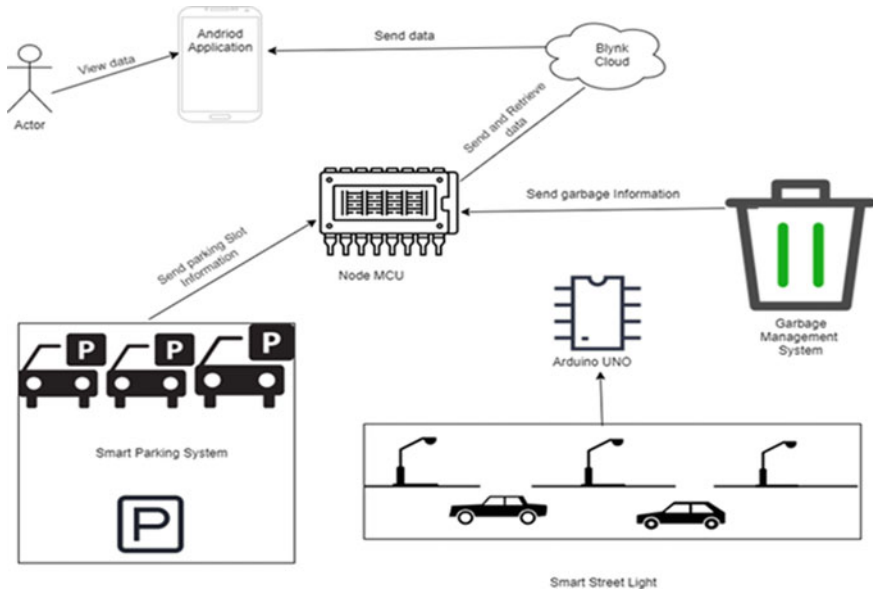


Fig. 1 Showing the working process of a model

types of problems, IoT-based smart city comes into the picture. The contribution of smart city is to make a healthy environment and less energy consumption from different areas. Smart city consists of three main models:

Smart street light.

Smart parking system.

Smart garbage management system

After getting data from IoT sensing devices and by using Arduino and nodeMCU, we send the data to Blynk cloud which shows the following results.

Figure 2 shows the working model of smart street light. So basically, whenever any vehicle will cross that street, the sensor will detect the vehicle, and as a result, the lights around the vehicle will automatically switch to a brighter setting. Figure 3 shows the fluctuation of green and red light. Green light means the parking spot is available, while red light means not available. Figure 4 shows the variation in green, yellow, and red light. As mentioned in figure, also green light shows that the bin is empty there is no garbage it is healthy for the environment. Yellow light shows the middle level that the dustbin is half filled, and red indicates that it is fully filled now and it creates an unhealthy atmosphere. The data obtained from dustbin will be received by the management system, and they take decisions as necessary.

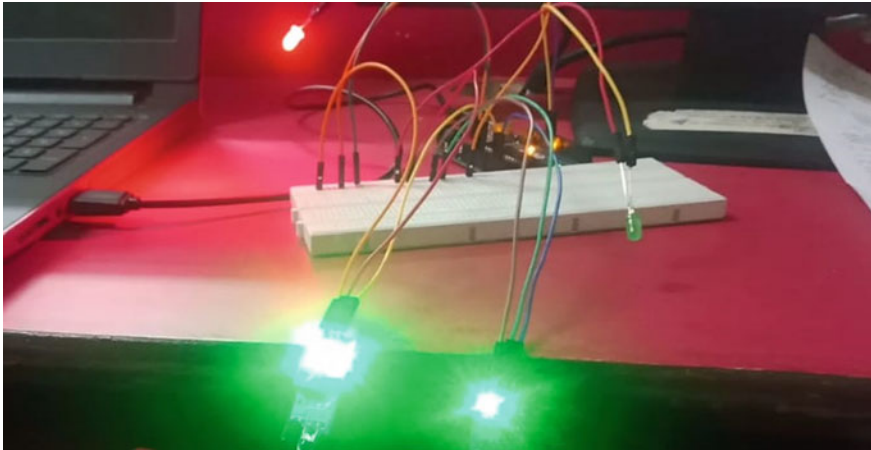
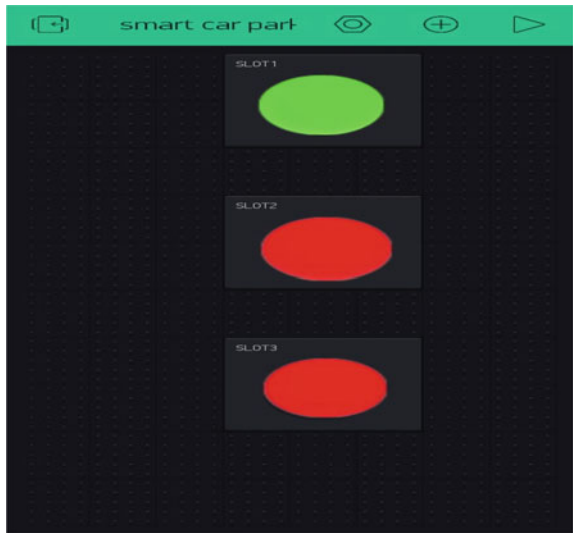


Fig. 2 Showing the working of a smart street light

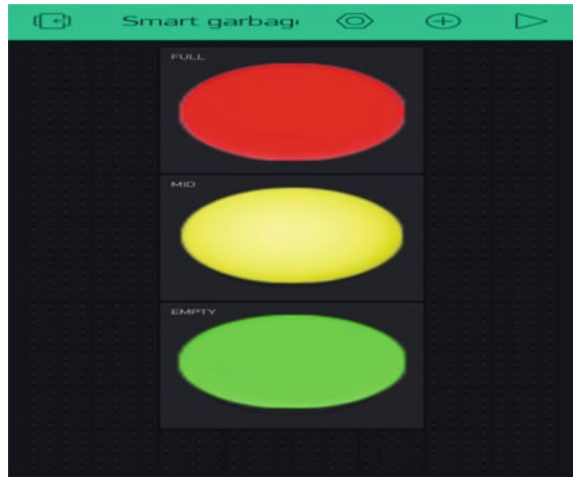
Fig. 3 Smart parking system result



5 Conclusion and Future Direction

IoT helps in every stage whether it is regarding human beings or the environment. As we know that the Government of India is working on many smart city projects. So, IoT is one of the main ingredients of this type of project. IoT provides several types of sensors that are cost-efficient as well as gives the best performance. Nowadays, sensors are commonly used products. Smart city has also used different types of sensors like ultrasonic sensors, IR sensors, etc. Some challenges which occurred

Fig. 4 Smart garbage system result



during the smart city project were to control the range of the sensors, work of the program in-efficient manner, and manage the model. The model solves daily-life problems such as parking slot problem, garbage problem, and so more.

In future, we plan to add some more models in the project such as smart traffic management, smart homes, buildings, and offices. We will plan to solve more challenges of IoT and smart cities. The main problem is security challenges. We will keep in mind about these types of issues as well.

References

1. Taylor L, Richter C, Jameson S, Perez de Pulgar C (2016) Customers, users, or citizens? Inclusion, spatial data, and governance in the smart city. Inclusion, spatial data, and governance in the smart city (June 9, 2016)
2. Bosch P, Jongeneel S, Rovers V, Neumann HM, Airaksinen M, Huovila A (2017) CITYkeys indicators for smart city projects and smart cities. CITYkeys report
3. Liu P, Peng Z (2013) China's smart city pilots: a progress report. *Computer* 47(10):72–81
4. Aldama-Nalda A, Chourabi H, Pardo TA, Gil-Garcia JR, Mellouli S, Scholl HJ, ... Walker S (2012) Smart cities and service integration initiatives in North American cities: a status report. In: *Proceedings of the 13th annual international conference on digital government research*, pp 289–290
5. Langham E, Downes J, Brennan T, Fyfe J, Mohr SH, Rickwood P, White S (2014) Smart grid, smart city, customer research report
6. Ramaswamy R, Madakam S (2013) The state of art: Smart cities in India: a literature review report. *Int J Innov Res Dev* 2(12):115–119
7. Gil-Castineira F, Costa-Montenegro E, Gonzalez-Castano F, López-Bravo C, Ojala T, Bose R (2011) Experiences inside the ubiquitous Oulu smart city. *Computer* 44(6):48–55
8. Huovila A, Penttinen T, Airaksinen M, Pinto-Seppä I, Piira K, Penttinen T (2016) Smart city performance measurement system. In: *Proceedings of the 41st IAHS world congress sustainability innovation for the future, Algarve, Portugal*, pp 13–16

9. Cilliers L, Flowerday S (2014) Information security in a public safety, participatory crowd-sourcing smart city project. In: World congress on internet security (WorldCIS-2014). IEEE, pp 36–41
10. Lövehagen N, Bondesson A (2013) Evaluating the sustainability of using ICT solutions in smart cities—methodology requirements. In: International conference on information and communication technologies for sustainability (pp 175–182)
11. Theodoridis E, Mylonas G, Chatzigiannakis I (2013, July). Developing an IoT smart city framework. In: IISA 2013. IEEE, pp 1–6

Smart IOT-Enabled Battery Management System for Electric Vehicle



Karan Gupta and Vilas H. Gaidhane 

Abstract There is an increased consciousness all over the world toward global warming, and several nations have joined hands to reduce the carbon emissions. The countries like India are aggressively promoting adoption of electric vehicles which significantly cause a reduction in carbon dioxide emission mainly due to ICE vehicles. Lithium-ion battery is one of the major components in an electric vehicle, and life of the battery inside determines the life of an electric vehicle. The paper presented a smart battery management system to prolong the battery life. It monitors and captures various parameters like voltage, current, SOH, SOC, number of cycles, GPS location, etc., and transmits all the data over Internet of Things (IOT) to a user-friendly application through a custom database. This will help user to keep track of their battery status and health all the time during their trips. It also has support for GPS to keep a track of battery location and onsite immediate support be provided in case of full battery discharge.

Keywords Internet of Things · Battery management system · Electric vehicle

1 Introduction

One of the big challenges faced in today's world is global warming which has created demand for cleaner and environmentally friendly fuel/energy. Carbon dioxide emissions from internal combustion engine (ICE) vehicles has been a major contributor in the increasing global warming across several nations [1, 2]. The countries like India have been rated as the most polluted countries in the world. To overcome the challenges, India has become one of the nations to sign the Paris accord which aims to bring about a climate change by adopting various methods to cut down carbon

K. Gupta (✉) · V. H. Gaidhane
Birla Institute of Technology and Science Pilani, Dubai Campus, UAE
e-mail: Karan1605@gmail.com

V. H. Gaidhane
e-mail: vilasgd612@gmail.com

emissions. In view of that lately, there has been a push from the Indian government toward electrification of all existing ICE vehicles. Under the honorable prime minister of India, a committee has been setup by the name of Niti Aayog which is responsible for releasing all incentives, standardization of protocols and adoption of policies related to electrification movement in India. In India, market for electrical vehicle is growing rapidly. The Government of India is promoting transformation from conventional ICE vehicles into electrical vehicles aggressively. The government has released a mandate making 50% of two-wheelers and three-wheelers to be electric by 2025. The manufacturers like Honda, Hero, Bajaj, Micromax have already entered into the electric vehicle (EV) technology and started manufacturing two-wheelers and three-wheeler EV. However, as per the report of Niti Aayog the biggest challenge that these manufacturers will face would be an establishment of fully fledged charging infrastructure across India.

In this paper, a “lithium-ion batteries with IOT-enabled battery management system with communication to charging stations” is presented which mainly focuses on solving the major challenge. This paper intends to provide a universal platform which will enable communication between lithium-ion batteries and the user via a user-friendly mobile application. The universal platform will be linked to our communication device interfaced with the BMS of lithium-ion batteries so users remain updated about their battery health and status all the time.

2 Battery Management System (BMS)

2.1 Smart BMS

The smart lithium-ion batteries have an IOT = based battery management system which is the brains of a lithium-ion battery pack. A BMS is responsible for monitoring peak voltages and surrounding environmental conditions to prolong the battery life.

2.2 Major Applications

Market for EV is growing rapidly in India with a lot of push from the Government of India. An electric vehicle has only dozen parts, whereas a conventional ice has over 200 moving parts causing wear and tear of these parts in just couple of years. A lithium-ion battery constitutes a major part of an electric vehicle and holds around 50% cost of an EV. For instance, if a two-wheeler electric scooter costs around 1 lakh rupees, then the cost of just the lithium battery used will be around 50 thousand rupees. Therefore, life and reliability of a lithium-ion battery used in an electric vehicle plays a vital role in determining the quality and life of an electric vehicle [3].

2.3 Need for Battery Management System in Electric Vehicle

Wei et al. [4] have discussed in their research about various issues associated with lithium-ion batteries used in EVs and smart grid systems. Even though lithium-ion batteries promise a greater number of charging and discharging cycles and power density over conventional lead acid batteries, they can undergo capacity degradation over time if peak voltages and surrounding temperature are not monitored time to time. A standard lithium battery capacity falls down to 80% of its original capacity once the battery finishes the total number of cycles declared by the manufacturer. The overcharging beyond permissible values or discharging below the allowed voltage values simply catalyzes the capacity degradation. Moreover, self-discharge is another issue associated with lithium-ion batteries which usually occurs when the battery is not in use. Surrounding temperature can result in increased discharge rate if not monitored properly. For a typical lithium iron phosphate (LFP) and nickel cadmium manganese (NCM) cell, the surrounding temperature should not exceed 45 °C. Therefore, BMS plays a vital role in thermal control as well by regulating current drawn limiting the power dissipated.

At a time, unbalanced cells in a battery pack can create issues, for instance, in a pack of battery pack having four cells with voltages 3.5, 4, 4 and 4 V. During charging, 3.7 V will be charged up to 3.7 V, whereas the remaining three will be charged up to 4.2 V, thereby making the 3.5-V cell to be the one first one to discharge compromising the full potential of entire battery pack. BMS has another functionality which ensures cell balancing [5].

2.4 Features of BMS

In this paper, the battery management system (BMS) has plethora of features incorporated. First of all, it supports charging and discharging control to monitor peak voltages. It is achieved in two stages, constant current and constant voltage. During constant current stage, constant supply of current is fed to the cells present till all cells are nearly full. Later, in the second stage, constant voltage is supplied making sure peak voltages do not cross the permissible values [6]. The presented accurate battery management system determines state of charge (SOC) of battery which follows a simple mathematical method called coulomb counting method. It is represented by a simple formula given as

$$\text{SOC} = \frac{\text{Total Current fed}}{\text{Total Capacity of the Battery}} \quad (1)$$

There are various other methods adopted to calculate the SOC, namely Ampere hour method, open-circuit voltage (OCV) method, machine learning algorithm

like extended Kalman filter and state-space model estimation using Kalman filter. However, it requires lot of computations [7].

It also features the determining state of health (SOH) which gives an estimate about the remaining cycles and expected age of the battery. Cell balancing is another feature supported by the battery management system. There are mainly two types of cell balancing, passive and active. In active cell balancing, cells which have higher potential will be used to charge the ones with lower potential so that voltages across all the cells are equalized, whereas in passive cell balancing with the aid of an external load resistor cells with excessive voltage are forced to discharge across these load resistors to equalize voltages across all the cells present [8, 9].

The BMS has multiple onboard thermistors (negative temperature coefficients) for external sensing of temperature. These sensors mainly help in determining the temperature of the lithium-ion battery pack so the temperature does not exceed 45 °C which is the typical maximum permissible temperature value for any lithium-ion battery pack.

The presented BMS is an IOT based which allows logging of various parameters of battery wirelessly to monitor status of the battery remotely. It is linked to database of Google firebase where all the parameters are stored which are then acquired by our front-end platform built on html, CSS and JavaScript, Node JS, express and MongoDB to perform analytics and display to the users. The BMS will also have ability to communicate with charging station battery swapping kiosks in close vicinity whenever the battery is running out of charge. The users can also pinpoint the location of the nearest available battery swapping kiosks on a user-friendly application.

3 Smart IOT Battery Management System Flowchart

Figure 1 shows the internal block diagram of battery management system. The analog front-end chip on the BMS will measure all the parameters mainly individual cell voltages, total current drawn using a high-value shunt resistor, and ensure cell balancing and distribute power to the load using an inbuilt power distribution unit (PDU). The parameters measured by the AFE are transmitted via serial communication protocol to the IOT module. The IOT module has MC60 GPRS/GPS chipset supporting low power 2G network. Figure 2 depicts the overall flow chart of smart BMS system.

The hardware implementation of the system mainly consists of the BMS module and a separate IOT module for cloud-based communication over TCP/IP protocol where the data is analyzed and stored in MongoDB database from Server 1 application. The processed and analyzed data is then sent to client-side applications, mainly Webapp, mobile app and CRM via Server 2. The Server 2 is also responsible for SOC estimation based on the parameters stored in MongoDB. Client-side applications can send data back to the BMS to perform.

Internet of Things (IoT) is a type of data communication that is performed and delivered wirelessly. The chipset used as an IoT is Quectel MC60 which is a quad-band full-featured GSM/GPRS/BLE module. The module supports several Internet

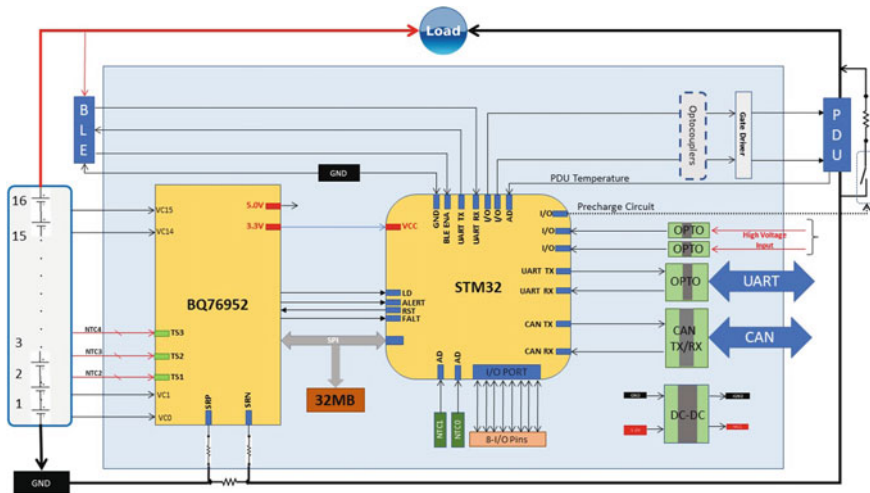


Fig. 1 Internal block diagram of battery management system

protocols Transfer Control Protocol (TCP), User Datagram Protocol (UDP), Point-to-Point Protocol (PPP), File Transfer Protocol (FTP), Hypertext Transfer Protocol (HTTP) and Secure Socket Layer (SSL). Based on the latest 2G module, it has a better performance in the case of SMS and data transmission frequency as well as audio service even in tough environments. It features dual SIM single standby function [10]. Figure 3 depicts how the data is transmitted among various devices.

The inbuilt low noise amplifier (LNA) [11] supports the module with improved radio frequency sensitivity and exceptional data acquisition performance or tracking performance even in weak signal areas. The close-packed form factor (FF), radiant positioning performance, low quiescent current (Qi) make MC60 an ideal choice for a wide range of machine-to-machine applications, such as automotive, telematics, handheld devices and logistics tracker [12, 13].

The GSM/GPS module is used in monitoring and managing an EV battery. Android application is used for an operation. The operating system can inform the user with a notification, SMS or e-mail. However, it always needs an active Internet connection or at least GPRS Internet connection in that place so that the device is ready to go online to suit user's needs. Furthermore, the system has a lot of backend processes that cause the wastefulness of batteries.

A battery management system is necessary for the safety of the battery. Several notions cause the breakdown of the battery such as dilapidation of battery and system failure. The prominent functionality of the BMS is to protect the battery from overcharging, deep discharging, overloading, under temperature, over-temperature and short circuit.

The proposed BMS incorporates a low-power analog front-end chip (AFE) to precisely monitor individual cell voltage values and current drawn and supports passive cell balancing. The AFE chip is controlled by the microcontroller chip STM32

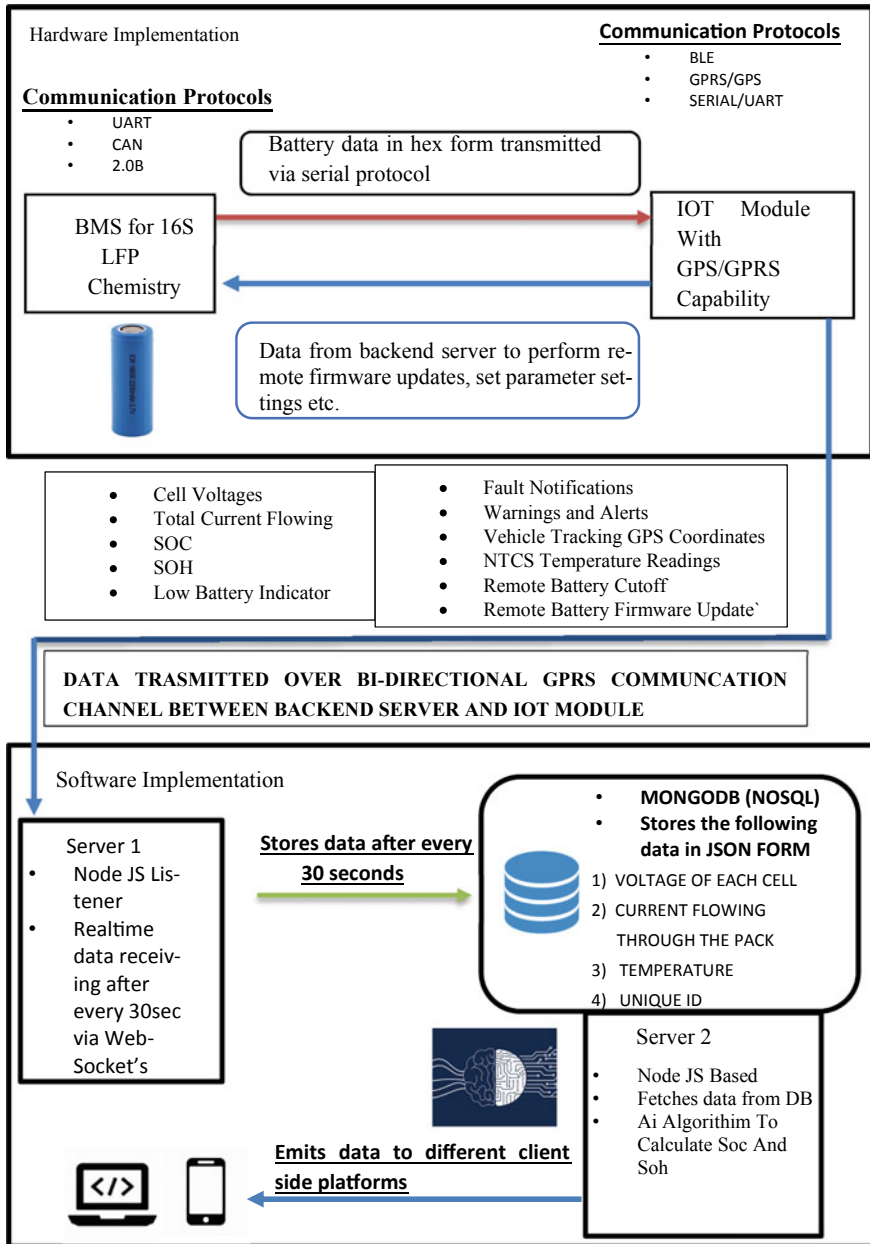


Fig. 2 BMS with IOT integration flowchart

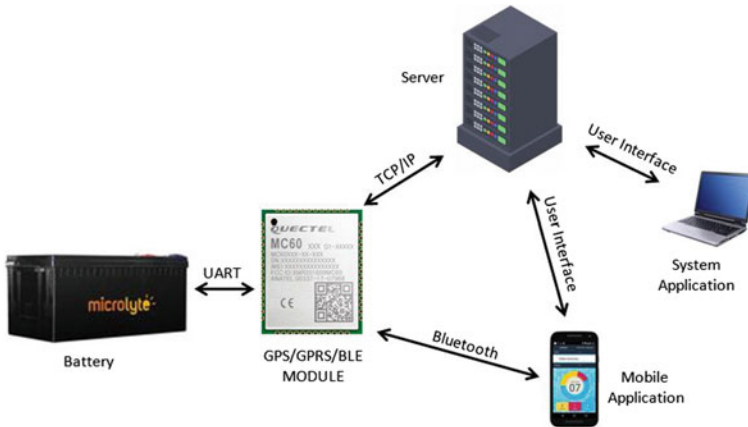


Fig. 3 IOT flowchart

which further communicates with the GPS/GPRS module MC60 over serial protocol. The battery management system ensures that all the cells in the battery are at the same state of charge (SOC) making the battery run at the full capacity. Therefore, it is important to remotely monitor the battery parameters using a wireless system. The data that is usually monitored is the minimum and maximum voltage, maximum current limits, minimum and maximum temperature for thermal management status, etc.

The BMS gives an ability to track, record and analyze how battery is performing in real time. These features lead to improvements in transit service through better on-time performance and quicker response time to emergencies. The location information along with other details such as the speed of the vehicle, the route followed, etc., is improving the efficiency of vehicle operation by generating various standard and exception reports (Fig. 4).

Some of the major parameters monitored and remotely transmitted to an android-based application or Webapp are as follows:

- Single cell voltage
- Current
- State of charge (SOC)
- Charge or discharge state
- Cycles
- Temperature
- Number of cycles left
- Remaining capacity (RMC)
- Design capacity (DCAP)
- Full charge capacity
- Average time to empty (ATTE)
- Serial number (SN)

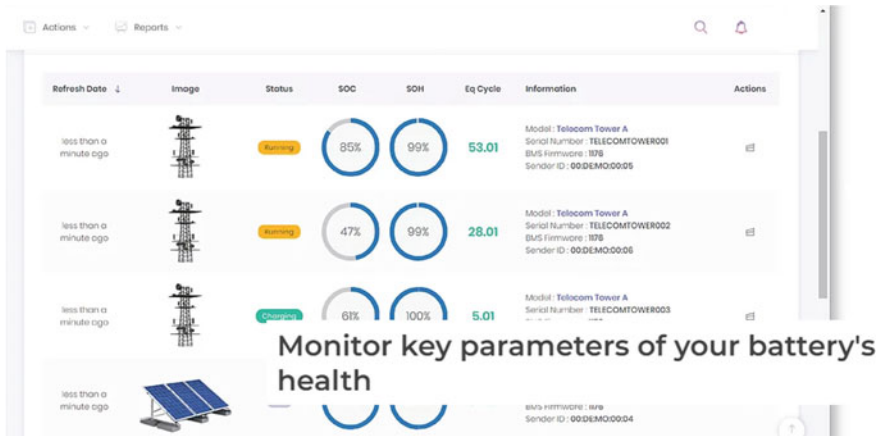


Fig. 4 Client-based Webapp application for IOT-based battery management system

- Manufacturer name
- Communication frequency
- Manufacture date
- GPS coordinates.

The presented smart BMS can find extensive application in electric two-wheelers and passenger three-wheelers battery pack with battery swapping support. The Webapp/mobile application is linked to the MongoDB database hosted on Amazon web services which stores all the relevant information and parameter including GPS coordinates, SOH, SOC, temperature conditions, lifecycles remaining. Our intelligent AI server engine extracts all this information from custom database to perform analysis. It helps to predict the time to complete discharge and remaining battery capacity. Users are auto-directed to the nearest available swapping station/charging stations whenever the battery capacity drops down to 10% of the total capacity on an interactive user-friendly application. The presented intelligent AI algorithm is precisely able to precalculated and analyze all the parameter so the user never runs out of battery at any point of time. The AI engine analyzes the data to model it against real-time digital simulation of the battery and charts the future course by automatically updating its configuration. The WebApp displays the current status of deployed status along with automated reporting of analytics and insights on battery performance.

4 Experiment and Analysis

This section discusses the experiment and analysis conducted using the presented BMS system with 15 series Li-ion battery pack, and voltage values for each cell in

Table 1 Voltage readings of 15 series Li-ion battery pack

Cell No.	BMS AFE voltage (V)	Voltage (V)	Accuracy (%)
1	3.33	3.38	98.52
2	3.34	3.37	99.109
3	3.332	3.35	99.46
4	3.334	3.40	98.05
5	3.338	3.35	99.64
6	3.566	3.58	99.60
7	3.564	3.57	99.831
8	3.558	3.56	99.94
9	3.43	3.44	99.709
10	3.508	3.51	99.94
11	3.568	3.57	99.94
12	4.365	4.365	100
13	4.444	4.5	98.75
14	3.333	3.350	99.49
15	3.18	3.19	99.6

the battery pack measured by BMS AFE chip and readings have been shown in Table 1 to estimate an average accuracy.

Figure 5 depicts the screenshot of custom-tailored application developed in JavaScript and Node JS hosted on a cloud-based server displaying the values received from the GPRS module MC60 module interfaced with the BMS over TCP/IP protocol. The BMS communicates with the module over serial protocol to transmit

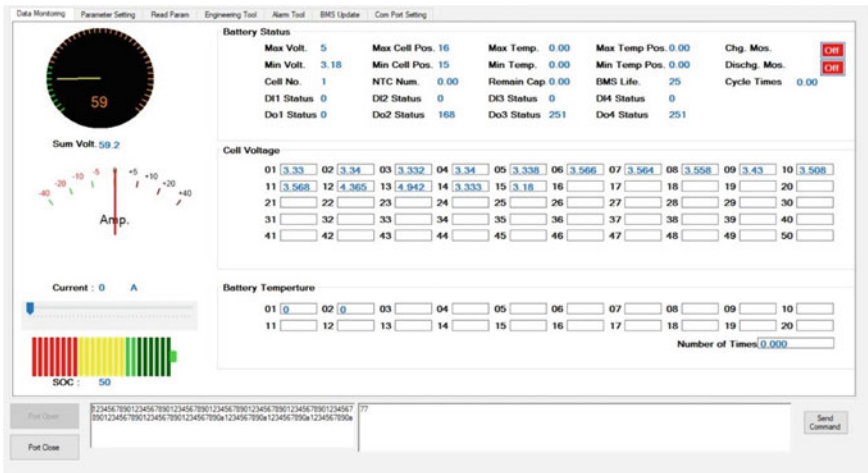


Fig. 5 Server-based application

parameter values like individual cell voltages of all the 15 cells connected in series of the Li-ion battery pack, total current drawn, onboard temperature sensor values and other which are then emitted to the custom-tailored server application. The server application incorporates a coulomb counting algorithm to estimate SOC of the battery.

Table 1 depicts the value of individual cell voltages received by the server application tabulated against the values manually measured by a multimeter to estimate accuracy percentage. The average accuracy comes out to be 99.57%. Therefore, it can be concluded the voltages transmitted from the BMS to the cloud-based server application are valid measurements.

5 Conclusion

In this paper, a smart battery management system for lithium-ion batteries which finds an extensive application in electric vehicles is presented. It determines various parameters like SOH, SOC, etc., and offers various kinds of protection like short circuit and over discharge protection. The BMS supports cloud-based communication using an IOT module which is based on MC60 GPRS/GPS Quectel chipset ensuring an uninterrupted flow of data after every 30 s. All the data is logged over the cloud into custom database which can be displayed on a user-friendly Webapp or application. Users can keep track of their battery status and view details like GPS location, battery health and age of the battery. Mobile application or Webapp can help user remotely access all the information about their battery and receive reminders like battery discharge and battery degradation notifications.

References

1. Duryea S, Islam S, Lawrance W (1999) A battery management system for standalone photo-voltaic energy systems. In: Conference record of the IEEE industry applications conference. IEEE, Phoenix, USA, pp 2649–2654
2. Siguang L, Chengning Z (2009) Study on battery management system and lithium-ion battery. In: International conference on computer and automation engineering. IEEE, Bangkok, Thailand, pp 218–222
3. Wen J, Jiang J (2008) Battery management system for the charge mode of quickly exchanging battery package. In: IEEE vehicle power and propulsion conference. IEEE, Harbin, China, pp 1–4
4. Wei C, Benosman M, Kim T (2019) Online parameter identification for state of power prediction of lithium-ion batteries in electric vehicles using extremum seeking. *Int J Control Autom Syst* 17(11):2906–2916
5. Ouyang Q, Chen J, Xu C, Su H (2016) Cell balancing control for serially connected lithium-ion batteries. In: 2016 American control conference (ACC), Boston, MA, USA, pp 3095–3100
6. Ali MU, Zafar A, Nengroo SH, Hussain S, Junaid Alvi M, Kim HJ (2019) Towards a smarter battery management system for electric vehicle applications: a critical review of lithium-ion battery state of charge estimation. *Energies* 12(3):446

7. Movassagh K, Raihan SA, Balasingam B (2019) Performance analysis of coulomb counting approach for state of charge estimation. In: 2019 IEEE electrical power and energy conference (EPEC). IEEE, Montreal, Canada, pp 1–6
8. Liu K, Li K, Peng Q, Zhang C (2019) A brief review on key technologies in the battery management system of electric vehicles. *Front Mech Eng* 14(1):47–64
9. Wang Z, Ma J, Zhang L (2017) State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the Gaussian process regression. *IEEE Access* 5(1):21286–21295
10. Mcgibney A, Rea S, Ploennigs J (2016) Open BMS—IoT driven architecture for the internet of buildings. In: 42nd annual conference of the IEEE industrial electronics society. Florence, Italy, pp 7071–7076
11. Yu Y, Liu H, Wu Y, Kang K (2017) A 54.4–90 GHz low-noise amplifier in 65-nm CMOS. *IEEE J Solid-State Circ* 52(11):2892–2904
12. Gaidhane VH, Mir A, Goyal V (2019) Energy harvesting from far field RF signals. *Int J RF Microwave Comput Aided Eng* 29(5):e21612
13. Mir A, Gaidhane VH (2017) Deriving energy from far field RF signal. In: IEEE international conference on electrical and computing technologies and applications. IEEE, Ras Al Khaimah, United Arab Emirates, pp 1–4

Admittance Measurements for Structural Health Monitoring in Metal Plates



T. Jayachitra  and Rashmi Priyadarshini 

Abstract Structural Health Monitoring (SHM) in civil and mechanical engineering fields plays a vital role in the current scenario. Electro Mechanical Impedance technique (EMI) is the most common method used in structural health monitoring. The impedance chip is small and low cost device used for structural health monitoring. In this paper, experiments are performed on a metal plate. Holes are induced in the metal plate, and metal bolts are used to fix the metal sheet. Piezo sensors (lead zirconate titanate patches) are used to detect the damage. Sensors are bonded to the metal plates. New metal bolt and corroded bolt are used to fix the metal sheet. The outputs are measured by using impedance chip. The bolts are removed, and the corresponding output is also measured using impedance chip. Comparison of conductance, admittance and susceptance are performed with the usage of new bolt, corroded bolt and removal of bolt. The results confirm the presence of damage in using the corroded bolt and as well as in the removal of bolt.

Keywords Structural health monitoring · Electromechanical impedance · Admittance · Conductance · Susceptance · Lead zirconate titanate patches

1 Introduction

Structural health monitoring (SHM) is widely used in civil, mechanical and aerospace structures. Damage detection is performed in four levels: detection, location, severity and remaining existence of the structure [1]. The most common method used in SHM is EMI technique. EMI techniques are based on piezoelectric property. EMI technique was first proposed by Liang et al. [2]. Damage detection was performed in different

T. Jayachitra (✉)

Electrical and Electronics Engineering Department, School of Engineering and Technology, Sharda University, Greater Noida, Uttar Pradesh 201310, India
e-mail: jayachitra.kishor@sharda.ac.in

R. Priyadarshini

Electronics and Communication Department, School of Engineering and Technology, Sharda University, Greater Noida, Uttar Pradesh 201310, India
e-mail: rashmi.priyadarshini@sharda.ac.in

structures like aluminum, concrete, etc., using EMI technique [3]. Any mechanical change of impedance in the structure is measured by the electrical impedance change and vice versa [4]. Usually, impedance and admittance are measured using impedance analyzers and LCR meter which is very bulky. Wireless sensor network system was proposed to measure the impedance which was complex circuitry [5]. The external low frequency vibration also plays a major role in predicting the damage in the structures [6]. The impedance chip is used for SHM due to its cost effectiveness and miniaturized size. Piezo sensors are bonded to the structure, and excitation is provided by the impedance chip. Detection of damage is performed in concrete beam and aluminum strip [7]. Piezo patch acts both as sensors and actuators [8]. For railway joints, bolts are used in track connection. Detection of damage was performed in bolted joint structure [9].

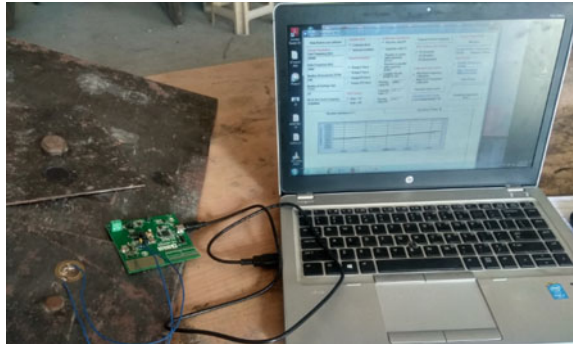
Damage detection is carried in an aluminum plate using piezo sensors by analyzing the output voltage using Ansys software [10]. Damage detection is monitored in bolted connections using signal response energy method and time reversal method [11, 12]. Bolts are unfastened, and the corresponding damage is identified using LibSVM [13]. Damage detection is performed in steel structure by loosening the bolts [14]. Impedance method is utilized, and comparison is made between the techniques of steady state and transient state and found steady state is accurate [15].

Impedance chip is an impedance converter which has an analog to digital converter, frequency generator, digital signal processor and a sensor to measure the temperature. Excitation is given by the frequency generator, and it is further processed by digital signal processing. At every sweep of frequencies, the real and imaginary data is generated. The clock frequency is 16 MHz for the impedance chip. Impedance chip is run by user defined software. Calibration of 200- Ω resistance is required at all times when it is restarted. Repeatability is required for verification of data obtained [16].

2 Experimental Analysis

The metal plate (mild steel) is chosen for experimentation. The size of the metal plate is $550 \times 400 \times 2$ mm. Hole is drilled in the metal plate, and metal sheet is used to cover the hole with the help of bolts. The usages of different bolts are taken into consideration for damage detection in the metal plate. Comparisons of measurements are done by using new bolt and corroded bolt. The piezo sensor is bonded in the metal sheet using epoxy resin. The new bolt is taken and tightened with the metal sheet. The output is noted with the help of impedance chip and considered as the base measurements. Then, the corroded bolt is used to fix the metal sheet, and the corresponding output is measured using impedance chip. The bolt is removed, and the output is measured under full damaged condition using impedance chip. Figure 1 shows the experimental setup of damage detection of metal plate with new bolt and corroded bolt.

Fig. 1 Experimental setup of metal plate with new bolt and corroded bolt



3 Results

The admittance, conductance and susceptance are measured in both the cases with new bolt and corroded bolt. Also, the bolt is removed, and the corresponding output is measured. Impedance chip is used for the measurement of these values. The comparison curves are shown in the following figures. Comparison is made between new bolt, corroded bolt and removal of bolt.

Figure 2 shows the comparison curves of conductance with new bolt, corroded bolt and removal of bolt. The conductance value decreases as the damage level is increased. The conductance value is reduced in both the cases of corroded bolt and removal of bolt.

Figure 3 shows the admittance signatures of different bolts. The admittance value also decreases as the damage detection is more in the case of corroded bolt and the removal of bolt.

Figure 4 shows the susceptance curves of different bolts. The curve clearly identifies that the susceptance value decreases when the bolt is removed and of the corroded bolts.

Fig. 2 Conductance signatures of different bolt

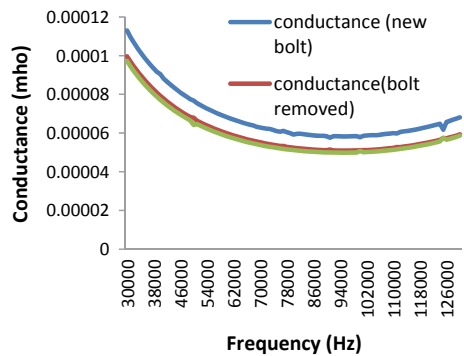


Fig. 3 Admittance signatures of different bolts

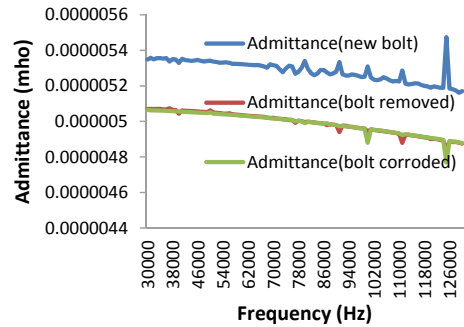
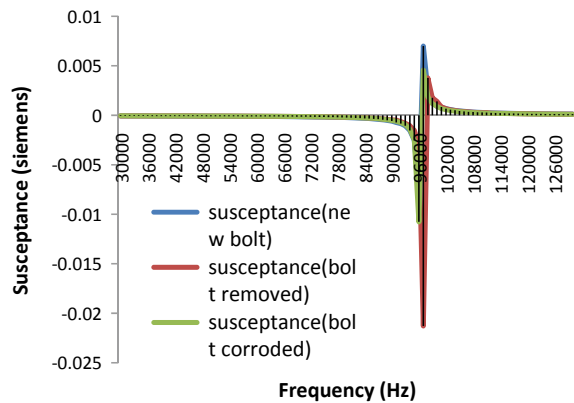


Fig. 4 Susceptance signatures of different bolts



4 Conclusion

This paper clearly indicates the change in their values as the damage takes place when the bolts get corroded or removed due to various reasons. Therefore, damage detection is reliable by using impedance chip which is very less cost and small size compared to impedance analyzer and LCR meter. Damage in any material or structure is dependent on temperature which can be detected by using impedance chip having temperature measurement too. The proposed method does not measure impedance rather measures admittance, conductance and susceptance.

References

1. Shanker R (2009) An integrated approach for structural health monitoring. Thesis
2. Liang C, Sun FP, Rogers CA (1997) Impedance method for dynamic analysis of active material systems. *J Intell Mater Syst Struct* 8(4):323–334
3. Annamdas VGM, Radhika MA (2013) Electromechanical impedance of piezoelectric transducers for monitoring metallic and non-metallic structures: A review of wired, wireless and energy-harvesting methods. *J Intell Mater Syst Struct* 24(9):1021–1042

4. Cortez NE, Filho JV, Baptista FG (2013) A new microcontrolled structural health monitoring system based on the electromechanical impedance principle. *Struct Health Monit* 12(1):14–22
5. Cortez NE, Filho JV, Baptista FG (2015) Design and implementation of wireless sensor networks for impedance-based structural health monitoring using ZigBee and global system for mobile communications. *J Intell Mater Syst Struct* 26(10):1207–1218
6. Priya CB, Reddy AL, Rao GVR, Gopalakrishnan N, Rao ARM (2014) Low frequency and boundary condition effects on impedance based damage identification. *Case Stud Nondestruct Testing Evol* 2(1):9–13
7. Kaur N, Bhalla S, Shanker R, Panigrahi R (2015) Experimental evaluation of miniature impedance chip for structural health monitoring of prototype steel/RC structures. *Exp Tech Soc Exp Mech* 1–12
8. Wang D, Song H, Zhu H (2014) Embedded 3D electromechanical impedance model for strength monitoring of concrete using a PZT transducer. *Smart Mater Struct* 23(11)
9. Smantaray SK, Mittal SK, Mahapatra P, Kumar S (2018) An impedance-based structural health monitoring approach for looseness identification in bolted joint structure. *J Civ Struct Health Monit* 8(5):809–822
10. Thomas AM, Pradeep KR, Mathew P (2016) Structural health monitoring using piezoelectric sensors and actuators. *Appl Mech Mater* 857:255–260
11. Tao W, Shaopeng L, Junhua S, Youroong L (2016) Health monitoring of bolted joints using time reversal method and piezoelectric transducers. *Smart Mater Struct* 25(2):025010
12. Wang T, Song G, Wang Z, Li Y (2013) Proof of concept study of monitoring bolt connection status using a piezoelectric based active sensing method. *Smart Mater Struct* 22(8):087001
13. Zhang Y, Zhang X, Chen J, Yang J (2017) Electromechanical Impedance based position identification of bolt loosening using LibSVM. *Intell Autom Soft Comput* 24(1):1–7
14. Yan S, Liu W, Song G, Zhao P, Zhang S (2018) Connection looseness detection of steel grid structure using piezoceramic transducers. *Int J Distrib Sens Netw* 14:1550147718759234
15. Budoya DE, Baptista FG (2018) A comparative study of impedance measurement techniques for structural health monitoring applications. *IEEE Trans Instrum Measur* 67(4):912–924
16. Devices A (1 April) 1 MSPS, 12 Bit impedance converter network analyzer. Available: <https://www.analog.com/en/index.html>

IoT-Based Smart Wristband



Spardha Kapil, Nikita Pandey, and Pardeep Garg

Abstract With the current increase in population, it has become difficult to manage crowds and ensuring their safety particularly at heritage sites and historical places where visitors are not allowed to carry their mobile phones for various reasons. Notifying the crowd timely is a high priority when incidents like fire/electrical accidents, building collapse, stampedes, etc., take place at such large crowd gatherings (religious, exhibition, fairs, musical/sports event, fests, etc.). Handling of such incidents and emergencies proposes a challenge to any organizer of such gatherings that need quick response and assistance of staff. This paper presents an IoT-based smart wristband which has been designed to strengthen the management of such events. This wristband will be given to each visitor to keep a track of their location. Each wristband continuously communicates with the security staff of the respective event through a message sent via Wi-Fi which will contain the location coordinates of the person wearing the wristband that can be used to track him/her through Google Maps when they require help, relocate people to safer places, notify the crowd of any hazardous incidents, and provide fastest routes to emergency exits.

Keywords Wristband · GPS · Wi-Fi · Crowd management · Texas Instruments

1 Introduction

India is a country full of heritage, religious places like temples, gurudwaras, mosques, etc. Heritage sites or places of worship in India are often crowded not only with Indian travelers but also with international travelers. Lakhs of people gather at events like Kumbh Mela and Hajj, which can lead stampede to take place or any other emergency scenario [1]. Be it children or senior citizens people of every age group visit these places. Losing your friends or family members in such crowded places is not very uncommon and often happens. This situation is not restricted to kids or senior people, it can occur with any age group. In such places, mobile phones or any other modes

S. Kapil · N. Pandey · P. Garg (✉)

Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Wakanaghat, Solan, Himachal Pradesh, India
e-mail: garg.pardeep22@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 192,
https://doi.org/10.1007/978-981-33-6546-9_18

171

of communication are not allowed and hence contacting our loved ones becomes difficult. In case of any natural calamity, it gets very late till the time information reaches everyone and uncontrollable damage has already occurred. Crowd safety at these events has always been a problem in India and has caused a lot of deaths in the past due to inefficient crowd management and untimely response to such incidents. To overcome this problem, a lot of advancements are made in the field of technology so that we can live our lives easily without worrying about the nitty-gritty of existence. Internet of things (IoT) has played a huge role in solving the problems that people face in day-to-day challenges of life. IoT has found its application in many different fields, such as smart home, smart city, agriculture [2], health care [3], wearables [4], and many others. These issues motivated us to design an IoT-based smart wristband that can help society in such a situation by providing a solution to these problems. This wristband not only sends the location coordinates but is also attached to the Google Map link so that the security officer can reach the person. Our wristband sends the alert message through the Wi-Fi module. We have also equipped our wristband with an emergency light so that if any disaster happens, the light will glow indicating the emergency exit. The remainder of the paper is organized as follows. Related work has been discussed in Sect. 2. The system components using which the wristband has been designed are presented in Sect. 3. The detailed description of the IoT-based smart wristband is given in Sect. 4. The results obtained are shown in Sect. 5. Finally, the paper is concluded in Sect. 6.

2 Related Work

To detect the location of an object, many researchers have implemented and used various technologies. M. Mohandes [5] developed a tracking and identification system for the pilgrims visiting the holy areas in Makkah-Saudi Arabia during Hajj (Pilgrimage). The author proposed an idea of using radio frequency identification technology (RFID). Each person will be given a mobile sensor having a global positioning system (GPS) unit and the coordinates would be sent periodically in coordination with the RFID system. In case of loss of Internet connectivity, the cell phone stores the location data and sends the data once Internet connectivity is restored [5].

Mohandes et al. designed a system that uses wireless sensor networks (WSN) [6]. In this approach, mobile devices are used which have to be carried by the pilgrims equipped with global positioning system (GPS) chip, antennas, and a microcontroller. The network is set up in the holy place for obtaining and redirecting the information. Now and then the mobile device sends its user's identification and location coordinates. A central server is used to map the coordinates on a geographic information system (GIS).

Dhaou proposed an idea of a client-server network architecture [7] for safe pilgrimage in Saudi Arabia. The system designed allows for regular tracking of the bus during its journey of the Holy area of Makkah using GPS. For identification

purpose, the system uses RFID technology, and the RFID technology is java implemented and allows both the bus driver and the travel agencies to prevent unsanctioned travelers to board the bus.

Sowmya and Santosh Kumar [8] designed a system that depends on a smart application installed on the Android cell phone. As the growth of smartphones is increasing almost every pilgrim would have a smartphone that has GPS and the application installed on the smartphone. It uses the GSM module for getting the pilgrim's information and also further forwarding the information about the rescue location to the missing traveler. Whenever the traveler is in some emergency, he can press the button on the application and the GPS & GSM modules work together to rescue the pilgrim [8].

Khilare et al. [9] developed real-time health control and pilgrim tracking by using wireless sensor networks. They have created a sensor network for tracking of the pilgrims. This sensor network comprised of both fixed as well as mobile units. The fixed units comprise of both software and hardware which are used to communicate with the mobile units. These mobile units are carried by the pilgrims to contact in case of any emergency.

Nafea et al. [10] proposed an approach to reduce the cost and time to deliver medical help by keeping up an online electronic health record. The approach is to make a cloud computing device that will be used to connect to the pilgrims. Each pilgrim will have their mobile phones and a unique magnetic card along with them. This magnetic card would have the pilgrim's health records. The pilgrims can use their mobile phones to contact the health center in case of emergency and with the help of the magnetic cards previous history of the patient can be known.

Shah et al. [11] developed a system that is used to identify, track, and monitor pilgrims. In this system design, a camera is used to regularly monitor and search for the high-density areas using image processing. The communication with the main server is done with the help of GSM. As soon as the prestamping starts, the camera detects the picture and sends it to the respective security personals. This system comprises of two parts: the stampede detection and pilgrim monitoring.

Rajwade and Gawali [12] developed a wearable unit based on Arduino which is used for location sensing and monitoring health parameters. The wearable unit uses GPS for location transmission and heartbeat sensors, the temperature sensor for health monitoring. The pilgrim can call for help in case of emergencies. Adverse health is notified by an alert message on the GUI.

Nair et al. [13] developed a system that has to transmit and receiving sections. The pilgrims have to carry an RFID tag. The transmitting section comprised of the RFID reader, microcontroller, and ZigBee. The transmitter unit sends the location coordinates along with the id of the pilgrim using the ZigBee transceiver. The receiver unit contains the ZigBee transceiver, external EEPROM unit, and microcontroller unit. The received location coordinates and ID will be stored into the external EEPROM regularly. A heartbeat sensing unit is connected to the transmitting section to look over the medical conditions in case of emergency.

3 System Components

The hardware components and software used in our designed system are as following:

3.1 Hardware Components

1. Microcontroller—CC3200
2. Wi-Fi in-built in CC3200
3. GPS module
4. In-built LED
5. In-built push button.

3.2 Software Used

1. Embedded C
2. Energia
3. Temboo
4. Twilio.

The detailed description of these components is as follows:

TI-CC3200 Microcontroller: It is a single-chip microcontroller that comes within an in-built Wi-Fi module and is manufactured by Texas Instrument (TI). The SimpleLink CC3200 [14] is a wireless microcontroller that combines a high-performance 32-bit ARM Cortex-M4 microcontroller, which allows anyone to develop a complete application with a single integrated circuit (IC). It can also be used for signal acquisition and processing.

Ublox NEO-6 M Global Positioning System (GPS) Module: The UbloxNEO-6 M GPS module is a good-performance GPS receiver. It has an in-built $25 \times 25 \times 4$ mm ceramic active antenna therefore providing a strong and stable satellite search capability. To monitor the status of the GPS module, it comes with power and signal indicators. Also if by chance, the main power gets shut down accidentally, then also the module can retain all of its data because of the presence of data backup battery.

Energia IDE: Energia is a network-driven integrated development environment (IDE) and an open-source software system. It provides an instinctive coding environment as well as a strong structure of simple-to-use functional application program interfaces (APIs) and libraries for programming a microcontroller, implemented on the wiring framework. Numerous TI processors are supported by Energia, mainly those which are accessible in the LaunchPad development ecosystem.

Temboo: Temboo is an expandable, fault-tolerant environment for managing and running the smart code snippets that are called Choreo. Choreos can call application program interfaces (APIs), dealing with the Open Authorization (OAuth) procedure, sending email messages, performing encoding, updating of databases, and many more.

Twilio: Twilio is a cloud communication platform. This platform is used for many applications like, for example, it can be used by programmers and software designers to create numerous software-based applications. Some of these applications are like for making and accepting calls, sending and getting text messages, and some other related to communication by utilizing its Web service application program interfaces (APIs).

Hypertext Markup Language 5 (HTML5): HTML5 is the standard markup language that is utilized for showing any document or file in an Internet browser. Numerous advancements like Cascading Style Sheets (CSS) can be utilized in this language. Other incredible assets such as JavaScript, which is a scripting language, can likewise be utilized [15].

Tiny GPS Library: Tiny GPS is a library utilized for analyzing the National Marine Electronics Association (NMEA) information streams. The GPS module is utilized to provide these data streams. This library is utilized for extracting the location coordinates from the available GPS device using compact methods. It gives output in the form of latitude, longitude, altitude, date, time, and speed.

Bootstrap: Bootstrap is a free source and open-source CSS-based system which is responsible for making responsive, easy, and interactive front-end Web page. The technologies used in this are CSS and alternatively, JavaScript implemented plans.

JavaScript: JavaScript is a very useful and powerful coding language. It is used in the development of Web pages mainly to build the front-end side. It allows the client-side interactively to cooperate with the main Web site. It is a coding language with object-oriented roles.

4 System Design and Implementation

We have used TI-CC3200 as it is more efficient than other microcontrollers like Arduino as it consumes very less power and also comes with a ton of inbuilt features like Wi-Fi connectivity, a couple of on-board sensors, and also LEDs. Also, to increase the efficiency we have used the inbuilt Wi-Fi module instead of the GSM module as in the case of GSM multiple users share the same bandwidth, due to which interferences occur. Another disadvantage of using the GSM module is that it can interfere with electronic devices such as hearing aids and pacemakers. The block diagram of IoT-based smart wristband using CC3200 module is shown in Fig. 1.

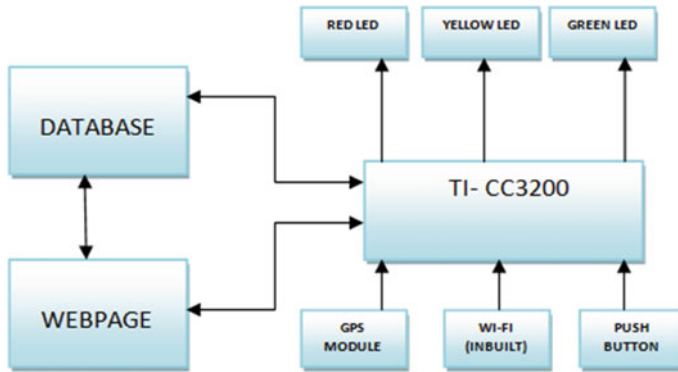


Fig. 1 Block diagram of the proposed model

The hardware of the model comprised of a microcontroller (TI-CC3200) [8], Wi-Fi module (inbuilt in the microcontroller), GPS module (NEO-6 M GPS module), three LEDs (red, yellow, green), and one push button.

A Web page has also been created named “Crowd Safety and Security System” which will be linked to the database (as shown in Fig. 1) that will contain the information of all the visitors including their emergency contact number for easy and quick retrieval of information. The Web page will further be linked to the wristband through Wi-Fi module. This Web page will have four buttons EMERGENCY EXIT: HIGH and LOW and REQUEST RECEIVED: HIGH and LOW.

During the event if any visitor requires help in case of any emergency or if the person is lost, he/she can press the push button on the wristband. A push button is used in case of a switch button to prevent accidental sending of help alerts. Pressing the push button continuously for 5 s will prompt the GPS module to render the GPS coordinates. Then the microcontroller will extract the coordinates (as depicted in Fig. 1) and send a message back to our Web page saying that “ PLEASE HELP ME” along with the Google Map link to his location using the GPS module. Then the nearest security officer would be alerted and he will go and find the person. The location would be updated every 5 s. The alert message is sent through the in-built Wi-Fi module in the wristband with different Wi-Fi hotspots set over different areas of the building. We have not used the GSM module as sometimes the signal transmission is not proper due to interference and causes errors.

The yellow LED on the wristband will glow indicating that the location has been sent successfully. If the request is received by the security official, the person will be notified through the green LED on the wristband.

The security official, after receiving the request, will press the REQUEST RECEIVED: HIGH button on the Web page. This will result in glowing of the green LED on the wristband notifying the person that the message has been received.

Also, during any incident such as fire or building collapse, the security officer will press EMERGENCY EXIT: HIGH button on the Web page by which the red LED will glow on the wristband notifying the person to evacuate the premises.

5 Results

The complete system setup of the model includes a computer that is interfaced with the TI-CC3200 microcontroller which in turn is interfaced with the GPS module as shown in Fig. 2. The Wi-Fi module, three LEDs, and a push-button are in-built in the microcontroller.

TinyGPS Library is used to parse the location coordinates which are provided by the GPS module through the microcontroller. The extracted location coordinates which include latitude, longitude, altitude, and time of the extraction are shown in Fig. 3.

The extracted location coordinates along with an alert message and Google Map link are sent to the security official as shown in Fig. 4.

The Google Map link when opened is depicted in Fig. 5.

The Web page is made using tools like HTML5, Bootstrap, and JavaScript and is named “Crowd Safety and Security System” which will be linked to the wristband through Wi-Fi module. This Web page will have four buttons EMERGENCY EXIT: HIGH and LOW and REQUEST RECEIVED: HIGH and LOW as shown in Fig. 6.

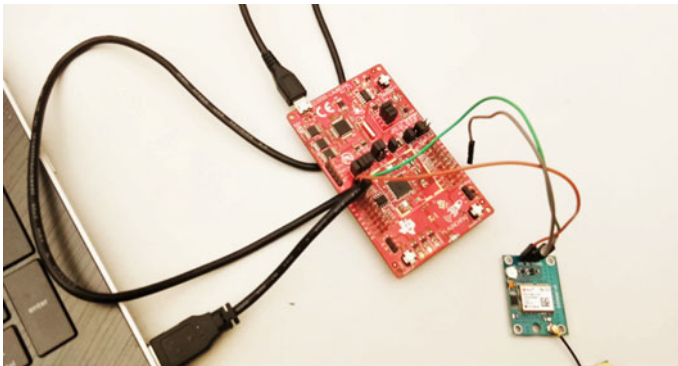
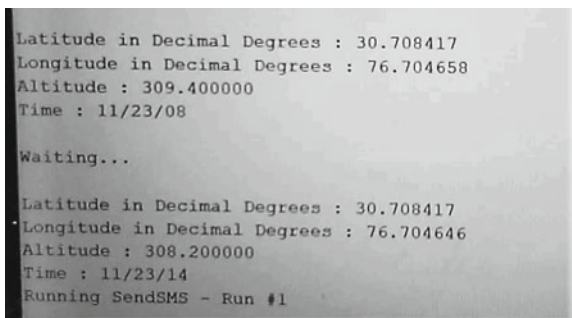


Fig. 2 Experimental setup of the system

Fig. 3 Acquiring GPS coordinates



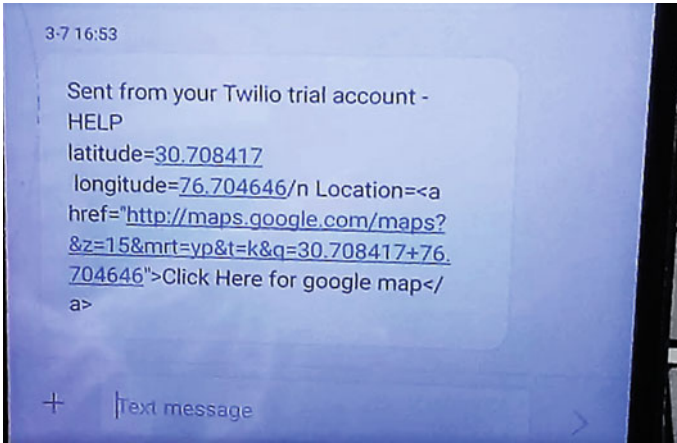
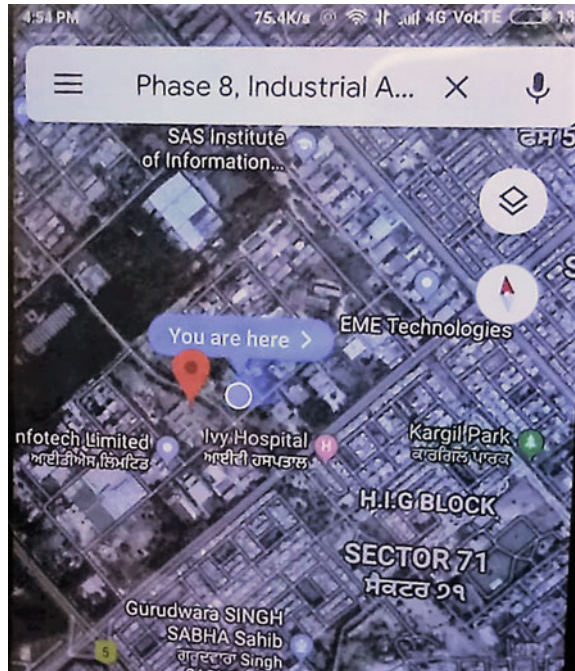


Fig. 4 Alert message

Fig. 5 Google Map location marker



6 Conclusion

A prototype of the wristband has been designed for the safety of the people visiting either the heritage site or the pilgrimage site. The wristband is built on TI-CC3200

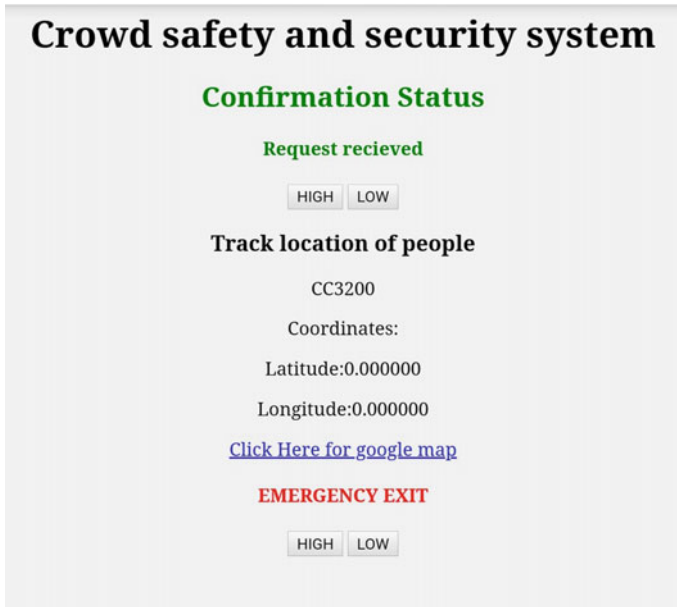


Fig. 6 Web page

for efficiency which may not be provided by the Arduino board. Often there is a difficulty at any heritage or pilgrimage site that languages do not match and people cannot be contacted. This is easily solved by our wristband as the only thing a person needs to do is to press the button for 5 s. This wristband is very easy to operate and can be even used by children or senior citizens. This can be operated by anyone.

In future, this work can be extended by adding an LCD screen to the model so that when the emergency button is pressed the LCD screen can show the emergency path.

References

1. Binhotan EA, Altameem A (2019) IoT-based RFID framework for tracking, locating and monitoring the health's rituals of pilgrims during Hajj. In: Luhach AK, Kosa JA, Poonia RC, Gao XZ, Singh D (eds) 1st international conference on sustainable technologies for computational intelligence (ICTSCI). Conference 2019, AISC, vol 1045. Springer, pp 37–46
2. Elijah O, Rahman A, Orikumhi I, Leow CY, Hindia MN (2018) An Overview of Internet of Things (IoT) and data analytics in agriculture: benefits and challenges. *IEEE IoT-J* 5(5):3758–3773
3. Baker SB, Xiang W, Atkinson I (2017) Internet of things for smart healthcare: technologies, challenges, and opportunities. *IEEE Access* 5:26521–26544
4. Islam SMR, Kwak D, Kabir MDH, Hossain M, Kwak KS (2015) The Internet of things for health care: a comprehensive survey. *IEEE Access* 3:678–708

5. Mohandes M (2010) RFID based tracking system for pilgrims identification and tracking. *ACES* 25:273–282
6. Mohandes M, Haleem M, Deriche M, Balkrishnan K (2012) Wireless sensor networks for pilgrims tracking. *IEEE Embedd Sys Lett* 4:106–109
7. Dhaou IB (2010) Client-server network architecture for safe pilgrim journey in the Kingdom of Saudi Arabia. *IEEE Int Veh Sympos* 1043–1048
8. Sowmya M, Kumar BS (2014) Smart way of tracking and assistance of pilgrims using android. *IJETT* 16:410–413
9. Khilare PA, Limkar MB, Kamble P (2014) Wireless sensor network for real time health control and tracking of pilgrims. *IOSR-JEEE* 8–12
10. Madra I, Bhairy N, Zeidan ZA (2014) Health tracking framework for Hajj pilgrims using electronics health records for Hajj. In: 2014 IEEE international conference on bioinformatics and biomedicine (BIBM), 2–5 Nov 2014, UK, pp 605–607
11. Shah SK, Kulkarni S (2015) A review: monitoring and safety of pilgrims using stampede detection and pilgrim tracking. *IRJET* 4
12. Rajwade KC, Gawali DH (2016) Wearable sensors based pilgrim tracking and health monitoring system. *ICCUBEA*
13. Anju NM, Joshua DS (2014) Design of wireless sensor networks for pilgrims tracking and monitoring. *IJISER* 1
14. Texas Instruments CC3200 SimpleLink™ Wi-Fi® and Internet-of-Things solution, a single-chip wireless MCU, SWAS032F datasheet, July 2013. Revised Feb 2015
15. Evans MB (2000) Challenges in developing research-based web design guidelines. *IEEE Trans Prof Comm* 43(3):302–312

Expert System on Smart Irrigation Using Internet of Things



Kratika Varshney, Sweta Tripathi, and Vaibhav Purwar

Abstract An upsurge in the population of India has resulted in a scarcity of resources. This problem escalates when unexpected rainfall occurs, especially during the time of harvesting that can result in crop loss and harms the nation's food security. This paper proposes Internet of things (IoT) involvement in rural development and introduces the model "Expert system on smart irrigation (ES2I)". ES2I is designed using Arduino and LoRa (long-range) technology. By managing the irrigation system through this model, water requirement of the field can be regulated. It will provide a mechanism to save our crops from unexpected rainfall and lays out a strategy to store the water (rain) for further use.

Keywords Arduino · Cloud computing · Internet of things (IoT) · LoRaWAN · Smart irrigation

1 Introduction

Farming plays an essential role in India's economy. Approximately, 54.59% of the population is involved in agriculture and allied activities as per the 2011 census and contributing 17.39% of India's gross value added (GVA) for the year 2016–17 [1]. But, unpredictable weather conditions, unregulated irrigation, mishandling of fertilizers, and pest attack lead to the failure of crops and farmers' debt situation.

In this scenario, Internet of things (IoT) is considered to be a promising technology that will automate agriculture by connecting every required entity through

K. Varshney (✉) · S. Tripathi
Department of Electronics and Communication Engineering, KIT, Kanpur, U.P, India
e-mail: kratikvarshney94@gmail.com

S. Tripathi
e-mail: swetatripathi16@gmail.com

V. Purwar
Department of Electronics and Communication Engineering, PSIT College of Engineering,
Kanpur, U.P, India
e-mail: vaibhavpurwar101@gmail.com

the Internet. It enables us to extract real-time data and will help us to utilize every inch of an arable field.

The concept of IoT was brought up by Kevin Ashton in 1999 [2, 3]. Fundamentally, it is an incorporation of multiple devices that sense, interact, and communicate with external and internal states through the integrated technology that the object contains. It can bring everything in connectivity. In IoT, sensors are used to collect the information, and the data so collected is processed using controllers. Then, the actuators are used to complete the automation process, i.e., take the electrical input and turn it into the physical action.

IoT in agriculture can be beneficial in the following ways [3]:

- By controlling the internal processes, production risk can be reduced.
- IoT can also reduce the waste level, and cost can be effectively managed. Hence, the farmer's condition can be improved.
- It can also mitigate the risk of diminished yield or even complete crop failure.
- It can reduce labor costs as automation will reduce the number of labor needed in the field.
- Regulated irrigation practices with the help of IoT will also reduce the electricity requirement in the field.
- Smart irrigation practices will also save the wastage of water in the field.
- IoT will reduce the stress level of crops and trees and will improve the yield.
- IoT in agriculture will lead to good quality and quantity of yield and that will propel the farmers to invest more in their field and purchase high-quality seeds.

Albeit, the use of the Internet of things in the agriculture sector are numerous, but the direct benefits are still not been drawn by the farmers. So, there is a need to focus on its accessibility, interoperability, and research.

Researchers [4, 5] have proposed smart agriculture sticks that can provide farmers with real-time data and can indeed solve many issues. Likewise, O.Pandithurai et al. [6] have developed the AGRO-TECH system using mobile technology to deal with agronomical issues such as improper power and water management, and poor crop growth.

Irrigation plays an important role in agriculture. Madushanki et al. [7] have stressed on water management. The author also recommends the use of LoRaWAN because of limited Internet connectivity and speed in developing countries. LoRa is low-power and long-range wide-area network technology and is widely used in smart agriculture. LoRa operates in license-free frequency bands. So, one does not need a license from the government to transmit data on these frequencies. Free or unlicensed frequencies vary from country to country. It has an advantage of low power requirements, excellent battery life, long-range connectivity, global availability, interoperability, and offers GPS-free tracking applications. But, the open frequency can result in disturbances, and the data rate may get low. It is used in smart agriculture, smart logistics, supply chain, smart cities, smart metering, smart homes, and etc. [8]. Rajeshkumar et al. [9] has proposed an automated irrigation system using Raspberry pi. Likewise, Gulati et al. [10] have also proposed a smart irrigation system that uses Arduino Uno, ESP-8266 Wi-Fi module chip, and sensor technology. The author

also suggested that evapotranspiration (ET) plays an important role in irrigation. ET monitoring can be done using different sensors, but remote sensing is the best option [11]. Also, research is going on to measure ET with great accuracy.

Irrigation can be made smarter if we get a precise knowledge of water-stressed areas that can be achieved by using unmanned aerial vehicles (UAVs)/drones equipped with different sensors and cameras [12].

Rain is a blessing for farmers as it naturally waters the crop. But, at the same time, unexpected rainfall can result in food crises, land problems, and even spoil the whole crop field. Researchers have used IoT to solve this problem. Roselin et al. [13] have proposed an automated system by using Arduino, Global System for Mobile Communication (GSM), Wi-Fi, and sensor technology. It will not only protect the crop from spoilage due to rain but also efficaciously recycle the rainwater for further irrigation purposes. The author has also used a passive infrared sensor (PIR) to sense animal intrusion in the field. Likewise, Giordano et al. [14] have developed a model to protect the crop from wild animals and extreme weather conditions. His proposed device would run on solar power and using Wi-Fi technology. This device also aware the farmers about lousy weather conditions in their farm through email. Furthermore, researchers [20–22] have also suggested the use of plastic sheet to cover the field automatically by using IoT that will provide protection to crops from heavy rainfall. But, they have used GSM and Wi-Fi in their respective models. Wi-Fi and GSM are not an idle choice in the agricultural field. Both of them require a good amount of power to operate and, therefore, cannot be used for battery-operated devices. Moreover, the agricultural field also has Internet connectivity issues.

In light of this issue, the paper proposes the model “Expert system on smart irrigation (ES2I).” The model will provide a mechanism to save crops from unexpected rainfall and provide a technique to use rainwater for further needs. The model will run on solar power and uses Arduino Uno and LoRa technology. Arduino Uno is connected with different sensors that are planted in the field. Data from Arduino Uno is passed through a comparator and is used to know the water requirement of the field. Then, appropriate action will be taken accordingly.

2 Model Descriptions

The challenges faced by the agriculture industry are stipulated by the inclusion of IoT technology in farms. The proposed model, “Expert System on Smart Irrigation” (ES2I), protects the crop from unexpected rainfall, as shown in Figs. 1 and 2. It is solar power enabled platform that works on 5 V DC supply. ES2I model is composed of the solar panel to recharge lithium batteries and provides regulated supply via 5 V DC converter to Arduino Uno and LoRa module. Several IoT sensors are installed on the farm, such as soil moisture sensor (KG003), rain sensor, temperature and humidity sensor (DHT11), pressure sensor, PH sensor (to check the soil salinity), and disdrometer (an audio sensor to measure the rain intensity). Passive infrared sensor (PIR sensor) is also used to check animal intrusion on the farm. The data from

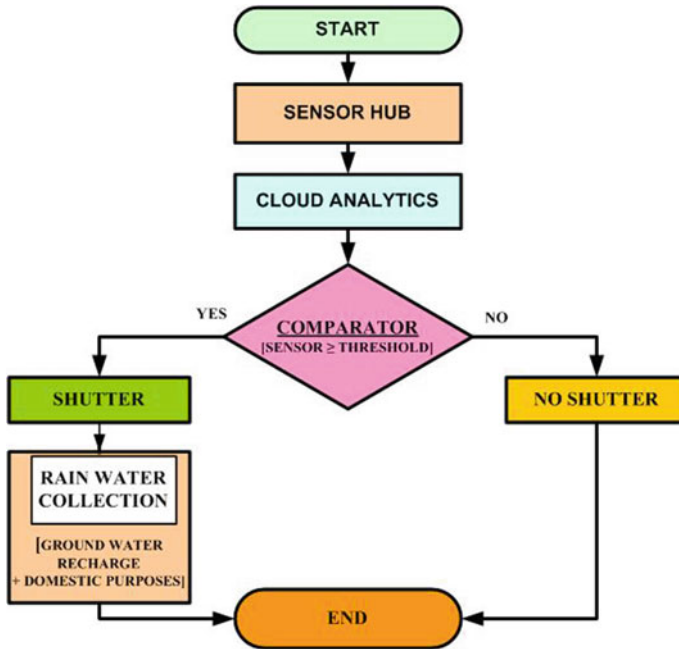


Fig. 1 Proposed model expert system on smart irrigation (ES2I)

each of them is pipelined to the sensor hub (Arduino Uno). This intelligent framework incorporates the LoRaWAN network for data transmission from the sensor.

Procured data from the sensor hub goes to the cloud analytics module where data processing and data analysis occur. The data pool includes the data of humidity, moisture level, temperature level and rain intensity. The weather forecasting data is fetched from the cloud and is compared with the data from the sensor hub using comparator via cloud analytics.

If the probability of rainfall is high and the soil moisture content is sufficient to meet crop needs, then farmers will be alerted through an alarm, and the shutter would be closed automatically. Therein water is collected to recharge the underground water. On the other hand, the shutter would remain open in the two conditions, i.e., when the rain probability is null or when there is less moisture in the soil. In this condition, irrigation can be done either through rain or by drip irrigation (if the probability of rainfall is low).

3 Results and Discussion

Moisture content in the soil directly impacts the growth rate and yield of the crops. Moisture levels in the soil can be tested through moisture sensors planted in the field.

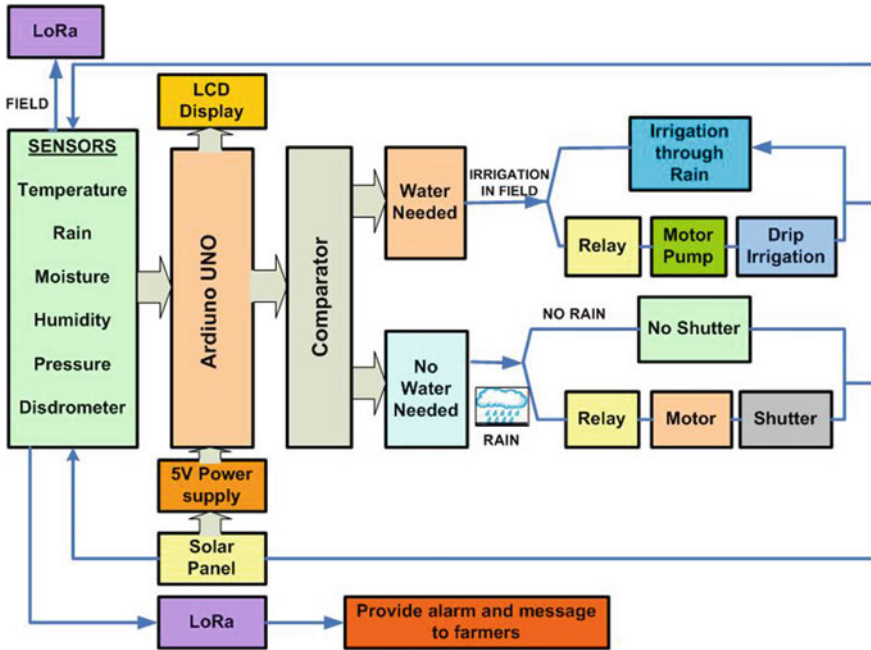


Fig. 2 Block diagram of smart irrigation using LoRa and Arduino Uno

Moisture sensor reading will be opposite in nature, i.e., its value will be higher when the soil has low water content or dry, and its value will be lower when the moisture content in the soil is high. The varying moisture level of the soil is depicted in the following Figs. 3 and 4, which have been drawn using ThingSpeak software. Graphs have been drawn by installing the moisture sensor (FC-28) in the soil where orange was growing.

Figure 3 has two shoots that are clearly visible at 23:57:49 and 00:59:30. At this point, the soil has low moisture content, i.e., below 65%, and it needs water. In the

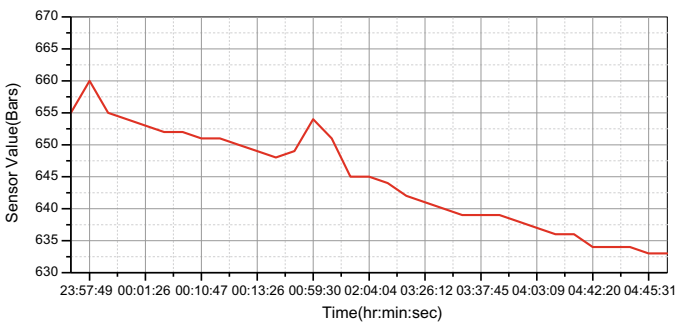


Fig. 3 Raw data of moisture sensor showing water content in the soil

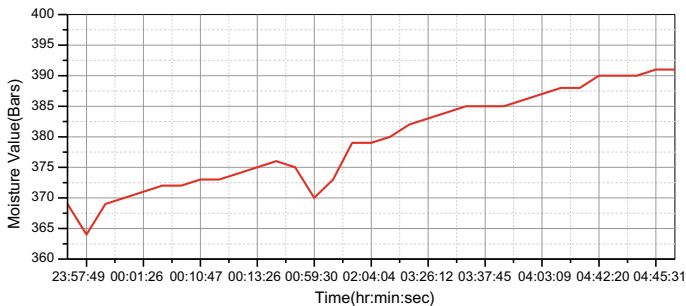


Fig. 4 Moisture count in the soil can be calculated from sensor data using formula Moisture level = $(1024 - \text{sensor value})$

initial phase, drip irrigation is used to remove the dryness of the soil, which has added approximately 2% moisture content in the soil. Then again, there is a dip in moisture content at 00:59:30. Rainwater has been used between the time interval 00:59:30 to 2:20:17, which has remarkably increased the soil moisture content by approximately 6%. The steepness in Fig. 3 depicts that the moisture content in the soil is rising.

The moisture need of the field depends upon the type of soil and crop. So, to set the threshold value of the moisture sensor, these parameters need to be measured at the field site concerned.

As per the data [15–19], untimely rainfall can result in colossal crop loss and can hamper the food security of the nation. A traditional manual method used to protect the crops possesses human error and leads to crop loss even if a minute delay exists.

Frameworks based on Wi-Fi [20] and GSM [21, 22] are less efficient than compared to the LoRa technique adopted in the proposed model. LoRa comes with long-range and less power consumption and can be used in battery-operated devices, unlike GSM and Wi-Fi technologies. Moreover, it provides excellent connectivity in agriculture fields and can establish reliable networking in rural areas where Internet connection is fragile. E2SI model will monitor the soil water content and takes preventive action against adverse weather conditions (rain/hailstorm/drought). It justifies the need for the automated methods available in the literature.

Precision agriculture is the need of an hour to meet the demand and supply. The automation trait of the expert model is capable of reducing human error, thereby attaining of higher productivity ratio as well as reducing the labor cost.

4 Conclusion and Future Scope

The proposed model will help farmers in the following ways:

- Providing crop protection from unexpected rainfall,
- Rainwater harvesting,

- Protection of crop from animal intrusion.

In designing the expert irrigation system, Arduino and LoRaWAN technology have played a crucial role.

Some of the future trends that are analyzed from the paper are as follows:

- The proposed IoT-based expert irrigation system involves a cost issue. In the future, the focus can be made on how to reduce the cost.
- To make the system more efficient, Arduino can be replaced by Raspberry pi, or both can be used in combination.

References

1. Chatterjee R (2019) India 2020. 64rdedn. New media wing, Ministry of Information and Broadcasting, Government of India, Lodhi Road
2. <https://iot-analytics.com/internet-of-things-definition/>. Last Accessed 16 June 2020
3. Farooq MS, Riaz S, Abid A, Abid K, Naeem MA (2019) A survey on the role of IoT in agriculture for the implementation of smart farming. In: Special section on new technology for smart farming 4.0 research challenges and opportunities. IEEE Access, pp 156237–156271
4. Shanmujam R, Saranya K (2020) IoT based smart farming stick-a survey. Int J Adv Res 8(01):2320–2547
5. Nayyar A, Puri V (2016) Smart farming: IoT based smart sensor agriculture stick for live temperature and moisture monitoring using Arduino, cloud computing and solar technology. In: The international conference on communication and computing system (ICCCS), pp 673–680
6. Pandithurai O, Aishwarya S, Aparna B, Kavitha K (2017) Agro-tech: a digital model for monitoring soil and crop using Internet of things (IoT). In: 3rd international conference on science technology engineering and management (ICONSTEM). IEEE, pp 342–346
7. Madhushanki AAR, Halgamuge MN, Wirasagoda WAHS, Syed A (2019) Adoption of IoT in agriculture and smart farming towards urban greening: a review. Int J Adv Comput Sci Appl (IJACSA) 10(4):11–28
8. Zourmand A, Hing ALK, Hung CW, Rehman MA (2019) Internet of things (IoT) using LoRa technology. IEEE international conference on automatic control and intelligent system (I2CACIS). IEEE, Selangor, Malaysia, pp 324–330
9. Rajeshkumar N, Vaishnavi B, Saraniya K, Surabhi C ((2019)) Smart crop field monitoring and automation irrigation system using IoT. Int Res J Eng Technol (IRJET), 7976–7979
10. Gulati A, Thakur S (2018) Smart irrigation using Internet of Things. In: 8th international conference on cloud computing, data science and engineering (Confluence). IEEE Xplore, pp 819–823
11. García L, Parra L, Jimenez JM, Lloret J, Lorenz P (2020) IoT based smart irrigation systems: an overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture. Sensors 20(4):1–48
12. Ayaz M, Uddin MA, Sharif Z, Mansour A, Aggoune EM (2019) Internet of things (IoT) based smart agriculture: towards making the field talk. In: Special section on new technology for smart farming 4.0: research challenges and opportunities. IEEE Access, pp 129551–129582
13. Roselin AR, Jawahar A (2017) Smart agro system using wireless sensor networks. In: International conference on intelligent computing and control systems (ICICCS). IEEE Xplore, pp 400–403
14. Giordano S, Seitanidis IN, Ojo MO, Adami D IoT Solution (2018) For crop protection against wild animal attacks. In: IEEE international conference on environmental engineering (EE). IEEE, pp 1–5

15. <https://weather.com/en-IN/india/news/2019-11-01-post-monsoon-rains-damage-crops-maharashtra-losses-5000-crores>. Last Accessed 16 June 2020
16. <https://www.grainmart.in/news/farmers-worried-as-wheat-damage-reported-due-to-untimely-rains-in-north-india/>. Last Accessed 16 June 2020
17. <https://economictimes.indiatimes.com/news/economy/agriculture/unseasonal-rain-damages-crop-in-north-india/articleshow/63756236.cms>. Last Accessed 16 June 2020
18. <https://www.indiatoday.in/mail-today/story/farmers-face-heavy-losses-due-to-untimely-rainfall-242737-2015-02-14>. Last Accessed 16 June 2020
19. <https://timesofindia.indiatimes.com/city/meerut/bijnor-stares-at-30-wheat-loss-due-unseasonal-rainfall/articleshow/75705258.cms>. Last Accessed 16 June 2020
20. Patil R, Gayathri J, Ashwini K, Gururaj KK (2018) Protection of crops and proper usage of rain water using IOT. *Int J Adv Res Electr Electron Instrum Eng (IJAREEIE)* 7(6):3017–3022
21. Kalyan TS, Madhumitha MS, Vodnala D (2020) Automatic crop protection from heavy rainfall and preserving rain water. *Int J Adv Sci Technol (IJAST)*: 6687–6696
22. Ajay A, Shivashankar PS, Sunil DM (2019) Agriculture crop protection with rain water harvesting and power generation. *Int J Sci Res Rev (IJSRR)* 7(3):2471–2476

Design and Analysis of Nonlinear PID Controller for Complex Surge Tank System



Jitendra Kumar

Abstract An adaptive nonlinear proportional plus integral plus derivative (NPID) controller is explained in the present exploration work for controlling a surge tank system. Dealing with a nonlinear system like surge tank is always a challenging research topic for scientists. Linear PID controller is unable to produce satisfactory result in this regard. NPID controller can be a suitable solution regarding this problem. A bio-inspired meta-heuristic algorithm, Cuckoo search optimization (CSA) technique, is explored in this research for finding the optimal gains of NPID controller. A performance criterion for optimization is chosen as integral of absolute error (IAE). The proposed controller is compared with state of the art for viewing its efficacy. Both the control techniques are tested for servo mechanism, regulatory operation, and parameter variations. Based on these criteria, NPID offers better performance on PID controller.

Keywords PID · Nonlinear PID · Surge tank · CSA

1 Introduction and Literature Survey

History of classical proportional plus integral plus derivative (PID) controller is about eight decades as reported by researchers [1]. Initially, pneumatic proportional and integral controller was introduced by Taylor Instrument Company in 1935. Also, in the same year, Foxboro Instrument Company introduced pneumatic PID controller which makes the process variable at a new set-point [2]. However, problem of gain adjustments of PID controller is yet to be solved. Earlier to that era, simply hand tuning was performed. This random selection of parameter adjustments of the PID controller was tedious and time taken. Taylor engineers, Ziegler, and Nichols have proposed a solution for this problem by developing “Ziegler and Nichols” method of tuning in 1942–43 [3, 4]. They suggested the tuning scheme first in open-loop method and second in closed-loop method. A control engineer, Cohen and Coon [5],

J. Kumar (✉)

Department of Electronics and Communication Engineering, GLA University, Mathura, India
e-mail: Jitendra.kumar@gla.ac.in

an employee of Taylor Instrument Company, during 1950s suggested an alternate method of parameter adjustment of PID controller.

Since last eight decades, classical PID controllers are literally extensively used control method adopted for the industries as it contains unpretentious structural design intension and economy in employment. Literature survey reveals that PID controller is used more than 90% for industrial applications nowadays [6–9]. Further, research study discovered that conventional PID controllers are generally insufficient for the systems containing nonlinearity, time-delays, significant oscillations, parameter variations, multi-inputs multi-outputs, time-varying, and complexity. Different control strategies have been evolved further for enhancement of control techniques. In this regards, nonlinear PID (NPID) controller came in the picture which adapts the parameters of NPID in run time allowing to the deviations of the inaccuracy [10]. The foremost design of NPID is constructed on PID controller where the gains are function of sudden inaccuracy. Numerous alternatives of NPID were investigated on various compound industrial difficulties such as feedback controller design for hydraulic systems [11], for controlling in car as anti-lock brake design [12], controlling the micro-positioning of piezoelectric motors [13], location control of a pneumatic muscle as a final control element [14], and controlling temperature of a CSTR [15]. In entire problems, NPID offers improved performance on PID. These literature reviews inspire the researchers to suggest an innovative NPID assembly for a complex system like surge tank. In the present work, an NPID controller is proposed and tested on the given system. To check efficacy of the system, the performance of the NPID controller is compared with linear PID controller. Gains of both the controllers are modified by Cuckoo search algorithm (CSA) [16, 17].

Further, the article is organized as follow. After literature survey in Sect. 1, surge tank mathematical model is presented in Sect. 2. Controller design is expressed in Sect. 3, while controller tuning and result analysis has been performed in Sect. 4. Finally, research work is concluded in Sect. 5.

2 Dynamic Model of Surge Tank

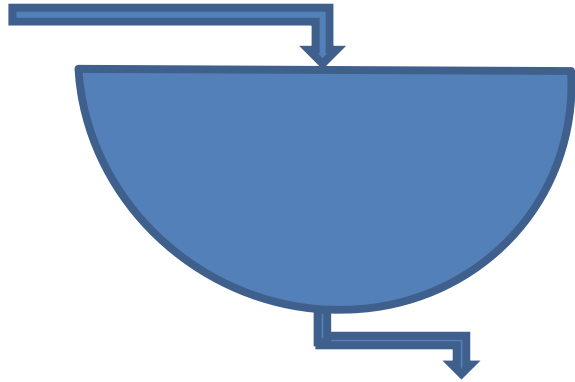
A spherical-shaped surge tank having one outlet is presented in Fig. 1 [18]. Due to spherical shape, its cross-sectional area is varying with time. Main objective in surge tank is to control the water level at a particular height $h(t)$. Input flow rate $u(t)$ is manipulating variable in surge tank level control strategy. The mathematical model representing the system is given as:

$$\frac{dh(t)}{dt} = \frac{-c\sqrt{2gh(t)}}{A[h(t)]} + \frac{u(t)}{A[h(t)]} \quad (1)$$

where

$u(t)$ is the input flow (m^3/s)

Fig. 1 Surge tank



- $h(t)$ is the water level (m)
- $A[h(t)]$ is the cross-sectional area of the tank
- $g = 9.8 \text{ m/s}^2$ is gravitational acceleration
- $c = 1$ is the cross-sectional area of output pipe.

In this model, area $A[h(t)] = ah^2(t) + b$ is assumed where a and b are variables. In simulation of surge tank model, parameters a and b have been taken fixed as $a = 1$ and $b = 2$ in this research work. As it can be observed that due to variations in parameters a and b with time, the system becomes non-stationary.

3 Structure of Controller

This section represents the fundamental design of linear PID and NPID controller. A fundamental feedback control design of surge tank system is shown in Fig. 2. Rest is explained in the subsections following.

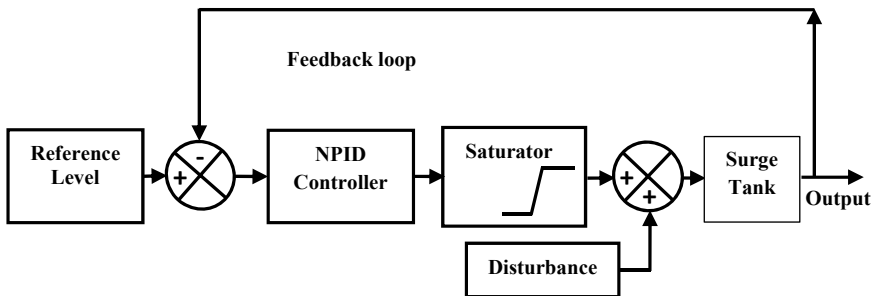


Fig. 2 Fundamental feedback control diagram of surge tank

3.1 PID Controller Structure

In this subsection, the basic strategy of PID controller is presented. In time domain, the structure of simple PID controller [23] can be stated as;

$$U_{\text{PID}}(t) = \left[k_p e(t) + k_i \int e(t) dt + k_d \frac{de(t)}{dt} \right] \quad (2)$$

where the parameters k_p , k_i , and k_d are the gains of proportional, integral, and derivative controllers. In s-domain, the output of PID controller can be expressed as;

$$U_{\text{PID}}(s) = \left[k_p + k_i * \frac{1}{s} + k_d * s \right] E(s) \quad (3)$$

3.2 NPID Controller Design

As surge tank is a nonlinear system, for a rapid modification in reference point, a small integral gain k_i offers a slow transient change toward set-point. Therefore, it has been observed that a technique which can offer a corresponding change in integral gain as well as proportional gain with rapid alteration in set-point can minimize the transient inaccuracy. In this context, a NPID controller is proposed in this work. It allows to change the gains of proportional and integral controller, i.e., k_p and k_i , respectively, in run time. It makes the NPID controller adaptive on the basis of current error occurred in the system. The mathematical expression of NPID controller can be presented as;

$$U_{\text{NPID}}(t) = \left[k_p e(t) f(e) + k_i \int f(e) e(t) dt + k_d \frac{de(t)}{dt} \right] \quad (4)$$

Here, $f(e)$ is a nonlinear proportional and integral gain factor which is a function of error e , and it can be formulated as;

$$f(e) = \cosh(k_0 e) \quad (5)$$

or,

$$f(e) = \frac{\exp(k_0 e) + \exp(-k_0 e)}{2} \quad (6)$$

Here,

$$e = \begin{cases} e; & |e| \leq e_{\max} \\ e_{\max} * \text{sgn}(e); & |e| > e_{\max} \end{cases} \quad (7)$$

In this research work, k_0 and e_{\max} are positive constants considered by user. The lower bound of $f(e)$ would be 1, when $e = 0$. Here, k_0 is optimized by CSA algorithm, whereas $e_{\max} = 20$ is assumed.

4 Controller Tuning and Result Investigation

As per its name, the fundamental design of CSA is based on the laying behavior of bird cuckoo. It is a meta-heuristic exploration procedure which is bio-motivated by the bird cuckoo and projected by Yang and Deb [16, 17]. Main features of CSA were mimicked by the behavior of bird cuckoo. CSA has the passionate aptitude to resolve multi-dimensional problems competently as well as it offers robust searching capability paralleled to alternative standard bio-inspired algorithm. A flow diagram of CSA is revealed in Fig. 3. Basic essential guidelines for CSA suggested by Yang and Deb are as follows:

- (a) Initially, each cuckoo puts one egg, i.e., one answer at a period and puts in a random nest.
- (b) The uppermost class of eggs drives to subsequent generation, and other nests are rejected.
- (c) The presented host nests are stable and host can discover the strange with a probability $p_a \in [0, 1]$. If the host can examine and observe alien egg, then it can either abolish or remove the nest and design a completely new nest as a new residence.

The greatest significant and attribute of CSA is the notion of Levy flight, which is a compound random walk tactic that explores it protuberant with the other optimization algorithms. The altered constraints of CSA which prominently used for optimizing different problems are as follows (Table 1):

Simulation works have been carried by MATLAB/SIMULINK software. A computer system is used with Intel core™ i5 processor with frequency 3.33 GHz, 8 GB RAM, and a 32-bit operating system which is for performing all the simulation studies. To find the solution of differential equations, Runge–Kutta method with order four is instructed, whereas sampling time for simulation study is reserved as 1 ms. Further, for entire simulation study, tuned gains are unaltered. The controller output boundaries are lies within -30 to $30 \text{ m}^3/\text{s}$. In the present study, considered reference level is shown in (8) as follows,

$$r(t) = \{2.5 + 0.25 \sin 0.2 \pi t\} \quad (8)$$

Integral of absolute error (IAE) has been referred as the objective function (OBF) in the present research work which is established in (9).

Fig. 3 Flowchart of CSA

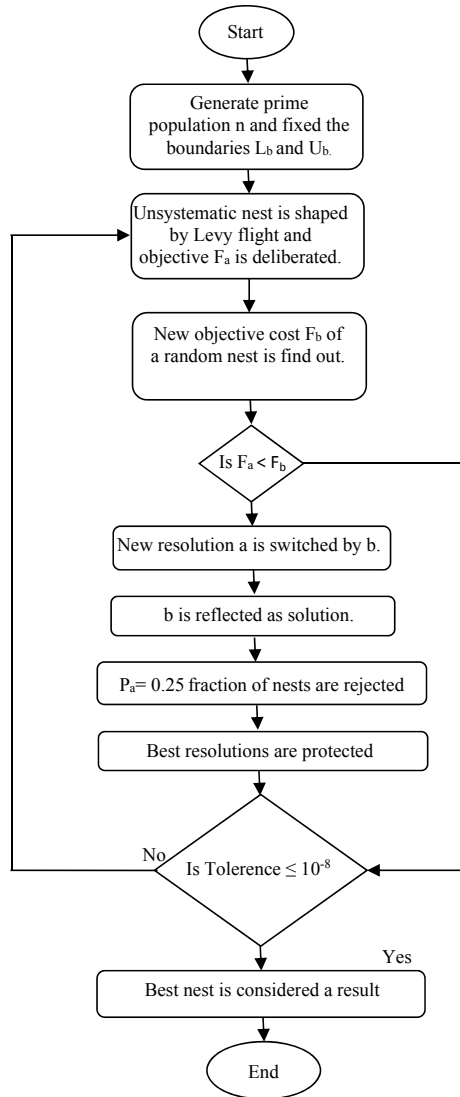


Table 1 Used parameters setting for CSA

Design parameter	Value
Nest available	25
Tolerance	1.0E-9
Step size	1.5
Likelihood of strange eggs	0.25
Total no. of iterations	100

$$IAE = \int_0^t |e(t)| dt \tag{9}$$

Simulation exploration has been carried to evaluate the performances of PID and NPID controllers for refcece level tracking. Firstly, to find the optimized gains, CSA is applied and objective function versus iteration curves for both the controllers which are shown in Fig. 4. Obtained gains are listed in Table 2. The curves which represent reference level tracking are shown in Fig. 5, whereas error curves and controller output curves are depicted in Figures 6 and 7, respectively. Obtained IAE values for PID and NPID controllers are 0.0059 and 0.0049, respectively. Founded the objective function values, the performance of NPID is 16.94% improved than PID control scheme.

To check the adaptability of NPID assembly, disturbance rejection task is also performed in the existing work. A disturbance is introduced to the output of controller which is given in (10) as;

$$dist(t) = u(t - 5) - u(t - 6) \tag{10}$$

A curve for disturbance signal is depicted in Fig. 8. Corresponding trajectory tracking curves, error curves, and controller output for PID and NPID controller are shown in Figures 9, 10, and 11, respectively. IAE values obtained for PID and NPID controllers are 0.018 and 0.012, respectively. After observing the obtained

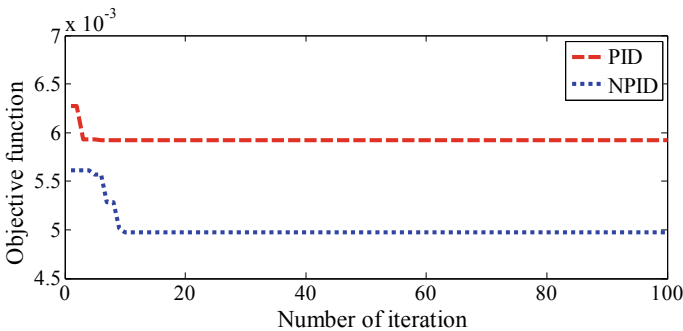


Fig. 4 OBF versus iteration plot

Table 2 Optimized controller parameters

Parameters	PID	NPID
k_p	3002.53	3001.83
k_i	1204.17	1203.27
k_d	0.10	0.10
k_0	–	501.68

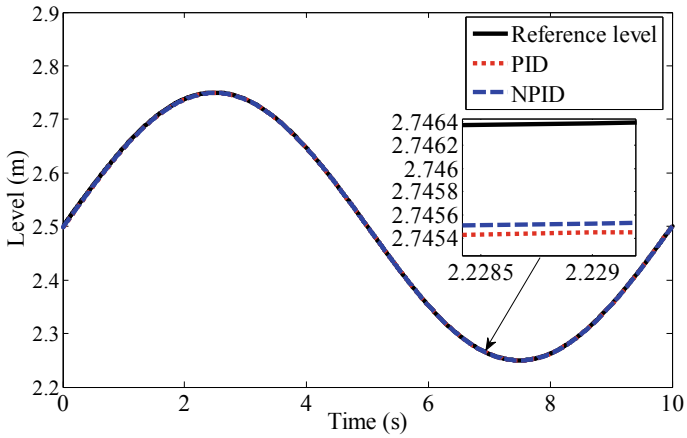


Fig. 5 Reference level tracking curve

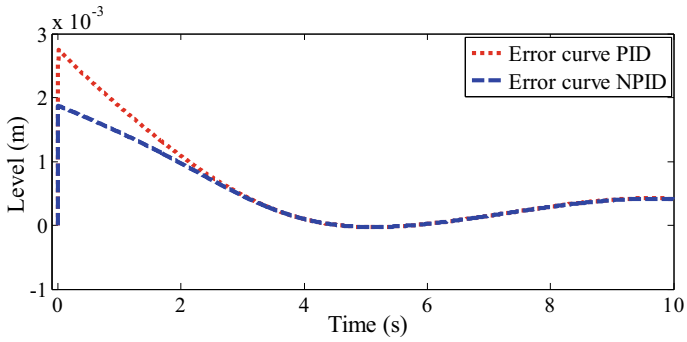


Fig. 6 Error curve

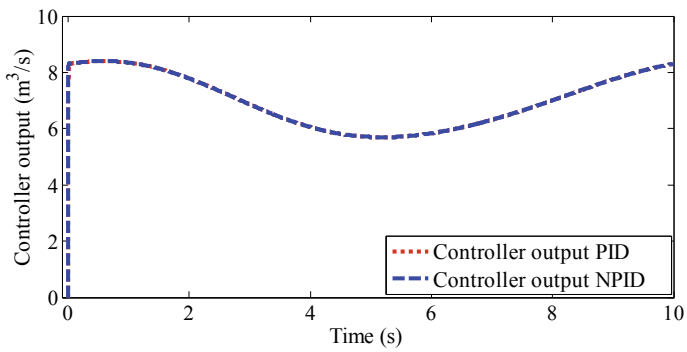


Fig. 7 Controller output curve

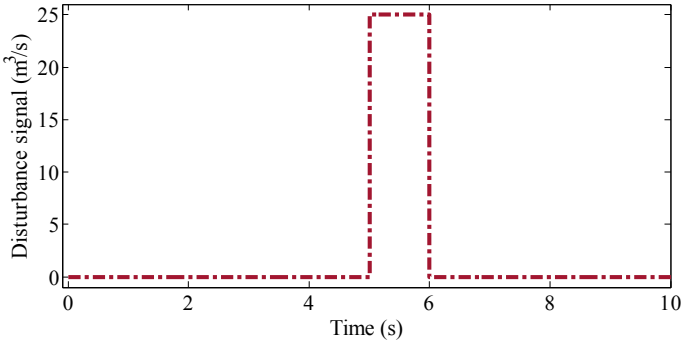


Fig. 8 Disturbance curve

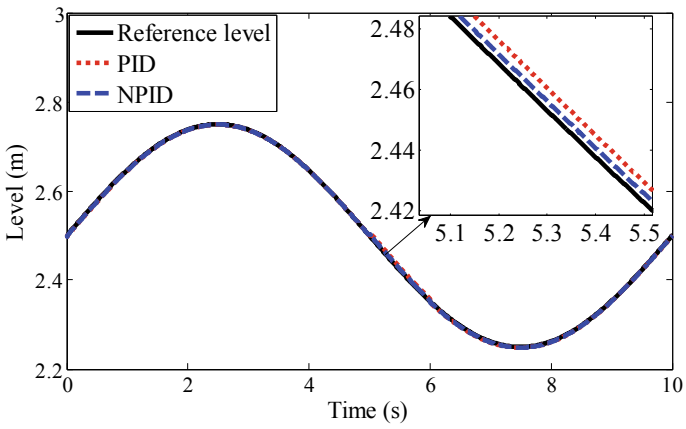


Fig. 9 Reference level tracking curve for disturbance rejection

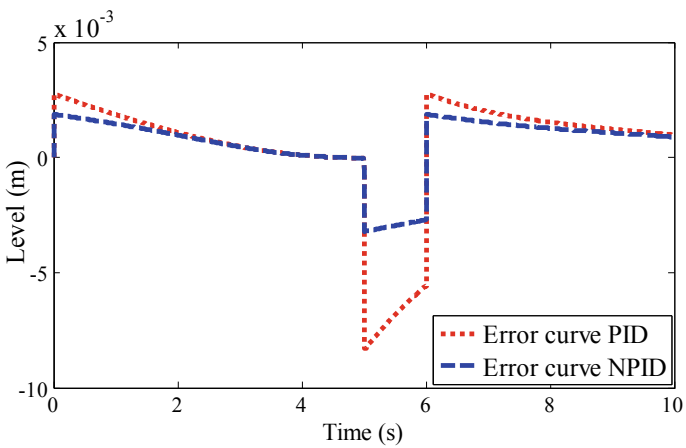


Fig. 10 Error curve

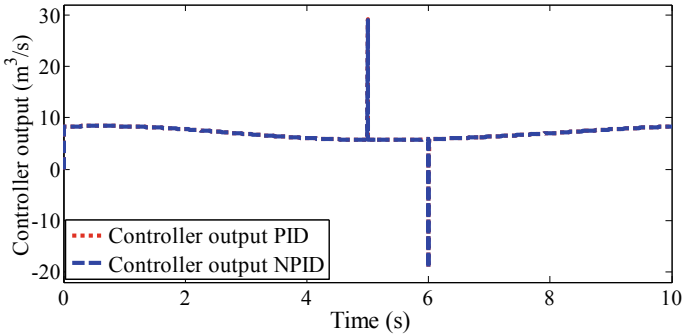


Fig. 11 Controller output

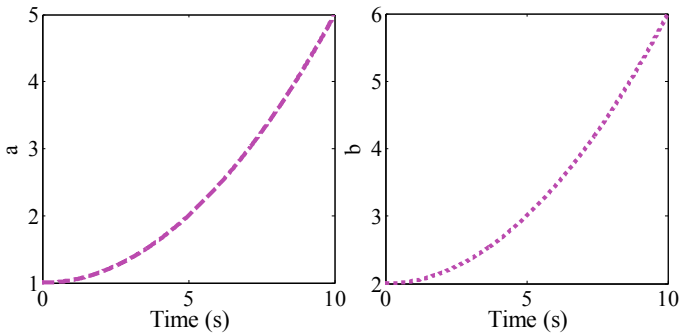


Fig. 12 Variations of system parameters *a* and *b* in run time

IAE values for this regulatory mode study, it can be easily inferred that NPID offers 33.33% better performance on PID controller.

Performance of both the controllers is also tested for parameters variations in real time. Surge tank parameters, *a* and *b*, are varying with time, and this variation is depicted in Fig. 12. Corresponding IAE values are obtained for PID and NPID controllers as 0.0080 and 0.0067, respectively. It is inferred that supremacy of 16.25% is obtained of NPID on PID controller in this case.

5 Conclusion

An adaptive controller is required for a nonlinear, non-stationary complex systems like surge tank. Linear PID controller is incapable to yield acceptable results for these complex systems. In this context, the research article presents nonlinear proportional integral and derivative (NPID) controller for controlling a surge tank. NPID controller

can be an appropriate solution concerning surge tank system. A bio-inspired procedure, specifically Cuckoo search optimization (CSA) technique, is applied in this research for defining the best gains of NPID control structure. A performance criterion for optimization is chosen as integral of absolute error (IAE). The proposed NPID is matched with state of the art for viewing its efficacy. In case of servo mechanism, performance of NPID is 16.94% better than PID controller. For regulatory mode operation, 33.33% better performance is obtained by NPID on PID controller. In case of system parameter uncertainty, 16.25% improvements have been observed by NPID on PID controller. Based on these criteria, NPID offers better performance on PID controller.

References

1. Kumar V, Nakra BC, Mittal AP (2011) A review of classical and Fuzzy PID controllers. *Int J Intell Control Syst* 16(3):170–181
2. Bennett S (2001) The past of PID controllers. *Annu Rev Control* 25:43–53
3. Ziegler, J.G. and Nichols, N.B., (1942). Optimum settings for automatic controllers. *Trans ASME* 64(11)
4. Ziegler JG, Nichols NB (1943) Process lags in automatic control circuits. *Trans ASME* 65(5):433–443
5. Cohen G (1953) Theoretical consideration of retarded control. *Trans ASME* 75:827–834
6. Stephanopoulos G (1984) *Chemical process control an introduction to theory and practice*. Pearson Education, USA. ISBN 81-7758-403-0
7. Kumar J, Kumar V, Rana KPS (2019) Fractional-order self-tuned fuzzy PID controller for three-link robotic manipulator system. *Neural Comput Appl*: 1–23
8. Åström KJ, Wittenmark B (2008) *Adaptive control*. Dover Publications, Mineola
9. Kumar V, Rana KPS, Kumar J, Mishra P (2018) Self-tuned robust fractional order fuzzy PD controller for uncertain and nonlinear active suspension system. *Neural Comput Appl* 30(6):1827–1843
10. Kler D, Rana KPS, Kumar V (2018) A nonlinear PID controller based novel maximum power point tracker for PV systems. *J Franklin Inst* 355(16):7827–7864
11. Liu GP, Daley S (2000) Optimal-tuning nonlinear PID control of hydraulic systems. *Control Eng Pract* 8(9):1045–1053
12. Jiang F, Gao Z (2001) An application of nonlinear PID control to a class of truck ABS problems. In: *Proceedings of 40th IEEE conference on decision and control*. IEEE, pp 516–521
13. Tan KK, Lee TH, Zhou HX (2001) Micro-positioning of linear-piezoelectric motors based on a learning nonlinear PID controller. *IEEE/ASME Trans Mechatron* 6(4):428–436
14. Thanh TDC, Ahn KK (2006) Nonlinear PID control to improve the control performance of 2 axes pneumatic artificial muscle manipulator using neural network. *Mechatronics* 16(9):577–587
15. Lee JY, So GB, Lee YH, So MO, Jin GG (2015) Temperature control of a CSTR using a nonlinear PID controller. *J Inst Control Robot Syst* 21(5):482–489
16. Yang XS, Deb S (2009) Cuckoo search via Lévy flights. In: *IEEE world congress on nature and biologically inspired computing (NaBIC)*. pp 210–214
17. Yang XS, Deb S (2013) Multiobjective cuckoo search for design optimization. *Comput Oper Res* 40(6):1616–1624
18. Agrawal A, Goyal V, Mishra P (2019) *Adaptive control of a nonlinear surge tank-level system using neural network-based PID controller*. *Applications of artificial intelligence techniques in engineering*. Springer, Singapore, pp 491–500

Realistic Assessment of Rural Broadband Requirements: Coverage Aspects



Pranjul Kumar, Ashwani Kumar, and Sanmukh Kaur

Abstract The scope of this paper is to conduct a study on the practical coverage requirements for extending the mobile broadband access to rural areas of India. The paper studies the vision of Digital India and National Digital Communication Policy 2018. It also outlines the challenges faced by telecom industry in India and subsequently presents the fundamentals to achieve the most cost-effective broadband penetration in the rural areas of India. The paper presents the mathematical analysis of the geographical locations of various Gram Panchayats (GPs) and evaluates the inter-GP distances between nearest pairs of GP's. We have chosen Kerala state to conduct this analysis of GP's geo-locations and inter-GP distances for all the GP's within District HQs (DHQ). The three categories of DHQs have been chosen for the analysis—densely-populated, moderately-populated, and scarcely-populated. The inter-site distances (ISD's) have been evaluated for these three distinct cases. This would facilitate the required broadband coverage to certain percentage of villages.

Keywords Mobile broadband · Rural India · Digital India · NDCP-2018 and telecom sector

P. Kumar (✉) · S. Kaur

Department of Electronics and Communication Engineering, Amity School of Engineering and Technology, Amity University, Uttar Pradesh, Noida, India

e-mail: pranjulk@yahoo.com

S. Kaur

e-mail: skaur2@amity.edu

A. Kumar

Huawei Technologies India Pvt. Ltd, New Delhi, India

e-mail: ashwani.kumar5@huawei.com

1 Introduction

1.1 Digital India

The vision of Digital India program is to have a nation which is digitally a strong society and an aware economy. Connecting every village through broadband and high-speed Internet and providing them with digital governance services, financial aids, and other social benefits will result in a well-connected nation [1].

NOFN/BharatNet Program

The Department of Telecom (DoT) established Bharat Broadband Network Ltd. (BBNL) to lay National Optical Fiber (NOFN) later renamed as BharatNet [2], so, that they can make sure that all the Gram Panchayats have high-speed connectivity. Bharat Broadband Network Ltd. will lay optic fiber network in 2,50,000 g Panchayats and provide with 100 Mbps link which will be used as information highway reaching all villages across the country. BharatNet would facilitate achievement of the stated objectives as outlined in the sub-section below:

- (a) E-Governance and services on demand
The National e-Governance Plan: to have a wholesome view of e-governance steps across the country, this was approved in 2006.
- (b) Digital empowerment of citizens

National Digital Communication Policy (NDCP-2018)

The vision of the policy is to support India's conversion as a digitally strong economy by satisfying the information and communication wants of people and businesses by establishing an overall fit digital communications infrastructure and services [3].

2 Fundamental Pillars of Optimal Approach to Rural Broadband Penetration

This topic would explain the key fundamental pillars for successful implementation of rural broadband connectivity in a cost-effective manner.

Key Aspects for Broadband Connectivity for Rural Areas

The architectural choice would be governed by the requirement to cover as many Villages as possible from a centralized location where a radio tower and power are available [4]. Such location naturally turns out to be the Gram Panchayat (GP) which usually has a telecom tower installed. Therefore, the objective is to install a base station at GP and provide coverage to adjoining villages.

- **Coverage**
There is an urgent need of coverage in every nook of rural areas and most importantly social institutions, business, and households.
- **Capacity**
The network should be able to address the minimum capacity requirements to the broadband connections in the rural areas serving from a centralized location, i.e., Gram Panchayats (GPs). In addition, it should be able to scale out to cater to the fast-growing future requirements.
- **Availability**
While provisioning broadband infrastructure, we should make sure that the broadband is available without disruption. As per targeted objectives, the broadband uptime must be taken care of [5].
- **Reliability**
The broadband infrastructure must offer the reliability of information in a secure manner.
- **Affordability**
To enhance and promote universal access, we should make sure that it is affordable for the citizens across both cost of service and device cost.

How to Ensure the Above Objectives

The technology we choose for broadband connectivity must ensure the following critical requirements for the most cost-effective and future proof on continuous basis:

- **Economy of scale**
Cost-effectiveness of deployment is the key lever of broadband plan for rural areas.
- **Global interoperability**
This goes without explanation that the access provided must be interoperable on a global basis which means that a villager should be able to avail the access while roaming nationally or internationally using the same device/handset [6].
- **Wide ecosystem availability**
Another key factor is timely and widest possible availability is ecosystem comprising devices and network infrastructure for the technology used.
- **No vendor-lock-in**
The technology should permit a choice of vendor selection initially and for up-gradation of the infrastructure with time.

3 Issues Faced by Indian Telecom Industry and Cost Pressures

Some of the key challenges and issues faced by the Telecom industry are [7]:

1. Policy and regulatory issues
Highly regulated sector, Quite high cost of compliance (CoC), Licensing provisions are restrictive and adoption of “Same Service Same Rule” is necessary for maintaining level playing field.
2. Financial condition and government levies
Despite being an essential service sector, it is a highly taxed sector, GST is 18% which is in higher slab despite being an essential basic service, LF & SUC ranges from 11–13% of operators’ revenue and AGR definition has been a contentious issue.
3. Lack of optical backhaul penetration
There is 70% excess of fiber networks in developed nations, whereas we only have 25% of mobile towers connected with fiber networks.
4. Declining average revenue per user (ARPU)
The data and ARPUs are having inverse relation, i.e. data demands are increasing and the ARPUs are decreasing. These decreasing ARPUs combined with losses has led to consolidation by Indian Telecom operators as the only way of survival.
5. Limited spectrum availability
 - The existing spectrum is less than 40 and 30% as compared to European nations and China, respectively. Also, the cost of the spectrum is expensive when compared with global average
6. Lack of telecom infrastructure in semi-rural and rural areas
 - The service providers do not see the same amount of revenue from the rural areas, which they incur on them while entering in the areas.
7. Pressure on margins due to stiff competition

As the competition is increasing with the new entry of Reliance Jio, other telecom operators are surely facing some problems, i.e. sudden decrease in tariff prices of both data and voice. Latter representing 60% of the revenue, which is completely lost now.

4 Compilation and Realistic Assessment of Inter-Site Distances (ISD) for Rural Broadband Planning

To connect the unconnected, broadband penetration of rural areas is a necessary step. To have the mobility at a cheap cost, we took the state “Kerala” to study the percentage of coverage over the distances between Gram Panchayats.

The methodology used in the evaluation is to calculate the inter-GP distances (IGD) for the nearest neighbor GP's and plot a CDF curve for the distribution of these IGDs. From the CDF curve, we can make an assessment on the required inter-site distance (ISD) required to for certain percentile of cases. We choose to find out the ISD requirement for 95th percentile of cases which can be reasonable as well as conservative measure to arrive at the representative ISD to plan a mobile broadband network in the cluster of GP's. This is a fair assumption to make that 95th of the villages would be covered if this ISD is used to plan the network coverage. As a next step to the outcome of this study, we can carry out the planning on a continuous grid of cells where the mobile broadband tower is installed at the centers of the GP's in a 19-cell layout. In this study, we have chosen Kerala state to conduct this analysis of GP's geo-locations and inter-GP distances for all the GP's within a District HQs (DHQ). We have chosen Kerala state as its population is divided uniformly over all the three categories, i.e. highly dense, moderately dense, and scarcely dense (Fig. 1).

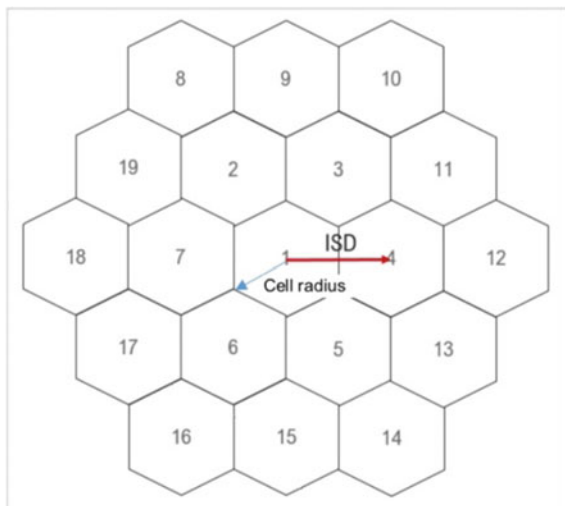
The geographical data was taken from BBNL website. The state was divided into three areas on the basis of density: high dense, semi-dense, and scarcely dense.

Studied the demographic data and develop a representative model comprising number of Gram Pachayats and districts in a particular state. After the data collection, performed a program (Python code) for calculating the distance between each Gram Panchayat that was falling under the same district for the state. These are some of the datasets of different districts whose distance was calculated (Figs. 2 and 3).

Highly Dense Gram Panchayats (GPs)

If we analyze the CDF plot of the ISD distribution, we observe that the corresponding value of the ISD for 95th percentile of cases is 5.05 km, which is a indication that in case of densely populated areas, roughly 95% of the villages would be covered with

Fig. 1 19-Cell geometry



```
[ ] import pandas as pd
import numpy as np
from pandas import ExcelWriter
from pandas import ExcelFile
df= pd.read_excel("Kerala(High).xlsx")
df
```

	STATE	DISTRICT	BLOCK	GP_NAME	LAT	LONG
0	KERALA	THIRUVANANTHAPURAM	ATHIYANNOOR	ATHIYANNOOR	8.393614	77.063736

Fig. 2 Code for the ISD calculation

```
R = 6373.0
for i in range(len(dataset2)) :

    print(dataset2.loc[i, "LAT"], dataset2.loc[i, "LONG"])

    lat1 = radians(dataset2.loc[48, "LAT"])
    lon1 = radians(dataset2.loc[48, "LONG"])
    lat2 = radians(dataset2.loc[i, "LAT"])
    lon2 = radians(dataset2.loc[i, "LONG"])

    dlon = lon2 - lon1
    dlat = lat2 - lat1

    a = sin(dlat / 2)**2 + cos(lat1) * cos(lat2) * sin(dlon / 2)**2
    c = 2 * atan2(sqrt(a), sqrt(1 - a))

    distance = R * c
    print("Distance:",distance)
```

Fig. 3 Code for ISD calculation

a base station deployed at a tower at GP level with sufficient coverage to reach about 2.9 km (Fig. 4).

Moderately Dense Gram Panchayats (GPs)

If we analyze the CDF plot of the ISD distribution, we observe that the corresponding value of the ISD for 95th percentile of cases is 6.7 km, which is an indication that in case of densely populated areas, roughly 95% of the villages would be covered with a base station deployed at a tower at GP level with sufficient coverage to reach about 3.87 km (Figs. 5, 6 and 7).

Scarcely Dense Gram Panchayats (GPs)

If we analyze the CDF plot of the ISD distribution, we observe that the corresponding value of the ISD for 95th percentile of cases is 7.9 km, which is an indication that in case of densely populated areas, roughly 95% of the villages would be covered with base station deployed at a tower at GP level with sufficient coverage to reach about 4.56 km (Figs. 8, 9 and 10).

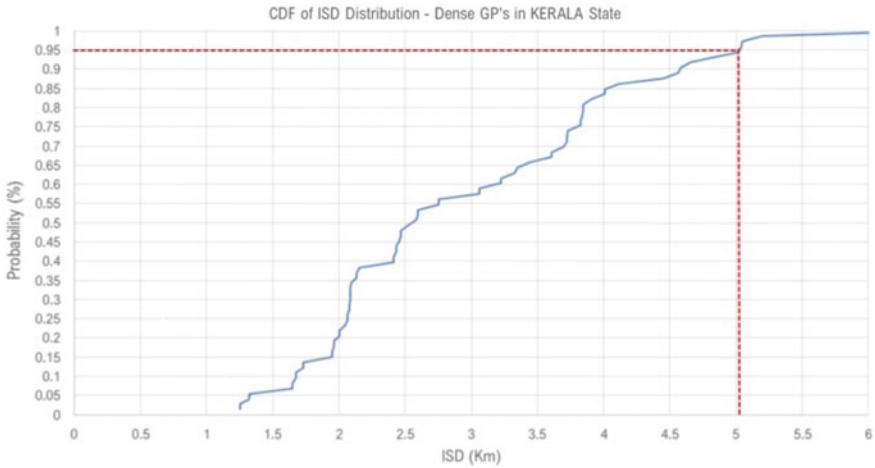


Fig. 4 CDF plot for ISD distribution

```
import pandas as pd
from math import sin, cos, sqrt, atan2, radians

dataset = pd.read_csv('showimg (19).csv')

dataset
```

	STATE	DISTRICT	BLOCK	GP_NAME	LAT	LONG
0	KERALA	MALAPPURAM	AREAKODE	AREEKODE	11.23451	76.05001

Fig. 5 CDF plot for ISD distribution

```
R = 6373.0
for i in range(len(dataset2)) :

    print(dataset2.loc[i, "LAT"], dataset2.loc[i, "LONG"])

    lat1 = radians(dataset2.loc[48, "LAT"])
    lon1 = radians(dataset2.loc[48, "LONG"])
    lat2 = radians(dataset2.loc[i, "LAT"])
    lon2 = radians(dataset2.loc[i, "LONG"])

    dlon = lon2 - lon1
    dlat = lat2 - lat1

    a = sin(dlat / 2)**2 + cos(lat1) * cos(lat2) * sin(dlon / 2)**2
    c = 2 * atan2(sqrt(a), sqrt(1 - a))

    distance = R * c
    print("Distance:", distance)
```

Fig. 6 Code for ISD calculation

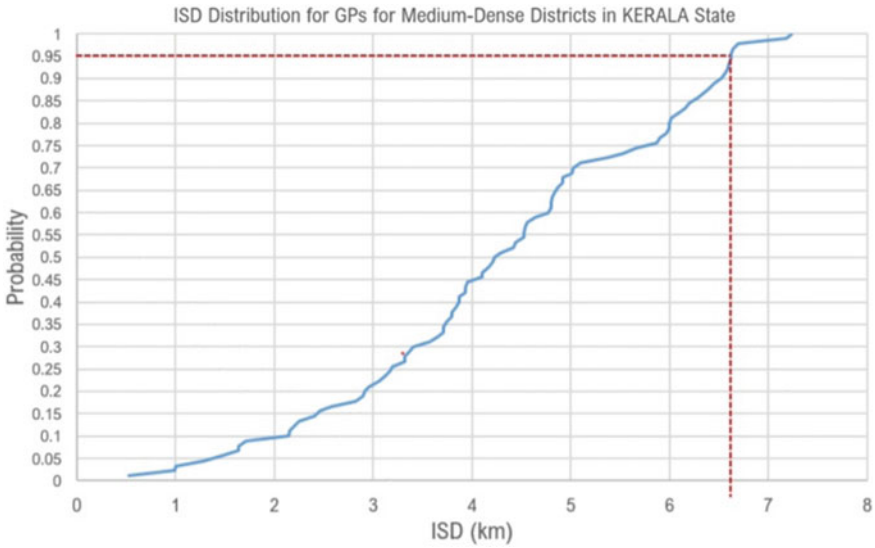


Fig. 7 CDF plot for ISD distribution

Fig. 8 Code for ISD calculation

```
[ ] import pandas as pd
import numpy as np
from pandas import ExcelWriter
from pandas import ExcelFile
df = pd.read_excel("showimg.xlsx")
df
```

	DISTRICT	GP_NAME	LAT	LONG
0	IDUKKI	Rajakkad	9.961191	77.096912

5 Conclusions

We analyzed the BharatNet fiber penetration data to Gram Panchayats. Then we calculated the intra-GP distance in the districts of a state. Through the analyzed data, we could plot a CDF of distance between the Gram Panchayats versus the percentile of coverage and therefore made a conclusion on range of area for the 95% coverage.

As the range would also depend over the density of the area therefore, we have classified our districts on the basis of CDF plotted through highly dense area, moderately dense area, and scarcely dense area. There is no as such incremental change in the technology as we are using the existing technology for the realistic assessment

```
from math import sin, cos, sqrt, atan2, radians
R=6367.0
for i in range(len(df1)):

    lat1=radians(df1.loc[50,"LAT"])
    lon1=radians(df1.loc[50,"LONG"])
    lat2=radians(df1.loc[i,"LAT"])
    lon2=radians(df1.loc[i,"LONG"])

    dlon= lon2-lon1
    dlat=lat2-lat1
    a = sin(dlat / 2)**2 + cos(lat1) * cos(lat2) * sin(dlon / 2)**2
    c = 2 * atan2(sqrt(a),sqrt(1-a))
    distance=R*c
    print(distance)
```

Fig. 9 Code for ISD calculation

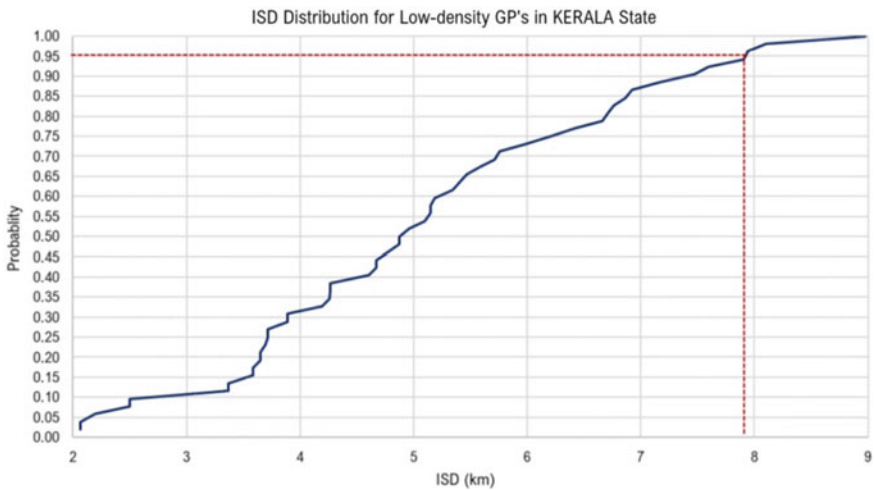


Fig. 10 CDF plot for ISD distribution

of the Rural Broadband with respect to the Coverage. From the analysis and evaluation, it is concluded that for the highly dense, moderately dense, and finally scarcely dense areas, the representative inter-site distances (ISDs) to cover 95% of villages are 5.5 km, 6.7 km, and 7.9 km, respectively.

References

1. Bharat Broadband Network Limited (2004) A Government of India undertaking. <https://www.bbnl.nic.in/>. Accessed 26 May 2020
2. Ramasetty P, Masilamani S (2019) 5G rural strategy in India. In: Paper presented at the optical fiber communications conference and exhibition (OFC), San Diego, California, United States, 3–7 March 2019
3. Gambhiri H, Singla K (2018) A review paper on broadband connectivity in rural India through TV white space. *Int J Adv Res Ideas Innov Technol* 4:91–96
4. Jha A, Saha D (2017) Why is 700 MHz band a good proposition for provisioning pan India 4G LTE services? A comparative techno-economic evaluation. In: Paper presented at the 9th international conference on communication systems and networks (COMSNETS), Bangalore, India, 4–8 January
5. Asif Ahmad AS, Keshavamurthy B, Abhay Narasimha KS, Mahesh N, Suma MN (2016) Communication system design for white-fi (802.11af). In: Paper presented at the IEEE international conference on advanced networks and telecommunications systems (ANTS), Bangalore, India, 6–9 November
6. Chaitanya Prasad N, Deb S, Karandikar A (2016) Feasibility study of LTE middle-mile networks in TV white spaces for rural India. In: Paper presented at the IEEE 27th annual international symposium on personal, indoor, and mobile radio communications (PIMRC), Valencia, Spain, 4–7 September 2016
7. Jha A, Saha D (2015) Techno-economic assessment of the potential for LTE based 4G mobile services in rural India. In: Paper presented at the 9th IEEE international conference on advanced networks and telecommunications systems (ANTS), Indian Statistical Institute (ISI), Kolkata, 15–18 December 2015

Fractional-Order Load Frequency Control of a Two-Area Interconnected Power System with Uncertain Actuator Nonlinearities



Akhilesh Kumar Mishra, Puneet Mishra, and H. D. Mathur

Abstract Nonlinearities in power system cause severe degradation in quality of power being generated. A robust controller is often needed to compensate for uncertainties caused by nonlinear components present in the control loop. In this paper, a robust fractional order controller is designed and investigated for tackling with inherent nonlinearities in an interconnected two-area non-reheated thermal power system. The presence of these inherent nonlinearities is due to generation rate constraint (GRC) and governor deadband (GDB), which can cause delayed disturbance rejection and/or sustained oscillations in power system variables. On considering both nonlinearities simultaneously into account, the complexity in controlling the power system increases, and conventionally optimized controllers fail to produce satisfactory dynamic performance. To address these issues, a robust fractional-order-proportional-integral-derivative (FOPID) controller has been designed utilizing a lately introduced Salp Swarm Algorithm (SSA). Extensive simulation studies have been performed and comparative studies have been drawn with controllers designed with available techniques in literature. Based on investigations carried out in the article, it is found that SSA optimized FOPID shows remarkable enhancement investigated in load frequency control performance of investigated plant in presence of significant parametric variations in comparison with particle swarm optimization optimized controllers.

Keywords Automatic generation control (AGC) · Actuator nonlinearity · Fractional-order calculus · Salp swarm algorithm (SSA)

A. K. Mishra · P. Mishra (✉) · H. D. Mathur
Birla Institute of Technology and Science, Pilani Campus, Pilani 333031, Rajasthan, India
e-mail: puneet.mishra@pilani.bits-pilani.ac.in

A. K. Mishra
e-mail: akhileshmishra.bits@gmail.com

H. D. Mathur
e-mail: mathurhd@pilani.bits-pilani.ac.in

1 Introduction

The frequency regulation is our prime concern in interconnected electrical power system (IEPS) to have a quality power. As the IEPS consists of diverse control areas, they are coupled via an electrical linkage commonly referred to as tie lines with one another. If the load demanded by the consumer from any control area gets altered, then it leads to deprivation of the dynamic performance of the IEPS due to variation in frequencies as well as scheduled electrical energy transfer via tie-lines from their predefined values. These deviations can be mitigated partially by the speed regulation mechanism present in IEPS by regulating the generator output by varying the valve position. The speed governing mechanism is also referred to as the primary control, but to augment the dynamic performance of IEPS, supplementary controller must add with primary controller, which can provide persistent frequency and tie-line power exchange (TLPE). Automatic generation control (AGC) or load frequency controller (LFC) is typically utilized to deal with such scenarios and their main objective makes the deviations of different control areas frequency as well tie-line power to zero [1].

In load frequency controller (LFC), the area control error (ACE) acts as an input to LFC, where ACE is the combination of deviation in TLPE and frequency deviation multiplied with frequency bias constant. In order to implement LFC in a realistic IEPS, the various inherent nonlinearities found must be incorporated. The utmost noteworthy nonlinearities present in load frequency control loops are governor dead-band (GDB) and generation rate constraint (GRC). The governor deadband (GDB), which mainly comes into existence due to either friction or physical geometry of rack and pinion arrangement in speed governing mechanism. The purpose of the speed governing mechanism is to rotate the camshaft that operates the control valves to manipulate steam input to the turbine. The GDB has disrupting nature on the transient as well steady-state response, as it can produce continuous oscillation in the response of frequency and TLPE [2]. Whenever an excessive steam is demanded from the boiler system to increase the generated power instantly, then due to adiabatic process occurring in boiler steam get condensed and causes reduction in life span of the turbine blades of thermal power plant. Hence for the acceptable operation of boiler system, with the help limiters, the maximum speed of opening and closing of the valve is constrained. The GRC curbs the ability of instant disturbance rejection in IEPS [3]. In IEPS, these nonlinearities can severely disrupt the performance of LFC. Many researches have proposed the LFC structure for IEPS with considering various aspects of modern IEPS and [4, 5] have presented an exhaustive literature review [6].

In the contrast of literature discussed and Table 1, we can infer that usually the nonlinearities associated with the interconnected power system have been either ignored or considered individually moreover its variation effect never considered simultaneously. Nevertheless, this is of prime importance to consider the simultaneous presence of GRC and GDB for the accurate implementation of LFC. Because their simultaneous presence has an undesirable effect on the dynamic performance of the system and can cause longer settling time with load frequency variation and

Table 1 A brief summary of LFC for IEPS with system nonlinearities

References	System under investigation	Non-linearities	Controller structure	Optimization algorithm	Performance index
[9]	2-area	GRC	PID	GA, BFOA	ITAE
[7]	2-area	None	FOPID	PSO	ITAE
[10]	3-area	GRC, GDB	Fuzzy FOPID	COA	ITSE
[11]	2, 4-area	None	FOPID	BBBC	ISE
[12]	4-area	None	Fuzzy FOPID	BBO	ITAE, ITSE
[13]	2-area	GDB, GRC	FOPID	IPSO	ITSE
[14]	3-area	GDB, GRC	PID	FA	ITAE
[15]	5-area	GRC	IDD	BFOA	ITAE
[16]	2-area	GRC, GDB	IMC-FOPID	None	ITAE
[17]	2-area	GDB	FOPID	GBMO	ITAE

large oscillations in tie-line power. Therefore, for load frequency control problem inherent nonlinearities associated with the power system needs further attention with variation in magnitude of GRC and GDB to achieve more refined control operation to have better power quality in terms of lesser frequency deviations and reduced tie-line power deviations. The integral order-based classical controller is not much efficient to provide satisfactory dynamic performance, under significant change in magnitude for step load perturbation [7]. To address the aforementioned issues, an effort has been made for investigation of two-area non-reheated thermal IEPS in this work. The simultaneous presence of GRC and GDB in the power system is addressed by using an optimal fractional order PID controller optimized by, i.e., Salp swarm algorithm (SSA) introduced by Mirjalili et al. in [8].

Further, the present work has been structured as follows; Sect. 2 deals with mathematical linearized model of system under investigation, i.e., a two-area IEPS system with GRC and GDB for LFC issues. In Sect. 3, brief introduction of fractional order controller, SSA as an optimization algorithm and selection criterion of objective function for LFC issue have been discussed. Section 4 focuses on simulation results and associated discussions. At last, concluding remarks are made in Sect. 5.

2 Dynamic Model of the Interconnected Power System

IEPS is a composite dynamical system with numerous generator and loads. Typically, IEPS investigated for LFC performance are subjected to very small load variations in contrast with their rated capacity. Therefore, a linearized model is usually utilized for the present investigation. A realistic IEPS with GRC and GDB nonlinearities incorporated simultaneously has been presented in Fig. 1. The extensively used two-area

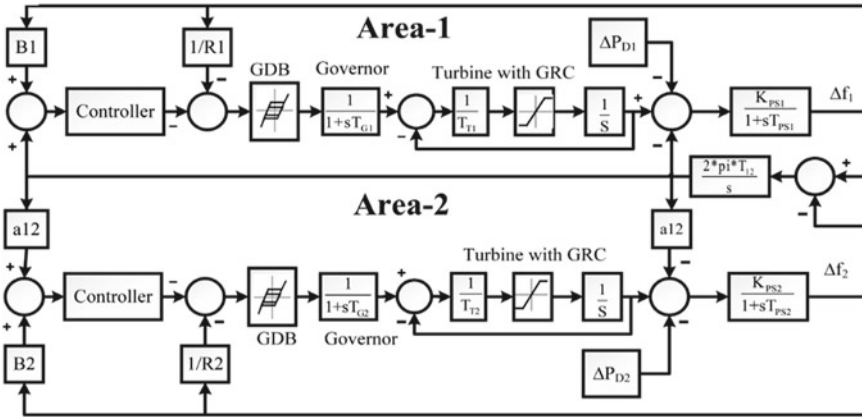


Fig. 1 Block diagram of the investigated IEPS, i.e., Non-reheated thermal-thermal two-area interconnected electric power system incorporating GDB and GRC

non-reheated IEPS for the investigation of the LFC performance under nominal operating condition adopted power from [9, 14]. The investigated IEPS for the presented work has been represented in Fig. 1. The system nominal parameters under investigation have been adopted from [9, 14] and presented in Appendix 1. Further in subsequent section, the control structure for the investigated LFC and its optimal controller parameters utilizing optimization algorithm has been presented to obtain the anticipated control objectives.

3 Load Frequency Controller Structure and Parameter Tuning

The fractional calculus offers the freedom to the operator for implementation of the non-integer or differential order operator, which can have a superior level of performance in terms of robustness when compared with integer order. In the literature, several approximation methods are proposed. Some simplified continuous-time domain approximations are ‘Crone or Oustaloup approximation,’ ‘Carlson approximation,’ ‘Matsuda approximation,’ etc. Present articles utilize the ‘Oustaloup approximation’ for approximating the fractional-order integrator and differentiator, because of its very decent fitting to fractional-order elements. This approximation technique performs the fractional-order integral and differential operators via a higher-order analog filter having an order of ‘ $2N + 1$ ’ within a specified frequency bound $[\omega_L, \omega_H]$ [16]. The value of ‘ N ’ and frequency bound $[\omega_L, \omega_H]$ for present work has been considered as 5 and $[10^{-3}, 10^3]$ respectively, and well accepted in the LFC application. For the sake of brevity, the details of implementation of fractional-order operator have been omitted here and can be found in [16, 18]. The fractional-order

PID controller can represent ($PI^\lambda D^\mu$) as:

$$G_c(s) = K_P + K_I S^{-\lambda} + K_D S^\mu \tag{1}$$

To obtain the desired control objective, the controller parameters must be accustomed. In the last decade, evolutionary computing and swarm intelligence based methods have grown substantial consideration from researchers around the world. Soft computing techniques have been used extensively for tuning the LFC parameter to achieve superior dynamic system performances [19]. Various popular optimization algorithms used for finding the optimal controller parameters to solve the LFC problem effectively have been presented in Table 1. Out of these algorithms, SSA has appeared as a robust optimization algorithm with only one tunable factor, rapid convergence to optimal solutions with the tendency to avoid local optimum points.

Like other optimization algorithms, a n -dimensional search space is created to represent the positions of salps, where n is the number of decision variable for a given problem. A two-dimensional matrix denoted by ‘ x ’ is used for storing the positions of salps. The food source ‘ S ’ has been assigned as the swarm’s target in the search space. Equation (2) is used for the updating for the swarm positions in the search space as follows:

$$x_i^1 = \begin{cases} S_i + A_1((ul_i - ll_i)A_2 + lbr) \text{ for } A_3 \geq 0.5 \\ S_i - A_1((ul_i - ll_i)A_2 + ll_i) \text{ for } A_3 < 0.5 \end{cases} \tag{2}$$

$$A_1 = 2e^{-(4l/Max_iter)^2} \tag{3}$$

where ‘ S_i ’ represents i th dimension the position of the food source, ‘ ul_i ’ designates the i th dimension of upper limit, ‘ ll_i ’ designates the i th dimension of lower bound of, x_i^1 represents the i th dimension position of the first salp and the uniformly distributed random numbers in interval [0, 1] represents the A_2 and A_3 . “ l ” is current iteration.

The follower’s (k^{th}) positions in the search space (in i^{th} dimension) modified utilizing the newton law of motion and given by the following equation considering initial speed (v_0) to be zero.

$$x_i^k = \frac{1}{2}\alpha t^2 + v_0 t \tag{4}$$

The ‘ i ’ will always be greater than or equals to two, further, $\alpha = \frac{v_{final}}{v_0}$, $v = \frac{x-x_0}{t}$ and ‘ t ’ the time taken for completing one iteration of optimization. Hence Eq. (4), can be restructured as

$$x_i^k = \frac{1}{2}[x_i^k + x_i^{k-1}] \tag{5}$$

Further, SSA algorithm outline [8] can be present as follows.

Set population of salp x_i ($i=1, 2, \dots, n$), subjected to 'ul' and 'll'
while (end condition is not fulfilled)
 Fitness of individual salp is calculated and allocate it as 'F' as best that individual salp
 Calculate and modify 'A₁' by Eq.3
 for each individual salp (x_i)
 if ($i=1$)
 By utilizing the Eq. 2 modify the leader salp position
 else
 Utilizing the Eq. 5 modify the follower salp position
 end
 On the basis of 'll' and 'ul' variables, modify the salps variables
end
Return S

3.1 Load Frequency Controller Design

It needs to mention here that various studies have been presented in the literature for LFC design, but very few of these have pertained attention to the existing actuator nonlinearities in the power system while focusing on the LFC design. In the current work, integral time absolute error (ITAE) is used as an objective function for obtaining the PID/FOPID controller because the ITAE-optimized controllers have lesser settling time as well overshoot as compared with others. Equation (6) represents the expression for the considered objective function.

$$J = \text{Objective Function (ITAE)} = \int_0^{t_{sim}} (|\Delta f_1| + |\Delta f_2| + |\Delta P_{tie12}|) t dt \quad (6)$$

where, $|\Delta f_1|$, $|\Delta f_2|$, $|\Delta P_{tie12}|$ and ' t_{sim} ' are the absolute values of deviation of frequency in control area-1, 2, tie-line power deviation, and simulation time, respectively. Figure 2a and b shows the convergence of the objective function for the PID and FOPID controller achieved by PSO and SSA for the design of LFC for a two-area IEPS. It may be noted that the convergence of SSA is more rapid in comparison with PSO for integer as well as fractional order PID controller. It also provided a better (lesser) value of the considered objective function (ITAE), i.e., a mere value of 3.928 and 3.203 in with SSA for PID and FOPID controller, respectively. Clearly, with reference to Fig. 3, the SSA-tuned FOPID (SSAFOPID) controller was a true winner under nominal conditions in comparison with PSO tuned FOPID. The optimized controller parameters of all the controllers under consideration are listed in Table 2.

The parameters of each controller employed in individual areas are considered as same to due identical structure of both control areas. Further, the frequency and tie-line power deviation suppression in both the areas for optimized controllers, under

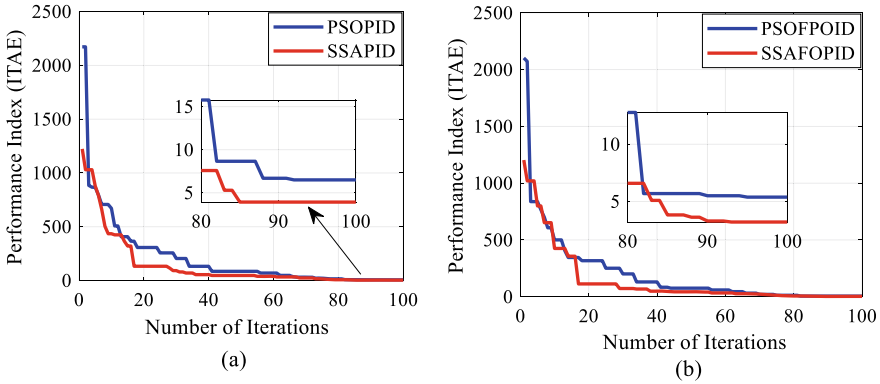


Fig. 2 Comparative convergence curves for PSO and SSA (a) PID controller (b) FOPID controller

Fig. 3 Values of 'J' for different controllers under nominal system parameters for realistic two-area interconnected power system

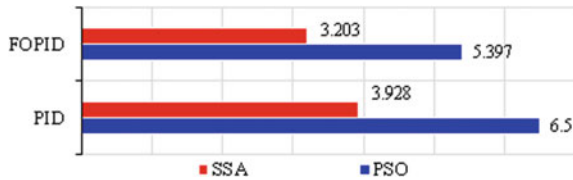


Table 2 Optimized controller parameters for realistic two-area interconnected power system with GRC and GDB

Parameter	PID		FOPID	
	PSO	SSA	PSO	SSA
K_p	13.085	8.9178	4.580	12.161
K_i	0.169	0.0374	7.0372	0.069
K_d	13.29	14.657	12.032	14.75
λ	–	–	0.024	0.951
μ	–	–	1.144	1.1912

nominal conditions, are shown in Fig. 4. Figure 4 shows the variation of frequency in control area-1, 2 and tie-line power exchange with PID controller structure. It may be noted from the above results that the SSA-tuned PID controller (SSAPID) provides minimal deviation in frequency deviations in comparison with PSO-tuned PID controller (PSOPID). However, this performance can be further enhanced with the use of fractional-order PID controllers, which provide more flexibility in the controller design. As can be seen from Fig. 5, the deviation of frequency in area-1, area-2 and tie-line power exchange, shows lesser deviations with FOPID controller as compared PID.

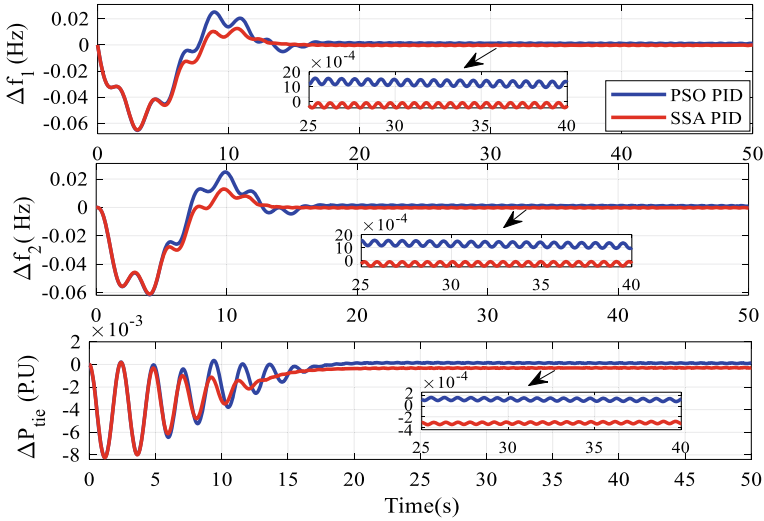


Fig. 4 Variation in Δf_1 , Δf_2 and ΔP_{tie} , under step load perturbation (SLP of 0.01 pu in area1 with nominal parameters for PID controllers

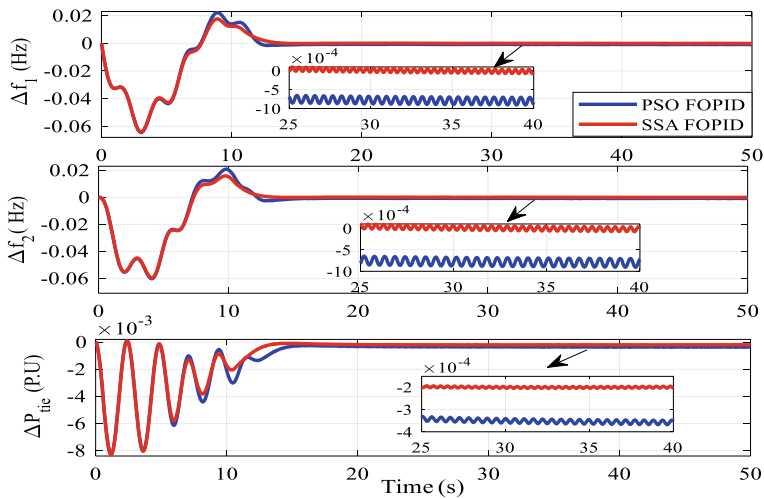


Fig. 5 Variation in Δf_1 , Δf_2 and ΔP_{tie} , under step load perturbation (SLP) of 0.01 pu in area1 with nominal parameters for FOPID controllers

4 Simulation Result and Discussion

This work presents a study on the use of optimally tuned fractional order controllers for the efficient design of LFC for system under investigation. As mentioned earlier,

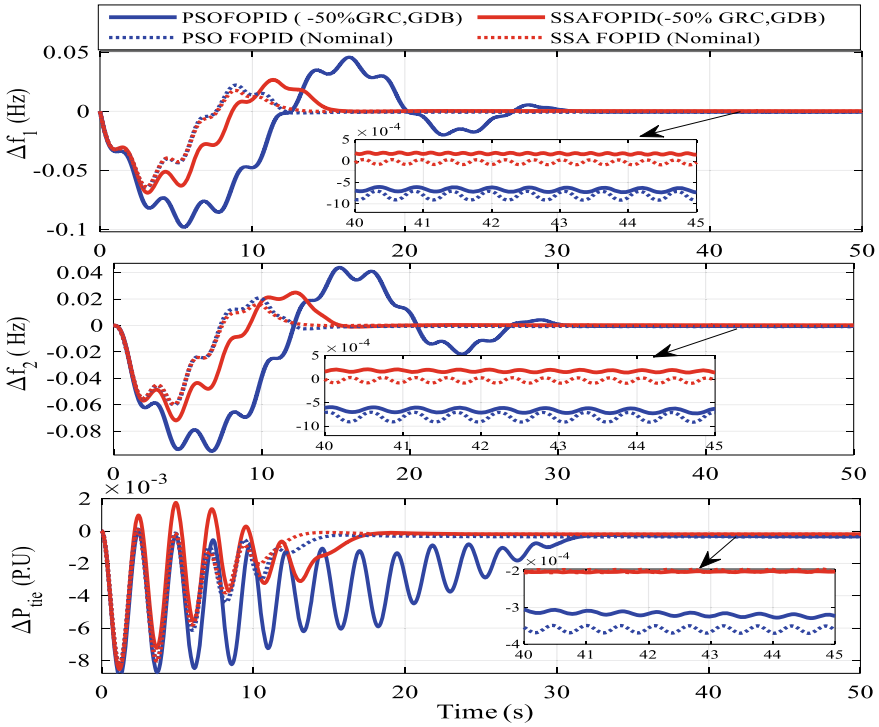


Fig. 6 The ‘ Δf ’ suppression in area-1, 2 and ‘ ΔP_{tie} ’ for -50% deviations in GRC and GDB from their nominal values

the presence of nonlinearities in the system can significantly affect the control performance.

4.1 LFC Performance Analysis with Uncertainty in GRC and GDB

This work is quite probable that the measurements of different nonlinear components characteristics be uncertain, and their actual values may be different from the one which is estimated by some means. This variation in parameter values can further bring a significant degradation of control performance or in some cases even instability. The variation of GRC and GDB has been recognized in the steps of 10% from their nominal values up to a maximum deviation of 50%. Negative deviations are considered in this study since GRC is a crucial parameter that specifies the control performance and a decrement in GRC will seriously affect the instant disturbance rejection capabilities of the system.

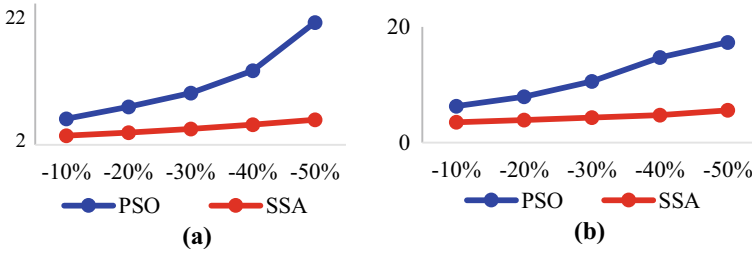


Fig. 7 Variation in ‘J’ values for parametric uncertainty in (a) GRC and GDB, (b) GRC, GDB and time constant values with different FOPID controllers

Figure 6 shows the variation in Δf_1 , Δf_2 and tie-line power exchange for a limiting case where a deviation of -50% is considered in GRC as well as GDB. This study proves the supremacy of the SSAFOPID controller for the robustness requirement, where a significant deviation in nonlinear component parameters can appear in the control loop. Further, a pictorial depiction of variation in cost function (J) is also presented in Fig. 7 for better interpretation of the control performance improvement achieved by SSAFOPID in comparison with PSOFOPID controller.

5 Conclusion

In the present work, a robust fractional order controller optimized via of PSO and salp swarm algorithm (SSA) has been explored for an enhanced load frequency control in nonlinear interconnected electrical power system (IEPS). The investigated IEPS is nonlinear in the sense, as we incorporated the nonlinear generation rate constraint (GRC) and governor deadband (GDB) in control loop. Comparative analyses based on simulation studies under nominal as well as against magnitude variation of nonlinearities, i.e. GRG and GDB show that SSA optimized FOPID controller is able to control the mitigate the any limit cycles originated due to governor deadband as well as reduces the settling time as well in comparison with PSO FOPID controller.

Appendix 1

The system nominal parameters under investigation are represented as follows [9, 14, 18] $P_{R1} = P_{R2} = 2000$ MW (rating), $PL_1 = PL_2 = 1000$ MW (nominal loading), $f = 60$, $T_{PS1} = T_{PS2} = 20$ s; $T_{T1} = T_{T2} = 0.3$ s; $2 * \pi * T_{12} = 0.545$ pu; $T_{G1} = T_{G2} = 0.08$ s; $K_{PS1} = K_{PS2} = 120$ Hz/pu MW; $a_{12} = -1$; $R_1 = R_2 = 2.4$ Hz/pu MW; $B_1 = B_2 = 0.425$ pu MW/Hz, GDB = 0.06 pu and GRC = 10% puMW/min.

References

1. Kundur P (2009) Power system stability and control. Tata McGraw Hill, New Delhi
2. Concordia C (1969) Effect of prime-mover speed control characteristics on electric power system performance. *IEEE Trans. Power Apparatus Syst* (5):752–756
3. Morsali J (2014) Appropriate generation rate constraint (GRC) modeling method for reheat thermal units to obtain optimal load frequency controller (LFC). <https://doi.org/10.1109/CTPP.2014.7040611>
4. Ibraheem A, Kumar P, Kothari DP (2005) Recent philosophies of automatic generation control strategies in power systems. *IEEE Trans Power Syst* 20(1):346–357. <https://doi.org/10.1109/TPWRS.2004.840438>
5. Pandey SK, Mohanty SR, Kishor N (2013) A literature survey on load-frequency control for conventional and distribution generation power systems. *Renew Sustain Energy Rev* 25:318–334. <https://doi.org/10.1016/j.rser.2013.04.029>
6. Debbarma S, Saikia LC, Sinha N (2014) Robust two-degree-of-freedom controller for automatic generation control of multi-area system. *Int J Electr Power Energy Syst* 63:878–886. <https://doi.org/10.1016/j.ijepes.2014.06.053>
7. Alomoush MI (2010) Load frequency control and automatic generation control using fractional-order controllers. *Electr Eng* 91(6):357–368. <https://doi.org/10.1007/s00202-009-0145-7>
8. 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. <https://doi.org/10.1016/j.advengsoft.2017.07.002>
9. Ali ES, Abd-Elazim SM (2013) BFOA based design of PID controller for two area load frequency control with nonlinearities. *Int J Electr Power Energy Syst* 51:224–231. <https://doi.org/10.1016/j.ijepes.2013.02.030>
10. Gheisarmejad M, Khooban MH (2019) Design an optimal fuzzy fractional proportional integral derivative controller with derivative filter for load frequency control in power systems. *Trans Inst Meas Control* 41(9):2563–2581. <https://doi.org/10.1177/0142331218804309>
11. Kumar N, Tyagi B, Kumar V (2017) Deregulated multiarea AGC scheme using BBBC-FOPID controller. *Arab J Sci Eng* 42(7):2641–2649. <https://doi.org/10.1007/s13369-016-2293-1>
12. Mohammadikia R, Aliasghary M (2018) A fractional order fuzzy PID for load frequency control of four-area interconnected power system using biogeography-based optimization. *Int Trans Electr Energy Syst* 29(2):1–17. <https://doi.org/10.1002/etep.2735>
13. Morsali J, Zare K, Tarafdar Hagh M (2018) Comparative performance evaluation of fractional order controllers in LFC of two-area diverse-unit power system with considering GDB and GRC effects. *J Electr Syst Inf Technol* 5(3):708–722. <https://doi.org/10.1016/j.jesit.2017.05.002>
14. Padhan S, Sahu RK, Panda S (2014) Application of firefly algorithm for load frequency control of multi-area interconnected power system. *Electr Power Compon Syst* 42(13):1419–1430. <https://doi.org/10.1080/15325008.2014.933372>
15. Saikia LC, Nanda J, Mishra S (2011) Performance comparison of several classical controllers in AGC for multi-area interconnected thermal system. *Int J Electr Power Energy Syst* 33(3):394–401. <https://doi.org/10.1016/j.ijepes.2010.08.036>
16. Saxena S (2019) Load frequency control strategy via fractional-order controller and reduced-order modeling. *Int J Electr Power Energy Syst* 104:603–614. <https://doi.org/10.1016/j.ijepes.2018.07.005>
17. Zamani A, Barakati SM, Yousofi-Darmian S (2016) Design of a fractional order PID controller using GBMO algorithm for load–frequency control with governor saturation consideration. *ISA Trans* 64:56–66. <https://doi.org/10.1016/j.isatra.2016.04.021>
18. Mishra AK, Mishra P (2019) Improved fractional order control of a nonlinear interconnected power system using Salp Swarm algorithm. In: 2019 IEEE 16th India council international conference (INDICON). pp 1–4. <https://doi.org/10.1109/INDICON47234.2019.9029023>

19. Kalavani F, Zamani-Gargari M, Mohammadi-Ivatloo B, Rasouli M (2019) A contemporary review of the applications of nature-inspired algorithms for optimal design of automatic generation control for multi-area power systems. *Artif Intell Rev* 51(2):187–218. <https://doi.org/10.1007/s10462-017-9561-7>

Robust Non-integer Control of a Nonlinear Two-Area Interconnected Power System Subjected to Large Parametric Variations



Akhilesh Kumar Mishra, Puneet Mishra, and H. D. Mathur

Abstract The article presents the load frequency control of an interconnected two-area non-reheated thermal (TANRT) power system in consideration of its inherent nonlinearity present in speed regulation mechanism as governor deadband (GDB) and generation rate constraint (GRC) found in turbine has been studied via utilizing a robust fractional order controller. The community conducts of Salpidae exist in the subterranean sea is the motive behind the newly introduced optimization algorithm known as salp swarm algorithm (SSA), which has been used to obtain the optimal controller parameters. In this work, exclusive simulation studies were performed to establish that SSA-optimized fractional-order controller provides noteworthy enhancement in the dynamic performance of the system with simultaneous presence of inherent nonlinearities, i.e., GRC and GDB under large parametric variations in contrast with GA optimized controllers.

Keywords Interconnected power system · Load frequency control · Fractional calculus · Salp swarm algorithm

1 Introduction

To maintain the nominal operating condition, the electric power systems must be integrated. The tie-line is used for integrating the multiple control areas of an interconnected power system (IPS). The alteration in active power demand results in deviation of tie-line power exchange and frequencies of various interconnected conventional generating units from their assigned values. As a result, the dynamic performance of the system can be significantly affected in terms of deviation in the frequency of

A. K. Mishra · P. Mishra (✉) · H. D. Mathur
Birla Institute of Technology and Science, Pilani Campus, Pilani, Rajasthan 333031, India
e-mail: puneet.mishra@pilani.bits-pilani.ac.in

A. K. Mishra
e-mail: akhileshmishra.bits@gmail.com

H. D. Mathur
e-mail: mathurhd@pilani.bits-pilani.ac.in

generated power and scheduled tie-line power exchange as well. The speed governor mechanism can partially compensate for these deviations by adjusting the position of the valve, which has installed at the power generation utility. Valve position changes the input for the turbine and consequently results in a change in generator power output. As the above control mechanism is not sufficient enough for maintaining the balance between the generation and load demand, a supplementary controller must be employed in the control structure [1]. In a supplementary controller, also known as load frequency controller (LFC), the area control error (ACE) is considered as input to LFC, where ACE results from a linear combination of tie-line power deviation and frequency deviation multiplied with frequency bias constant. The LFC operates in such a manner that it can force ACE to become zero, which in turn tries to make deviations in frequency and tie-line power to zero.

Many researches have proposed the LFC structure for IPS with considering various aspects of modern IPS and [2, 3] have presented an exhaustive literature review. Further, Ali and Abd-Elazim have proposed the PID control structure for LFC considering the approximate value for the generation rate constraint (GRC) nonlinearity with PID controller optimized bacterial foraging optimization but for actual practical values of GRC it fails [4]. It is evident from these literature reviews that the performances of the conventional PID controllers may be deplorable when the system parameter variation or a shift in the system's operating point tend to alter often. The control performance of these controllers degrades because of their synthesis dependency on the plant model. Hence integer-order controller sometime fails to provide satisfactory dynamic performance under system parametric uncertainty and continuous load disturbances [5, 6]. Sondhi and Hote have designed a fractional order controller for LFC of a single area power system. The simulation results obtained show that the FOPID has reduced steady-state error, robust toward plant gain variations, and excellent disturbance rejection in comparison with integer order controller [7]. The internal model control (IMC) has been explored to design the non-integer order controller for two-area interconnected power systems for load frequency control in [8]. Simulation results claim that IMC-based fractional-order controller can have improved performance in terms of disturbance rejection.

In contrast to above literature review, we can infer that the present highly complex IPS requires a fractional calculus-based robust control, which can efficiently handle the system parametric disturbances. Further, in realistic IPS to effectively solve the LFC problem the inherent nonlinearities, i.e., governor deadband (GDB) present in speed governing mechanism and GRC found in turbine of IPS needs to incorporate for have better power quality in terms of lesser frequency deviations and reduced tie-line power deviations.

To address the abovementioned issues, an effort has been made to investigate the LFC performance of the considered TANRT interconnected power system in this work. The simultaneous presence of GRC and GDB in the power system is addressed by using an optimal fractional order PID controller optimized by a newly introduced, i.e., Salp Swarm algorithm (SSA) a bio-inspired optimization algorithm proposed by Mirjalili et al. [9, 10]. In this work, firstly, the FOPID controller is optimized with SSA and compared with the well-established genetic algorithm (GA) under nominal

system parameters. To validate the proposed controller, individual and simultaneous variations of turbine time constant, governor time constant, and synchronizing power coefficient for investigated system have been considered. The dynamic performance has been then compared with GA. Further to evaluate the efficacy of fractional order control structure has been compared with integer order controller optimized with GA and SSA. The optimization problem is formulated as a design problem by evaluating integral multiplied by time absolute error (ITAE) criterion.

Further, the outline of present article is as follows; Sect. 2 deals with brief mathematical description LFC issue for a TANRT in consideration GRC and GDB for IPS. In Sect. 3, brief description of fractional-order controller, utilized optimization algorithm and selection criterion of objective function for LFC problem have been presented. In Sect. 4, the simulation results and discussion have presented. At last, concluding remarks are made in Sect. 5.

2 Dynamic Model of the Investigated IPS

The deviations in load occurring in IPS are typically exposed to very small in comparison with their rated capacity. Hence, it is usually characterized by the linear model. Interconnecting individual control areas can constitute a multi-area power system through the tie-line. A realistic version of interconnected power system with GRC and GDB nonlinearities has been included and shown in Fig. 1. The nominal system parameters of system under investigation have been adopted from [4, 11] and provided in Appendix 1.

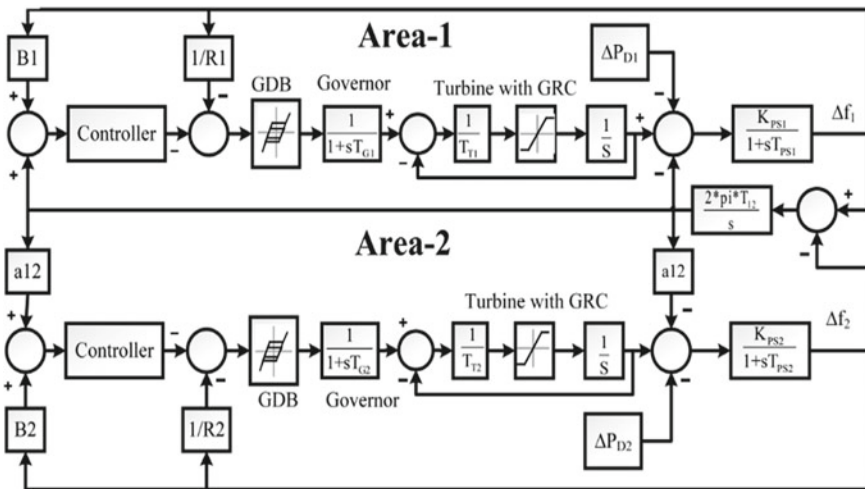


Fig. 1 Block diagram of the investigated IPS, i.e., Thermal-thermal non-reheated incorporating GDB and GRC

Further in subsequent section, the control structure (fractional order) for the LFC and its optimal parameters via different optimization techniques has been presented to achieve desired control objectives.

3 Load Frequency Controller Structure

Fractional calculus is a branch of mathematical calculus, which is used to change the integral and differential operator from integer-order to fractional order. Hence fractional-order calculus provides the liberty to the user for implementation of the non-integer or differential order operator, which can have a superior level of performance in terms of robustness when compared with integer order. In the literature, several approximation methods are proposed, ‘Carlson approximation’, ‘Matsuda approximation’, ‘Crone or Oustaloup approximation’, are few simplified continuous-time domain approximations techniques, etc. In the present work, ‘Oustaloup approximation’ is used to approximate the non-integer order calculus, because of its very good fitting to non-integer order elements. This approach fits the approximation utilizing a higher-order filter having an order of ‘ $2N + 1$ ’ inside a specified frequency bound (ω_L, ω_H).

Approximation of s^λ can be written as,

$$S_\lambda \approx G' \prod_{k=-N}^{k=N} \frac{s + \omega_{zk}}{s + \omega_{pk}} \tag{1}$$

where G' is gain, ω_{zk} and ω_{pk} , characterizes as the zeros and poles of analog filter respectively and can be represented as

$$\omega_{zk} = \omega_L \left(\frac{\omega_H}{\omega_L} \right)^{\frac{k+N+0.5(1-\lambda)}{2N+1}}, \omega_{pk} = \omega_L \left(\frac{\omega_H}{\omega_L} \right)^{\frac{k+N+0.5(1+\lambda)}{2N+1}} \text{ and } G' = (\omega_H)^\lambda \tag{2}$$

The effectiveness of fractional-order approximation of this technique primarily depends upon the value of ‘ N ’. With smaller values of ‘ N ’, it is easy to implement fractional-order operator approximation in hardware but deteriorates the approximation performance by producing the ripple in magnitude as well in phase response. The present article uses ‘ N ’ as 5, and frequency bound of [ω_L, ω_H] is considered as [$10^{-3}, 10^{-3}$], which is suitable in the LFC application [8, 10]. The fractional-order PID controller can represent ($PI^\lambda D^\mu$) as

$$Gc(s) = K_P + K_I S^{-\lambda} + K_D S^\mu \tag{3}$$

To obtain the desired system performance, the parameters of controllers must be adjusted. Soft computing techniques have been used extensively for tuning the LFC parameter to achieve superior dynamic system performances [12]. Some of these

algorithms are grasshopper optimization algorithm, water cycle optimization, bat algorithm, dolphin echolocation-based optimization algorithm, whale optimization algorithm, firefly optimization algorithm, harmony search, salp swarm optimization algorithm, etc. Out of these algorithms, SSA has emerged as a robust optimization algorithm with only one tunable factor, rapid convergence to optimal solutions with the tendency to avoid local optimum points. Due to these significant qualities present in SSA to solve optimization problems, this work has adopted SSA to design the fractional-order controller as LFC for interconnected power systems. Further, a brief detail about SSA is presented in the next subsection for better readability of the work presented in this article. SSA is a population-based heuristic optimization technique, which mathematically replicates the navigating and foraging performance of salps exists in subterranean seas, proposed by [9]. An effort has been made in [9] for the same where mathematical foundations of an efficient optimization algorithm have been laid. For the conciseness of the present article, the detailed description has not been portrayed and can be found in [13].

3.1 Load Frequency Controller Design

LFC in IPS are employed to perform the following objectives: (i) restore the frequency deviation to its nominal value, i.e., zero; and (ii) maintain the tie-line power transfer to its schedule value. Hence objective function should be designed in such a way so that it can address the above issues.

The integral time-weighted absolute error (ITAE) supported controller parameter optimization results in faster settling in contrast with other methods. LFC analysis performed in [11] shows that ITAE as a superior performance criterion. Therefore, in the current work, ITAE is used as an objective function for designing the PID/FOPID controller. Equation (4) represents the objective function for controller parameter optimization purpose.

$$J = \text{Objective Function(ITAE)} = \int_0^{t_{sim}} (|\Delta f_1| + |\Delta f_2| + |\Delta P_{tie12}|) t dt \quad (4)$$

where, $|\Delta f_1|$, $|\Delta f_2|$, $|\Delta P_{tie12}|$ and ' t_{sim} ' are the absolute values of deviation of frequency in control area1, control area2, tie-line power deviation, and simulation time respectively. Figure 2a and b shows the convergence for the PID and FOPID controller achieved by GA and SSA for the design of LFC for a two-area IPS.

It may be noted that the convergence of SSA is more rapid in comparison with GA for integer as well as fractional order PID controller. It also provided a better (lesser) value of the considered objective function (ITAE), i.e., a mere value of 3.928 and 3.203 in with SSA for PID and FOPID controller, respectively. Clearly, with reference to Fig. 3, the SSA-tuned FOPID (SSAFOPID) controller was a true winner under nominal conditions in comparison with GA-tuned FOPID. The optimized

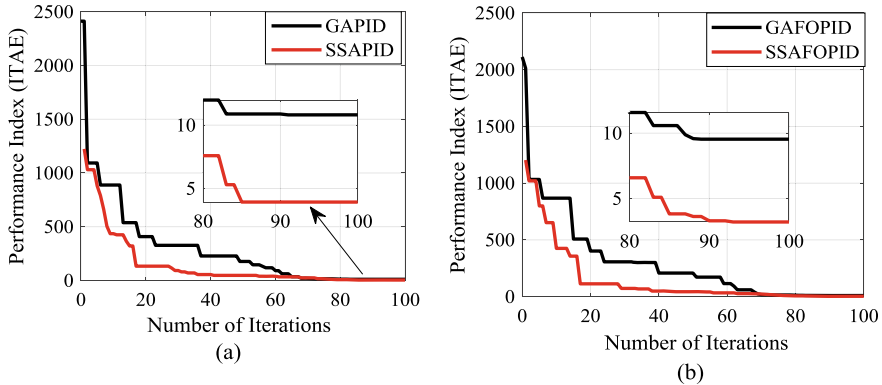
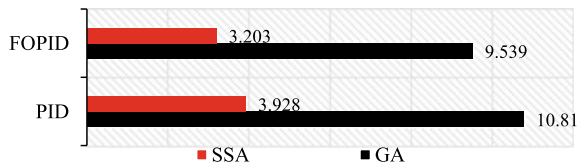


Fig. 2 Comparative convergence curves for GA and SSA (a) PID controller (b) FOPID controller

Fig. 3 Values of ‘*J*’ for different controllers under nominal system parameters for realistic two-area interconnected power system



controller parameters of all the controllers under consideration are listed in Table 1. The parameters of each controller employed in individual areas are considered as same to due identical structure of both control areas.

Further, the frequency and tie-line power deviation suppression in both the areas for optimized controllers, under nominal conditions, are shown in Figs. 4 and 5 with PID and FOPID controllers. Figure 4 shows the ‘ Δf ’ in area-1, area-2 and ‘ ΔP_{tie} ’ with PID controller structure.

As can be seen from Fig. 6, the ‘ Δf ’ in area-1, area-2 and ‘ ΔP_{tie} ’, shows lesser deviations with FOPID controller as compared PID, which establishes the supremacy of SSA optimized FOPID controller. Hence for the conciseness of the article further we have shown the FOPID controller results only, while other are presented in tabulated form in subsequent section.

Table 1 Realistic two-area interconnected power systems optimized controller with GRC and GDB

Parameter	PID		FOPID	
	GA	SSA	GA	SSA
K _p	10.0	8.9178	-2.2113	12.161
K _i	0.556	0.0374	6.971	0.069
K _d	13.63	14.657	13.399	14.75
λ	-	-	0.0016	0.951
μ	-	-	1.0682	1.1912

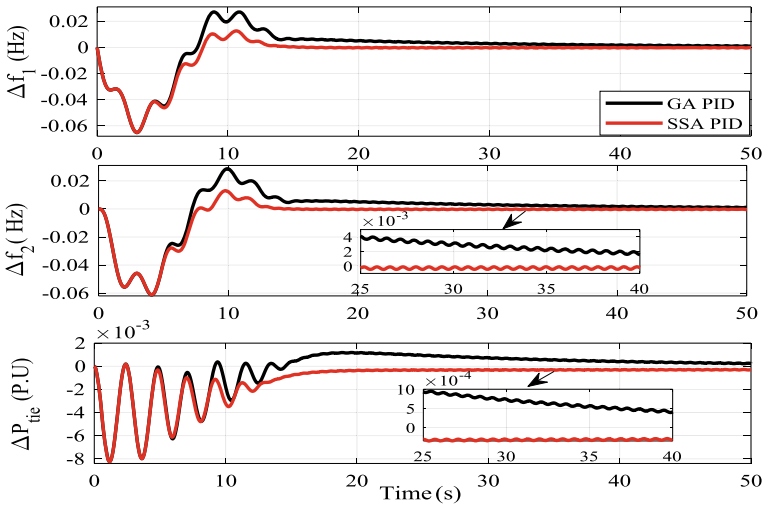


Fig. 4 Variation in Δf_1 , Δf_2 and ΔP_{tie} , under step load perturbation (SLP) of 0.01 pu in area1 with nominal parameters for PID controllers

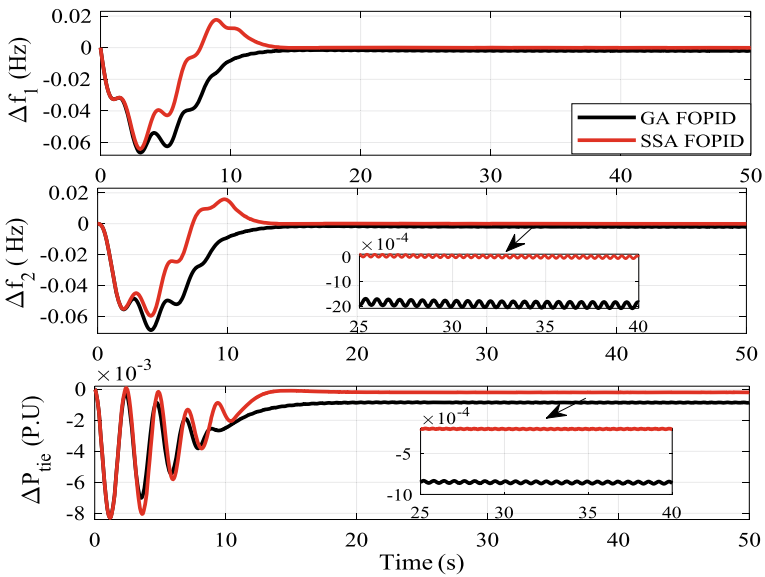


Fig. 5 Variation in Δf_1 , Δf_2 and ΔP_{tie} , under SLP of 0.01 pu in area1 with nominal parameters for FOPID controllers

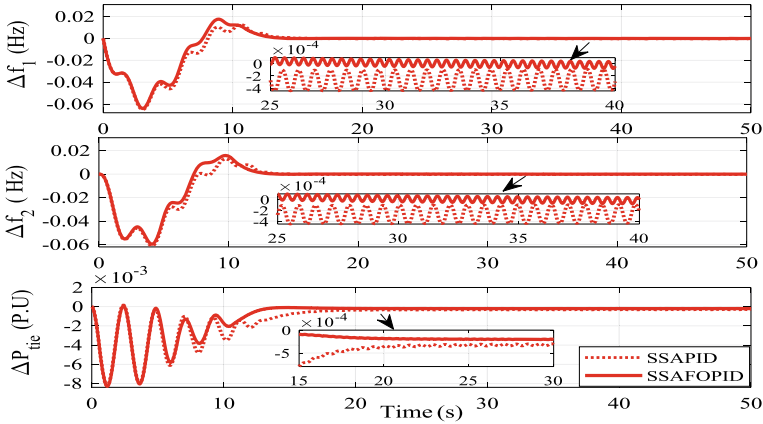


Fig. 6 Comparative representation of PID and FOPID controller for variation in Δf_1 , Δf_2 and ΔP_{tie} , under SLP of 0.01 pu in areal optimized with SSA

4 Simulation Results and Associated Discussion

This work presents a study on the use of optimally tuned fractional-order controllers for the effective design of LFC for two-area power system. As mentioned earlier, the presence of nonlinearities in the system can significantly affect the control performance. To address this issue, various simulation studies have been carried out and are presented in detail in this section. A detailed analysis of the designed optimal fractional-order controller for parametric variation has been carried out.

4.1 LFC Performance Analysis with Uncertain Time Constant System Parameters

The power system components such as governor, generator, turbine, tie-line components are used for an extended period and may withstand a deviation from their nominal values. If the controller design is not robust enough, then the control operation may become unstable which can, in turn, cause significant damage to the interconnected power system and thereby higher losses.

An extreme case has been carried out, in which all the time constant parameters, viz. T_G , T_T , and T_{12} are deviated simultaneously from their nominal values by $\pm 50\%$. Figure 8 shows the variation in ‘ Δf ’ of area-1, area-2 and ‘ ΔP_{tie} ’ for a simultaneous -50% in T_G , T_T , and T_{12} under step load perturbation (SLP) of 0.01 pu in control area-1. The ITAE achieved by the SSAFOPID controller is mere 3.144, while the ITAE value for GAFOPID controller 8.782, respectively, which are way higher than SSAFOPID controller.

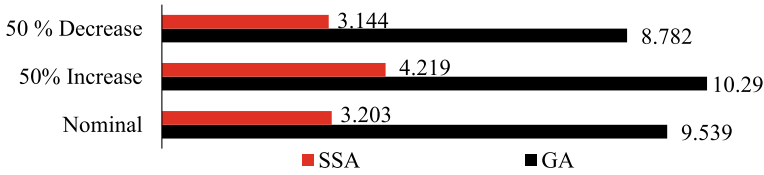


Fig. 7 Comparative chart of ‘J’ for FOPID controllers, under different parametric variation (T_G , T_T and T_{12})

Further, Fig. 9 shows the variation in Δf_1 , Δf_2 and tie-line power exchange for a simultaneous +50% all three time constants. It may be noted that due to increment in time constant parameters, the oscillations have been significantly increased in the transient period of response. The SSA tuned FOPID controller performance is superior in comparison to GAFOPID. This claim is also supported by the quantitative comparison, as shown in Fig. 7, which indicates that the SSAFOPID controller is providing the least ‘J’ at a value of 4.219 in contrast of 10.29 obtained by GAFOPID controller. The quantitative comparison for all considered cases is presented in Table 2, where it may be noted that SSAFOPID controller is providing best performance in all the cases.

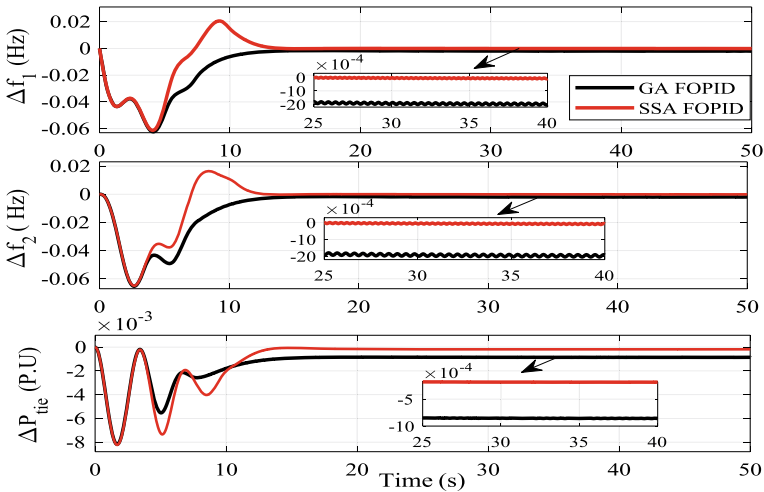


Fig. 8 The ‘ Δf ’ in area-1, area-2 and ‘ ΔP_{tie} ’ under –50% deviations in T_T , T_G , and T_{12}

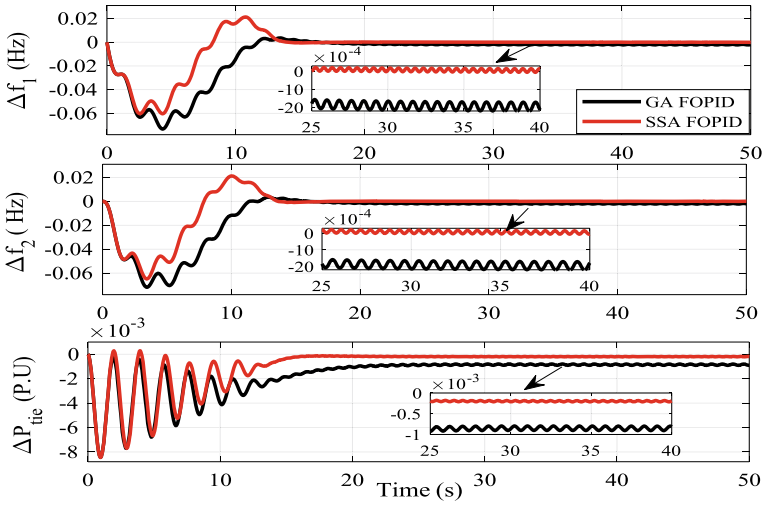


Fig. 9 The ‘ Δf ’ in area-1, area-2 and ‘ ΔP_{tie} ’ under +50% deviations in T_T , T_G , and T_{12}

Table 2 Values of ‘ J ’ for different controllers under system parametric variation

Robustness analysis type	PID		FOPID	
	GA	SSA	GA	SSA
Nominal	10.81	3.928	9.539	3.203
+50% T_G	10.84	4.469	9.367	3.423
-50% T_G	10.67	3.594	9.531	3.100
+50% T_{12}	10.89	4.558	10.17	3.885
-50% T_{12}	9.818	3.319	8.79	3.423
+50% T_T	10.82	3.929	9.56	3.206
-50% T_T	10.8	3.927	9.56	3.201
-50% T_T , T_G and T_{12}	9.645	3.21	8.782	3.144
+50% T_T , T_G and T_{12}	11.58	5.839	10.29	4.219

5 Conclusion

In the present work, an effort has been made to showcase the efficacy of a fractional-order, i.e., FOPID controller for enhanced dynamic performance of a nonlinear interconnected power system. The two-area interconnected power system considered in this work has nonlinearities, i.e., generation rate constraint (GRC) and governor dead-band (GDB), which hampers with the efficient control operation of the power system. Uncertainties in the power system component can cause additional complexity in controlling the power system, and the effectiveness of FOPID controllers in such scenarios has been addressed thoroughly in this work. FOPID controllers have simple

architecture yet if designed properly can provide enhanced control capabilities. The same is verified in this article through extensive simulation studies and comparative analyses. Salp swarm algorithm (SSA)-based FOPID design was critically investigated and compared with genetic algorithm (GA) optimized controllers based on a curated performance index. It was found that the SSA-tuned FOPID controller was capable of controlling the system more efficiently even under significant parametric variation.

Appendix 1

Nominal parameters of IPS under investigation are given as follows [4, 10, 11];

$P_{R1} = P_{R2} = 2000$ MW (rating), $PL_1 = PL_2 = 1000$ MW (nominal loading), $f = 60$ Hz, $T_{PS1} = T_{PS2} = 20$ s; $T_{T1} = T_{T2} = 0.3$ s; $2 * \pi * T_{12} = 0.545$ pu; $T_{G1} = T_{G2} = 0.08$ s; $K_{PS1} = K_{PS2} = 120$ Hz/puMW; $a_{12} = -1$; $R_1 = R_2 = 2.4$ Hz/puMW; $B_1 = B_2 = 0.425$ puMW/Hz, GDB = 0.06 pu and GRC = 10% puMW/min.

References

1. Kundur P (2009) Power System Stability and Control. Tata McGraw Hill, New Delhi
2. Ibraheem A, Kumar P, Kothari DP (2005) Recent philosophies of automatic generation control strategies in power systems. *IEEE Trans Power Syst* 20(1):346–357. <https://doi.org/10.1109/TPWRS.2004.840438>
3. Pandey SK, Mohanty SR, Kishor N (2013) A literature survey on load-frequency control for conventional and distribution generation power systems. *Renew Sustain Energy Rev* 25:318–334. <https://doi.org/10.1016/j.rser.2013.04.029>
4. Ali ES, Abd-Elazim SM (2013) BFOA based design of PID controller for two area Load Frequency Control with nonlinearities. *Int J Electr Power Energy Syst* 51:224–231. <https://doi.org/10.1016/j.ijepes.2013.02.030>
5. Alomoush MI (2010) Load frequency control and automatic generation control using fractional-order controllers. *Electr Eng* 91(6):357–368. <https://doi.org/10.1007/s00202-009-0145-7>
6. Debbarma S, Saikia LC, Sinha N (2014) Robust two-degree-of-freedom controller for automatic generation control of multi-area system. *Int J Electr Power Energy Syst* 63:878–886. <https://doi.org/10.1016/j.ijepes.2014.06.053>
7. Sondhi S, Hote YV (2014) Fractional order PID controller for load frequency control. *Energy Convers Manage* 85:343–353. <https://doi.org/10.1016/j.enconman.2014.05.091>
8. Saxena S (2019) Load frequency control strategy via fractional-order controller and reduced-order modeling. *Int J Electr Power Energy Syst* 104:603–614. <https://doi.org/10.1016/j.ijepes.2018.07.005>
9. 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. <https://doi.org/10.1016/j.advengsoft.2017.07.002>
10. A. K. Mishra and P. Mishra, Improved fractional order control of a nonlinear interconnected power system using salp swarm algorithm. In: 2019 IEEE 16th India Council International Conference (INDICON), 2019, pp. 1–4, doi: <https://doi.org/10.1109/INDICON47234.2019.9029023>.

11. Padhan S, Sahu RK, Panda S (2014) Application of firefly algorithm for load frequency control of multi-area interconnected power system. *Electr Power Components Syst* 42(13):1419–1430. <https://doi.org/10.1080/15325008.2014.933372>
12. Kalavani F, Zamani-Gargari M, Mohammadi-Ivatloo B, Rasouli M (2019) A contemporary review of the applications of nature-inspired algorithms for optimal design of automatic generation control for multi-area power systems. *Artif Intell Rev* 51(2):187–218. <https://doi.org/10.1007/s10462-017-9561-7>
13. Guha D, Roy PK, Banerjee S (2019) Maiden application of SSA-optimised CC-TID controller for load frequency control of power systems. *IET Gener Transm Distrib* 13(7):1110–1120. <https://doi.org/10.1049/iet-gtd.2018.6100>

Availability and Optimization of Continuous Manufacturing System Using Markov Modelling and Genetic Algorithm



Amit Jain , Vikrant Aggarwal, Rakesh Kumar, and Harmeet Singh Pabla

Abstract A mathematical model based on Markov death birth process of pulp manufacturing system is presented for improving the availability of this process. Here, the system passes through a series of preventive as well as corrective maintenances on its different transitions to pending-to-failed and failed states correspondingly. The probability recommendations at all stages of this system have numerous differential equations; those have to be solved by applying Laplace Transformations to compute state probabilities. Genetic algorithms have been developed to optimize the availability with varying input variables. The analysis boosted researcher's energy for the identification of key factors and there exists good scope to improvement in the system availability by controlling the contributing factors.

Keywords Markov modelling · Preventive maintenance · Genetic algorithm · Availability

1 Introduction

Reliability engineering emerges as a powerful concept in analysing the capabilities of the process industry. With increase in competition and much-defined schedules, every industry has to come up with the difficulties when the failure in any machinery

A. Jain (✉)

Department of Computer Science Engineering, GNDEC, Ludhiana, Punjab, India
e-mail: amitjain_17@live.com

V. Aggarwal

Production, Planning and Control Manager, MSME Consulting, Ludhiana, Punjab, India

R. Kumar

Department of Mechanical Engineering, Ludhiana Group of Colleges, Ludhiana, Punjab, India

H. S. Pabla

Department of Mechanical Engineering, Chandigarh University, Gharuan, Punjab, India

occurs. The various machinery and plant setup may be of series, parallel or mixed arrangement [1] depending on the industry and process. Availability is one of the major parameters to evaluate the plant performance. Markov technique is one of the most widely and acceptable in performing reliability and availability analysis. Most of the industries have a number of stages starting from the raw material to the final product. As the number of stages grows, there is a need a high computational effort to find its availability. It is difficult to compute analytical expression for the availability, when time of failure and/or time of repair is not exponential [2, 3]. Many researchers discuss the availability characteristics of degraded repairable systems under various assumptions on failure and repair rate with two or three components [4]. The literature suggests the use of several approaches to analyse and optimize the behaviour of various systems. Markov approach has [5] been applied to chewing gum making plant. Barabady et al. [6] proposed an effective optimal maintenance strategy is required to achieve performance. The model has been designed using language equations and the availability has been found. A simulation model on the availability of carbon dioxide cooling system in a fertilizer plant has been discussed by [7]. Genetic algorithm optimization has been applied [8] for crystallization processing part of any sugar plant. Deb [9] has discussed the optimization techniques and there use in analysing various complex engineering problems. Availability of combed yarn production system [10] has been studied, in consideration to preventive maintenance (PM) and corrective maintenance (CM). An effective PM planning guarantees the fewer breakdowns of the machines and effective production planning [11]. Udoh et al. [12] proposed a Complementary Optimal Age Maintenance (COAM) for repairable systems. The adaptive maintenance scheme for any bleaching process of pulp manufacturing plant has been [8] studied using steady-state behaviour. Kahle et al. [13] explained optimization modelling in the context of statistical maintenance of all repairable systems. Ram et al. [14] suggested various reliability computing measures for representing any industrial system with three sub-systems.

2 Process Description

The present study focusses on the availability and optimization of the Pulp Making System of Paper Manufacturing Plant under various conditions of its operation. The straw from different regions is procured and dried in a yard. This dry straw is seasoned. With the help of cutting machine, the straw is cut into small pieces of 10–12 mm length and 2–3 mm thin. This chopped straw is fed into two continuous rotary spherical digesters with fixed proportion of different cooking chemicals and water. The steam is injected into digester continuously from boiler to raise the temperature

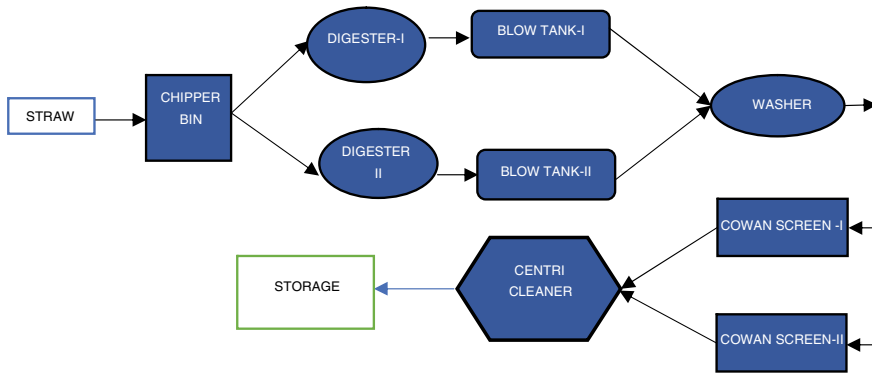


Fig. 1 Flow diagram of pulp making system

and pressure. Then this cooked pulp is sent to blow tanks. Then the pulp is passed through Knotters. After this, washing of cooked pulp is done to separate unbleached pulp from liquid in a washer. This washed pulp is passed through Cowan Screen after refining to remove fine shires. The screened pulp is cleaned through three-stage, the centric cleaner and the diluted pulp is thickened in vacuum thickener and stored in tower. The block diagram/model of Pulp Making System is presented in Fig. 1. The system comprises of five different components.

1. Component A_1 : It consists of one chipper bin. As chipper bin has a negligible failure rate, it will not affect the system’s availability.
2. Component A_2 : It consists of two units (A_{21} and A_{22}) arranged in parallel. Each unit comprises of digester and blow tank arranged in series. The component works with one unit in full capacity. Other unit acts as a hot standby.
3. Component A_3 : It consists of Washer. If it fails, the overall system failure takes place.
4. Component A_4 : It consists of two Cowan Screens (A_{41} and A_{42}) which are arranged parallel to each other. Failure of any one of the Cowan Screen decrements the overall processing capacity up to 50%. And, the complete failure of the system is considered only when both Cowan Screens remain in the fail state.
5. Component A_5 : It consists of Centric Cleaner without replacement. If it fails, the complete failure of the system takes place.

3 Mathematical Modelling

Notations:

Superscript “o”	: represents component is operative.
Superscript “r”	: represents component is under repair.
Superscript “g”	: represents component is good but not operative.
Superscript “qr”	: represents component is in queue for repair.
Superscript “qm”	: represents component is in queue for preventive maintenance.
$\lambda_1, \lambda_2, \lambda_3$: depicts failure rate of the component A_3, A_4 and A_5 respectively.
μ_1, μ_2, μ_3	: depicts repair rates of the component A_3, A_4 and A_5 respectively.
θ_1	: depicts transition rate leading the component A_2 to go for PM.

Assumptions:

1. At any particular point of time, the component maybe in any of the operating or reduced or may be in failed state.
2. Failure, preventive maintenance, repair and transition rates all are constant.
3. The plant/system always employs one maintenance team for the handling of both preventive maintenance and corrective maintenance.
4. At a time, preventive maintenance can be performed on one component.
5. The component is as good as new after preventive maintenance.
6. When any one component of the system is under repair, then the other component cannot be initiated for preventive maintenance.
7. Preventive maintenance is a regular exercise component A_2 , which make either unit to remain operational all time.
8. If component A_2 is in preventive maintenance and another component comes in a failed state, then the corrective maintenance of the failed component will be initiated at priority.
9. The repair is always done on the basis of some priority like, if number of failed items is more than some threshold value. For any repair task, A_3 and A_5 have first priority whereas component A_4 is on second priority.
10. While working at reduced capacity, the failure and repair rate of component A_4 remains unchanged. Figure 2 represents the State Transitions of Pulp Manufacturing System.

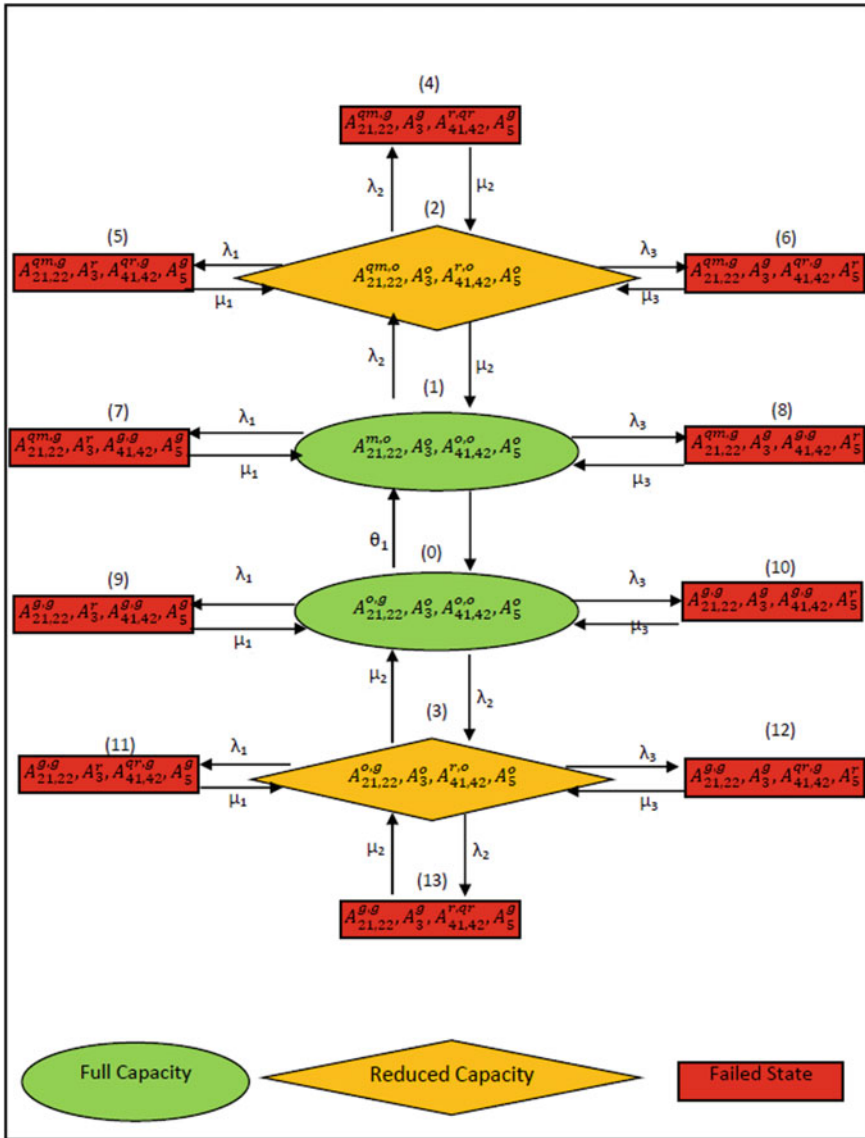


Fig. 2 State transition diagram of pulp making system

4 Formulation of the System

For the determination of reliability and availability of any pulp manufacturing system, the mathematical formulations are as follows:

$$\begin{aligned}
 P'_0(t) + B_0 P_0(t) &= \beta_1 P_1(t) + \mu_2 P_3(t) + \mu_3 P_9(t) + \mu_3 P_{10}(t) \\
 P'_1(t) + B_1 P_1(t) &= \theta_1 P_0(t) + \mu_1 P_7(t) + \mu_3 P_8(t) + \mu_2 P_2(t) \\
 P'_2(t) + B_2 P_2(t) &= \lambda_2 P_1(t) + \mu_1 P_5(t) + \mu_3 P_6(t) + \mu_2 P_4(t) \\
 P'_3(t) + B_2 P_3(t) &= \lambda_2 P_0(t) + \mu_1 P_{11}(t) + \mu_3 P_{12}(t) + \mu_2 P_{13}(t) \\
 P'_4(t) + \mu_2 P_4(t) + \lambda_2 P_2(t) & \\
 P'_5(t) + \mu_1 P_5(t) + \lambda_1 P_2(t) & \\
 P'_6(t) + \mu_3 P_6(t) + \lambda_3 P_2(t) & \\
 P'_7(t) + \mu_1 P_7(t) + \lambda_1 P_1(t) & \\
 P'_8(t) + \mu_3 P_8(t) + \lambda_3 P_1(t) & \\
 P'_9(t) + \mu_1 P_9(t) + \lambda_1 P_0(t) & \\
 P'_{10}(t) + \mu_3 P_{10}(t) + \lambda_3 P_0(t) & \\
 P'_{11}(t) + \mu_1 P_{11}(t) + \lambda_1 P_3(t) & \\
 P'_{12}(t) + \mu_3 P_{12}(t) + \lambda_3 P_3(t) & \\
 P'_{13}(t) + \mu_3 P_{13}(t) + \lambda_2 P_3(t) &
 \end{aligned}$$

where, $B_0 = \theta_1 + \lambda_1 + \lambda_2 + \lambda_3$, $B_1 = \beta_1 + \lambda_1 + \lambda_2 + \lambda_3$, $B_2 = \mu_2 + \lambda_1 + \lambda_2 + \lambda_3$.

Solving recursively, the above equations after applying Laplace transformations, the following state probabilities are obtained:

$$P_n(s) = M_n P_0(s), \text{ For } n = 1 \text{ to } 13 \tag{1}$$

where,

$$M_1 = \frac{\theta_1}{B_4}, M_2 = \frac{\lambda_4}{B_4} \cdot \frac{\theta_1}{B_5}, M_3 = \frac{\lambda_2}{(s + B_2) - \mu_1 M_9 - \mu_3 M_{10} - \mu_2 B_6},$$

$$M_4 = B_6 M_2, M_5 = M_9 M_2, M_6 = M_{10} M_2, M_7 = M_9 M_1, M_8 = M_{10} M_1,$$

$$M_9 = \frac{\lambda_1}{s + \mu_1}, M_{10} = \frac{\lambda_3}{s + \mu_3}, M_{11} = M_9 M_3, M_{12} = M_{10} M_3, M_{13} = B_6 M_3$$

Also,

$$B_4 = \left(s + B_1 - \mu_1 M_9 - \mu_3 M_{10} - \mu_2 \frac{\lambda_2}{B_5} \right), B_5 = (s + B_2 - \mu_1 M_9 - \mu_3 M_{10} - \mu_2 B_6),$$

$$B_6 = \frac{\lambda_2}{s + \mu_2}$$

Taking Laplace transform of Eq. (1)

$$P_0(s) = \frac{1}{A_3} \tag{2}$$

where, $B_3 = s + B_0 - \beta_1 M_1 - \mu_2 M_3 - \mu_1 M_9 - \mu_3 M_{10}$.

Full capacity availability function $A_{FC}(s)$ of the Pulp Making System is given as,

$$A_{FC1}(s) = P_0(s) \tag{3}$$

Reduced availability function $A_{RC}(s)$ of the Pulp Making System is given as,

$$A_{RC1}(s) = P_2(s) + P_3(s) \tag{4}$$

Inversion of $A_{FC1}(s)$ and $A_{RC1}(s)$ gives the availability function $A_{FC1}(t)$ and $A_{RC1}(t)$ respectively.

5 Steady-State Behaviour of the Plant/System

The planning and management activities for an industry are mostly connected to compute the availability/long life of the system [15]. To solve this purpose, there is a further need to compute steady-state probability of that system under investigation. The above-mentioned equations are helpful for computing steady-state probabilities. Table 1 represents the failure, repair, transition and PM rates of various components of the system. Considering the values we get:

$$A_{FC1} = 0.8335, A_{RC1} = 0.0613$$

The Pulp Making System is running at full capacity or reduced capacity. While running at reduced capacity, the availability reduces to 50%. Therefore, the total availability (A_{TC1}) of the system can be computed as:

$$A_{TC1} = A_{FC1} + 0.5 * (A_{RC1}) = 0.8335 + 0.5 * (0.0613) = 0.8641$$

Table 1 Failure, repair, transition and PM rates of the components of Pulp Making System

Component	Failure rate	Repair rate	Transition rate	PM rate
Digester and blow tank	–	–	0.05	0.04
Washer	0.008	0.15	–	–
Cowan screens	0.018	0.3	–	–
Centri cleaner	0.015	0.25	–	–

6 Applicability of Genetic Algorithm for Optimizing the Results

GA is one of the efficient and effective heuristic search-based optimization technique. This is also based on the principles of natural genetics and natural selection. Very complex engineering problems where optimization is needed to be solved with GA. Basic steps in GA are shown in Fig. 3 [9, 16–20].

The presentation enhancement of the mash making framework is profoundly impacted by the disappointment and fix boundaries of every framework. Hereditary algorithm is proposed to arrange the disappointment and fix boundaries of every framework for stable framework execution under ideal PM and broken PM. The quantity of factors is eight. For genetic algorithm enhancement, the chromosomes are to be coded in genuine structures. Not at all like, unsigned fixed point number coding boundaries are planned to a predefined span $[H_{min}, H_{max}]$, where H_{min} and H_{max} are the base and most extreme estimations of framework boundaries. The most extreme estimation of the accessibility work compares to ideal estimations of framework boundaries. These boundaries are upgraded by the accessibility level. The presentation of the mash making framework is assessed by utilizing the planned estimations of the unit boundaries as appeared in Table 2.

The effect of number of ages on the AOC under ideal PM Mutation likelihood = 0.1, Population size = 100, Crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is appeared in Fig. 4a and the Effect of number of populace on the AOC under ideal PM Mutation likelihood = 0.1, age size = 100, Crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is appeared in Fig. 4b.

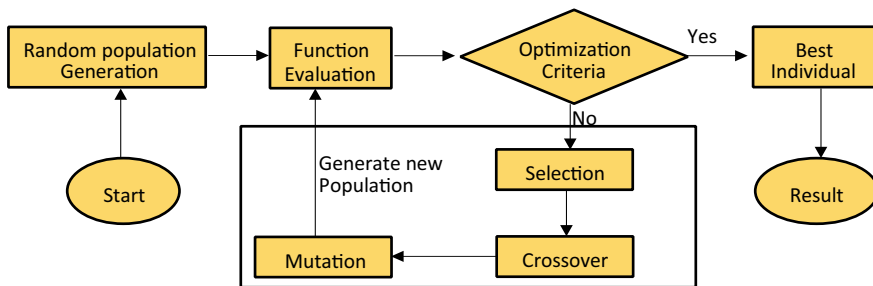


Fig. 3 Methodology of genetic algorithm

Table 2 Maximum and minimum value of parameters for genetic algorithm

Parameter	λ_1	λ_2	λ_3	μ_1	μ_2	μ_3
H_{min}	0.008	0.018	0.018	0.01	0.15	0.3
H_{max}	0.04	0.09	0.09	0.05	0.65	0.7

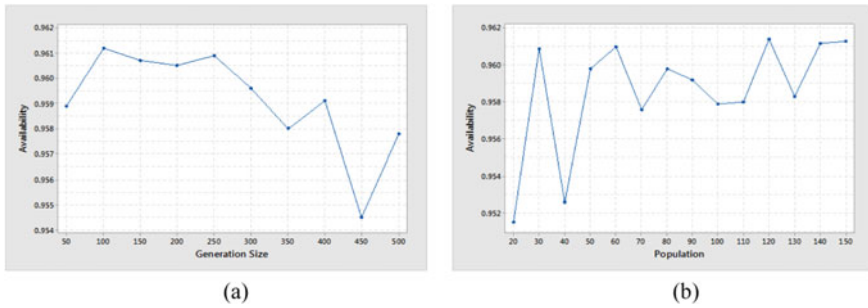


Fig. 4 AOC varies with **a** generations in GA optimization technique under ideal PM **b** populations in GA optimization technique under ideal PM

The impact of number of age on the AOC under broken PM Mutation likelihood = 0.1, Population size = 100, Crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is appeared in Fig. 5a. The effect of number of populace on the AOC under flawed PM mutation likelihood = 0.1, generation size = 100, crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is appeared in Fig. 5b

The effect of number of populace on the AFC under ideal PM, mutation likelihood = 0.1, population size = 100, crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is appeared in Fig. 6a. The effect of number of populace on the AFC under ideal PM taking mutation likelihood = 0.1, generation size = 100, crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is appeared in Fig. 6b.

The effect of number of age on AFC under broken PM taking mutation likelihood = 0.1, population size = 100, crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is indicated in Fig. 7a. The effect of number of populace size on AFC under defective PM taking mutation likelihood = 0.1, generation size = 100, crossover likelihood = 0.8, $\alpha = 0.25$, $\varphi = 0.2$ is appeared in Table 10 and graphical portrayal in Fig. 7b

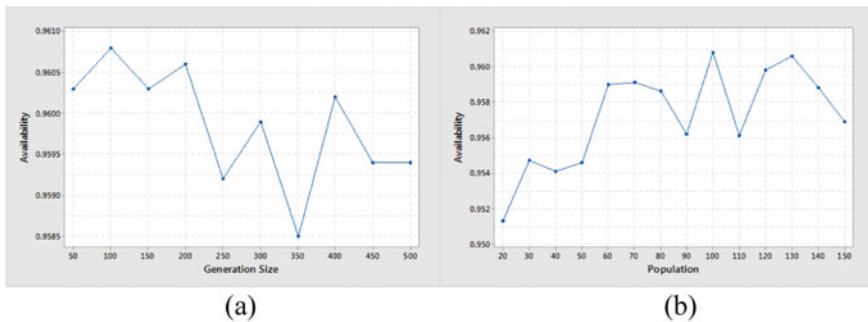


Fig. 5 AOC varies with **a** generations in GA optimization technique under faulty PM **b** populations in GA optimization technique under faulty PM

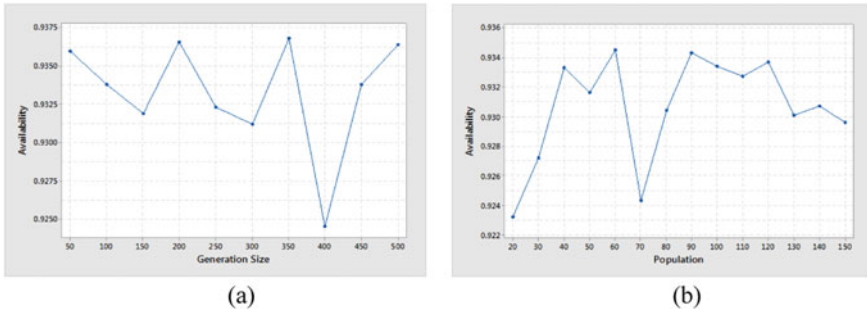


Fig. 6 AOC varies with **a** generations in GA optimization technique under faulty PM **b** population in GA optimization technique under ideal PM

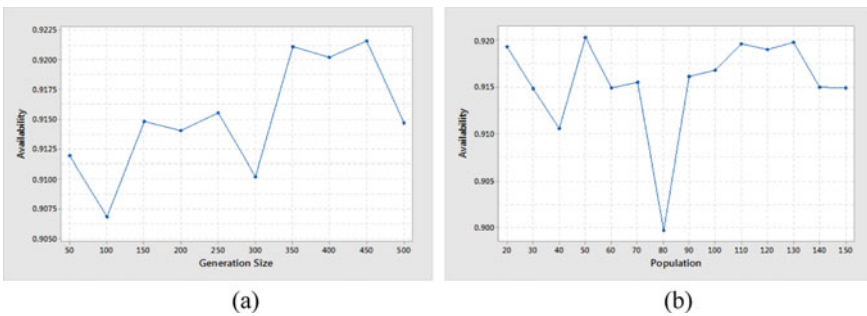


Fig. 7 A_{FC} varies with **a** generations in GA optimization technique under ideal PM **b** population in GA optimization technique under faulty PM

7 Results and Discussion with Performance Optimization

The most extreme present accessibility of the framework is 89.54% if there should be an occurrence of ideal PM and 89.48% in the event of broken PM. The point of streamlining is to upgrade the accessibility of the framework. The general accessibility of the framework can be expanded up to 96.14% with ideal PM and 96.08% in the event of broken PM. Tables 3 and 4 sums up the estimations of factors comparing to the accessibility.

8 Conclusion and Result Validation

The work present in the paper is unique in terms of application. Although genetic algorithm is used as a powerful tool in various system optimizations but it had never been applied to a paper manufacturing system with ideal and faulty PM. The researchers never compared proposed technique in terms of results in consideration to

Table 3 Performance optimization of AOC genetic algorithm

β	A _{OC} (%)	Variation in population	A _{OC} (%)	Variation in generation
		Failure rates and repair rates		Failure rates and repair rates
0	96.14	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.019, \lambda_6 = 0.01, \mu_3 = 0.75, \mu_4 = 0.699, \mu_5 = 0.697, \mu_6 = 0.65$	96.12	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.024, \lambda_6 = 0.01, \mu_3 = 0.75, \mu_4 = 0.698, \mu_5 = 0.696, \mu_6 = 0.65$
1	96.08	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.021, \lambda_6 = 0.01, \mu_3 = 0.749, \mu_4 = 0.693, \mu_5 = 0.689, \mu_6 = 0.65$	96.08	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.021, \lambda_6 = 0.01, \mu_3 = 0.749, \mu_4 = 0.693, \mu_5 = 0.689, \mu_6 = 0.65$

Table 4 Performance optimization of A_{FC} genetic algorithm

β	A _{FC} (%)	Variation in population	A _{FC} (%)	Variation in generation
		Failure rates and repair rates		Failure rates and repair rates
0	93.45%	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.019, \lambda_6 = 0.012, \mu_3 = 0.75, \mu_4 = 0.699, \mu_5 = 0.599, \mu_6 = 0.643$	93.68%	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.02, \lambda_6 = 0.01, \mu_3 = 0.749, \mu_4 = 0.7, \mu_5 = 0.666, \mu_6 = 0.649$
1	92.03%	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.023, \lambda_6 = 0.01, \mu_3 = 0.728, \mu_4 = 0.691, \mu_5 = 0.697, \mu_6 = 0.641$	92.16%	$\lambda_3 = 0.018, \lambda_4 = 0.018, \lambda_5 = 0.02, \lambda_6 = 0.01, \mu_3 = 0.75, \mu_4 = 0.698, \mu_5 = 0.69, \mu_6 = 0.65$

optimization. But authors can easily see that optimization enhances the availability of the system. The concept behind to find the availability considering various population and generation size is to iterate the system so as to find the best possible results.

References

1. Srinath LS (1994) Reliability Engineering, 3rd edn. East-West Press Pvt. Ltd., New Delhi
2. Crevecoeur GU (1993) A model for the integrity assessment of ageing repairable systems. IEEE Trans Reliab 42(1):148–155
3. Srinivasa MR, Naikan VNA (2011) A hybrid Markov system dynamics approach for availability, analysis of degraded systems. In: Proceedings of the 2011 international conference on industrial engineering and operations management. Kuala Lumpur, Malaysia, January 22–24
4. LeSanovskf A (1988) Multistate Markov models for systems with dependent units. IEEE Trans Reliab 37(5):505–511
5. Goyal A, Gupta P (2012) Performance evaluation of a multi-state repairable production system: a case study. Int J Performability Eng 8(3):330–338
6. Barabady J, Markeset T, Kumar U (2007) Improvement of production plant performance using production assurance programs. In: Swedish production symposium
7. Kumar S, Kumar D, Mehta NP (1999) Maintenance management for ammonia synthesis system in a urea fertilizer plant. Int J Manage Syst 15(3):211–214
8. Tewari PC, Khanduja R, Gupta M (2012) Performance enhancement for crystallization unit of a sugar plant using genetic algorithm technique. J Ind Eng Int
9. Deb K (1995) Optimization for engineering design: algorithms and examples. Prentice Hall of India, New Delhi, India

10. Garg S, Singh J, Singh DV (2010) Mathematical modelling and performance analysis of combed yarn production system. *Appl Math Model* 34(11):3300–3308
11. Yuting J, Xiaodong F, Chuan L, Zhiqi G (2014) Research on preventive maintenance strategy optimization based on reliability threshold. In: 2014 Prognostics and system health management conference (PHM-2014 Hunan). Zhangjiajie, pp 589–592
12. Udoh N, Effanga E, Onwunke C (2020) Complementary optimal age maintenance (COAM) policy for repairable systems. *Int J Reliab Saf* 14:1–13
13. Kahle W, Love CE (2003) Modelling the influence of maintenance actions. *Mathematical and statistical methods in reliability*. In: Lindqvist BH, Doksumeds KA Series on quality, reliability and engineering statistics. World Scientific Publishing, Singapore
14. Ram M, Manglik M (2016) Reliability measures analysis of an industrial system under standby modes and catastrophic failure. *Int J Oper Res Inf Syst* 7:36–56
15. Arora N, Kumar D (1997) Availability analysis of steam and power generation system in thermal power plant. *Microelectron Reliab* 37(5):795–799
16. Castro HF, Cavalca K (2003) Availability optimization with genetic algorithm. *Int J Qual Reliab Manage* 20(7):847–863
17. Chales C, Kondo A (2003) Availability allocation to repairable systems with genetic algorithms: a multi-objective formulation. *Reliab Eng Syst Saf* 82(3):319–330
18. Goldberg DE (2001) *Genetic algorithm in search, optimization and machine learning*. Pearson Edition, Asia
19. Khanduja R, Tewari PC, Chauhan RS, Kumar D (2010) Mathematical modeling and performance optimization for the paper making system of a paper plant. *Jordan J Mech Ind Eng* 4(4):487–494
20. Luo X, Qian Q, Fu YF (2020) Improved genetic algorithm for solving flexible job shop scheduling problem. *Procedia Comput Sci* 166:480–485

Indoor and Outdoor Localization Methods for Advanced Navigation Systems



P. Kanakaraja, Sarat K. Kotamraju, L. S. P. Sairam Nadipalli,
S. V. Aswin Kumer, and K. Ch. Sri Kavya

Abstract The indoor and outdoor localization plays an important role in the domain of the Internet of Things, Advanced IoT, and its Applications to enhance the performance and the accuracy of the system. It can be possible by using some wireless and MEMS sensor networks, which help to track and tracing the location of the persons and objects. In the olden days, it is very difficult to find the location of particular objects and persons. After introducing the GPS, the location finding becomes very easy, but in most of the situations, we cannot get the GPS co-ordinates, and we cannot depend on the GPS. In the absence of GPS, the hall-effect sensor, pressure sensor, Bluetooth beacons, accelerometer, magnetometer, and gyroscope are used to find the location with combined effects on both indoor and outdoor localization. This chapter also deals with LIDAR used in autonomous vehicle localization.

Keywords Navigation systems · Localization methods · Internet of Things (IoT) · Inertial sensors · Sampling rate · LIDAR · Absolute positioning system · Coordinate system · Autonomous vehicles

1 Introduction

IoT sensors such as RFID, IMU, Hall Sensors, Pressure Sensors these are the main sensors that we are going to work with localization [1] it appears strange to the

P. Kanakaraja · S. K. Kotamraju (✉) · L. S. P. S. Nadipalli · S. V. A. Kumer · K. Ch. Sri Kavya
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
e-mail: kksarat@gmail.com

P. Kanakaraja
e-mail: pamarthikanakaraja407@gmail.com

L. S. P. S. Nadipalli
e-mail: sai.nadipalli@gmail.com

S. V. A. Kumer
e-mail: svaswin@gmail.com

K. Ch. Sri Kavya
e-mail: kavyakorada@gmail.com

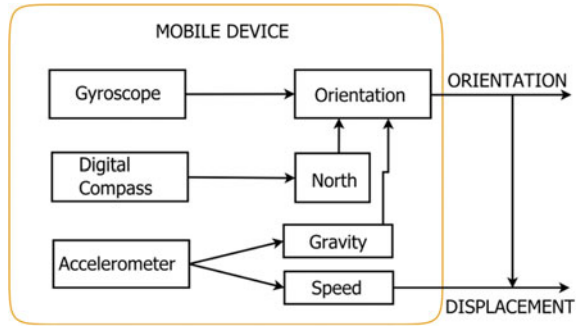
designer perhaps that now here GPS actually shown and that's the key with respect to localization [2, 3] with GPS Systems because GPS [3–5] has its own problems and it has its own advantages it may have to use an indirect way of measuring distance and from distance [6] try to arrive at localization [7] of these sensors. So, now if the designer look at RFID and look at what are these applications that will require RFID, to start putting some of applications in forward to the designer take a large warehouse, when the designer say warehouse it could be a warehouse where there are garments there can be jewelry or it could be a departmental store where we go and by groceries. So, all these essentials have lot of items and keeping inventory of these items is a tedious task; today, there is a what is known as a barcode; each of these items have a barcode, and they are barcoded and placed in a particular rack and each time a customer picks an item, the item is put into a basket and brought in front of the point of sale system; there the barcode is read; the amount is deducted, and there is a billing which is all familiar with; this is fine as far as barcodes are concerned; supposing the designer want to do inventory, this barcode is not going to help the designer at all. It is only for billing; it is only meant to know the item and so on so forth. It cannot tell the designer where and how many of these numbers are available, which means anything with respect to inventorying them are not going to work. So far that what people have done is a dumb way of doing barcode is good, but it is dumb; let us try and replace it with something much more useful something more interactive. So, they started replacing the barcodes with RFID tags. These are all passive, the costs are not down to earth, but costs are fall the time the designer can buy a RFID inlay for less than 15 rupees or the inventory business can be solved, the designer can do inventory and the designer can get to know the numbers.

2 Inertial Sensors (IMU)

Walk with the mobile phone in hand, allowing the user to interact with the screen of their terminal or allowing the mobile terminal to work without the user being aware of the device, it can be used to receive a large amount of information from the environment in a very graphic and intuitive way. The latest generation phones are already considered in many cases as IMUs, due to Since mobile inertial sensors are already incorporated, therefore, it is logical to think of a mobile terminal as a census device to obtain movement data, since In addition to being a very widespread technology in society, it integrates as part of its own hardware the inertial sensors necessary to analyze movement of people and power exploit this information as a complement to more powerful positioning systems. The mobile terminal will not constitute the positioning system itself but will be a support tool that will provide the true absolute positioning system, giving us additional information about the orientation and acceleration of the individual.

As shown in Fig. 1, the latest generation mobile terminals already integrate in many cases these sensors can be considered in themselves as IMUs. This fact greatly facilitates the possibility of developing localization [8, 9] solutions for precision in

Fig. 1 Mobile device like IMUs



indoor settings, technically feasible and acceptable from the point of view economic and functional since it will not be necessary to provide the user with additional hardware. Considering that inertial systems are not advisable from the point of view to be used as absolute positioning systems since they present an important limitation that make them unsuitable for this use which is drift or error cumulative, due to which the errors obtained in the data processing of a Position, accumulate and affect the calculation of the next position, making the system highly imprecise over a period of time. This work leads to consider IMU inertial systems as autonomous systems that complete the information from other more powerful absolute positioning systems, contributing to this dynamic and complementary information on the movement made.

For the acquisition and analysis of sensory data (accelerometer, magnetometer, gyroscope), we are going to use the Samsung phone model GT-I8160, Android operating system version 4.3.6 or Apple I Phone; the choice of this device was made because of the built-in sensors it has as shown in Fig. 2.

The sensors are error tested to consider the need for calibration before using such sensors. To know the accuracy and behavior of inertial sensors, we are going to carry out different types of tests repetitively; we will carry out these tests using the eclipse software, and the samples will be taken at approximately 10, 15 and 1.2 ms for each of the inertial sensors that we are going to study The response from the sensor module depends on the movement of the device, which we will call the device’s coordinate system as seen in Fig. 3, which is in relation to the orientation of the screen and the axes which do not vary when the motion of the device has other dimensions.

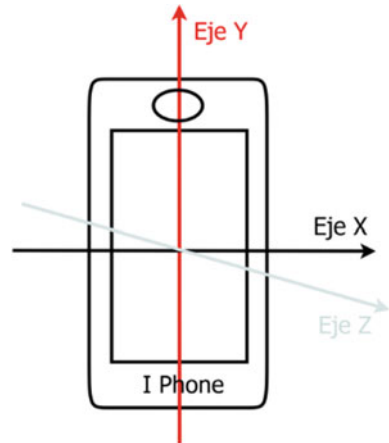
The coordinate system reference is defined by:

1. X-axis, by the cross-product between Y, Z, and this is tangential to the earth at the location current device and points west.
2. Y-axis is tangential to the ground at the current location of the device and points to the North pole.
3. Z-axis, it is perpendicular to the ground and points toward the center of the earth.

Fig. 2 I Phone IMU sensor



Fig. 3 Coordinate system of the mobile device



3 Indoor Multimodal Locations

After taking into account the coordinate system that we are going to use in the mobile device that will be our source of information for inertial sensors; next, we are going to carry out two types of tests: In the first one, the mobile phone is motionless on a flat surface and on the second holding the mobile in hand while walking on different

Fig. 4 IMU of the mobile device

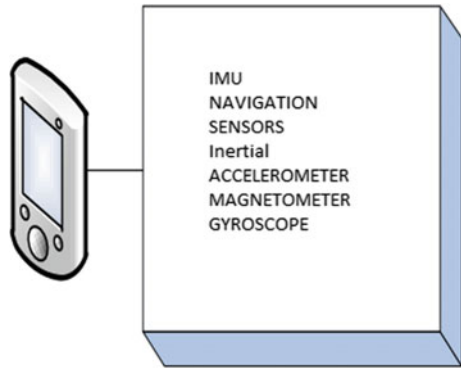


Table 1 Sampling frequency of inertial sensors

Sensor module	Mean value of sampling rate	Upper threshold	Lower threshold	Square root of variance
Gyroscope	1.484824	2.982	0.578	0.2823342
Magnetometer	15.92276	18.9	13.96	0.3501344
Accelerometer	19.97665	22.802	19.292	0.3702467

directions. In this part of the project, we are going to give more preference to the phone always pointing north. Ideally, the constant sampling rate with high values is chosen to reduce the loss of data because of the unselect able sampling frequency. Otherwise, the inertial sensors can be observed directly as shown in Fig. 4.

The device having limitation of sampling rate based on the condition for sampling so the observation from the sensors connected with the device based on their own sampling frequency. Then, the measurement from the sensor by the device is stable. After that the results produced by these two tests having more correlation which are listed in Table 1.

The tests carried out with each type of inertial sensors are detailed below. The information was obtained from the Smartphone mobile device that has sensors as shown in Fig. 5.

4 Stationary Device on a Flat Surface

This test will be carried out on a flat part, leaving the device unmoving [10, 11] for a period 20 s, so that the only force that interacts is gravity. For each test, we will have the outputs of each of the sensors the accelerometer, gyroscope and magnetometer.

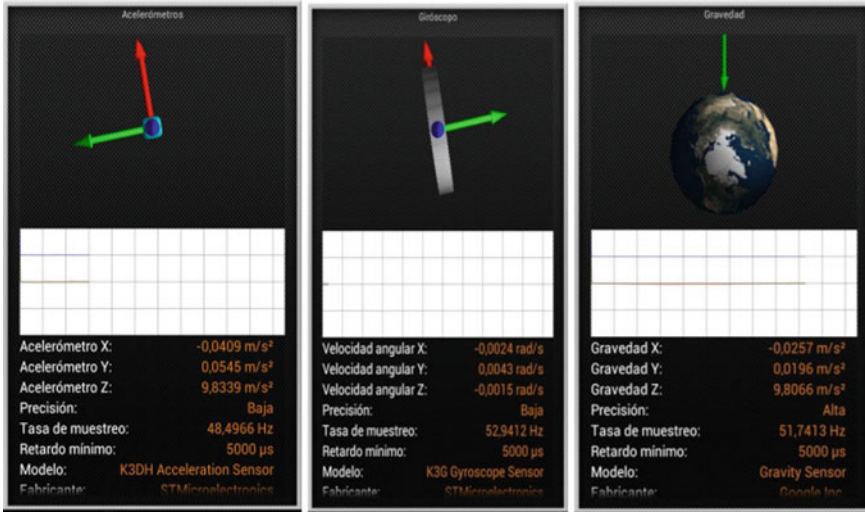


Fig. 5 Values obtained from IMUs on the mobile device

4.1 Accelerometer

The force applied to the sensor can be measured by applying acceleration in the device, which can be determined by the accelerometer with the help of the relationship following:

$$A_d = - \sum F_s / \text{mass} \tag{1}$$

Also, the relation between the measurement of acceleration and the force of the gravity is also determined by the following Eq. (2).

$$A_d = -g - \sum \frac{F}{\text{mass}} \tag{2}$$

where.

A_d = Applied acceleration.

g = gravitational force.

F = Device force.

m = Device mass.

\sum = mean of axes x , y , and z .

If the device is the flat position, then no acceleration is on the device; then, the magnitude read by the accelerometer is equal to 9.81 m/s^2 . Then, if the device is falling, the magnitude read by the accelerometer is equal to 0 m/s^2 because of the acceleration 9.81 m/s^2 . The exact acceleration of particular device can be measured with the help of the accelerometer data of the gravitational force as shown in Fig. 6.

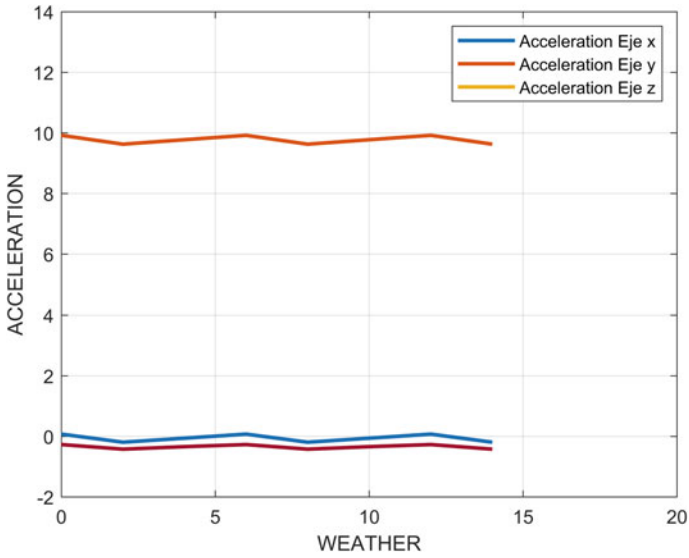


Fig. 6 Accelerometer output on each axis (x, y and z)

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} \tag{3}$$

In the Table 2, we have the statistical data of the outputs of each axis of the accelerometer of the measurements that were carried out.

Based on the precision level of the accelerometer, the measured acceleration when the device is in stable position is approximately equal to 9.66 m/s² during testing, and instead of 9.81 m/s², the variance of the total acceleration is 0.0064 m/s² which is equal to 1/100th part of total acceleration, which generates error periodically.

Table 2 Statistical values of the accelerometer output

AXES (m/s ²)	Mean value of sampling rate	Upper threshold	Lower threshold	Square root of variance
x	-0.4371	-0.3480	-0.521	0.0453
y	-0.0491	0.0960	-0.291	0.0644
z	9.7766	9.9210	9.6292	0.0615
Total acceleration	9.6556	9.9240	9.0150	0.08288

4.2 Gyroscope

The speed of rotation around the devices in all the three axes can be measured by the gyroscope sensor with the units (rad/s). If the device is in stable position, then the output of the gyroscope is 0 (rad/s) as shown in Table 3.

In the results, it can be seen that there is a displacement called—bias|| in all the three axes, but especially in Y-axis (positive) and the Z-axis (negative). The angular velocity can be obtained from the gyroscope output which integrates with time which is used to evaluate angle α .

$$\alpha_n = \sum_{i=0}^n (W_i * \nabla t) \tag{4}$$

The different tests that were performed rotating the device 360° on the z-axis in different time intervals showed us an output similar to the one shown in the table previous. The value of the gyroscope is approximately equal to 0.3 radians after a complete rotation in clockwise direction, and similarly, it is -0.3 radians in anticlockwise direction. The value may vary based on the speed of the rotation, so the error correction equation as following is used to avoid this problem.

$$W_i = W_i - W_i(1 - \epsilon_{x,y,z}) \tag{5}$$

where

$$\epsilon = \text{Estimation error}$$

The relation between estimated error and the drift error can be related as following Eq. (6).

$$\epsilon = \frac{0.33}{2\pi} \tag{6}$$

$$\epsilon = 0.053(5.3\%)$$

Table 3 Statistical values of the gyroscope output

AXES (rad/s)	Mean value of sampling rate	Upper threshold	Lower threshold	Square root of variance
x	-0.00013	0.05230	-0.0810	0.009586
y	-0.005925	0.19538	-0.1948	0.007688
z	-0.008653	0.03965	-0.0526	0.008219

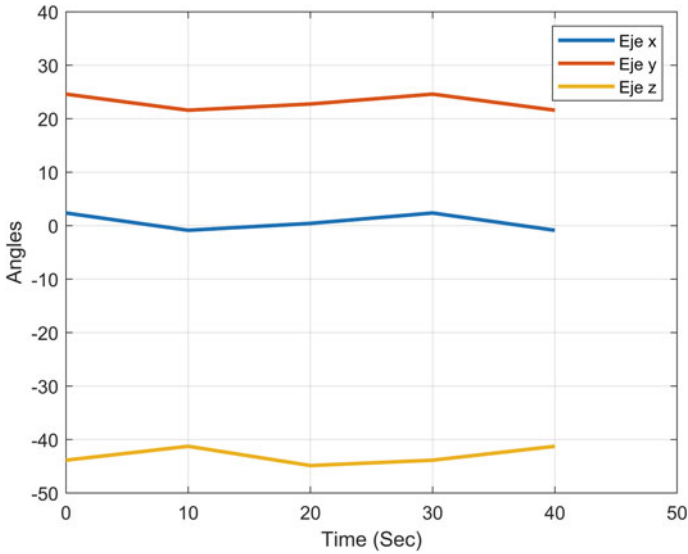


Fig. 7 Magnetometer output at rest

This correction significantly reduces the gyroscope drift that accumulates with the time in the corners. Even though the error value is negligible, it creates a great difference in angle at the output of gyroscope periodically. To compensate this error, the magnetometer are also considered by azimuth value.

4.3 Magnetometer

The ambient magnetic field can be measured by using magnetometer with the unit micro-tesla (μT) in all three coordinate axes. The output of the magnetometer together with the values of the accelerometer is merging to get the orientation. The magnetometer sensor is used along with the accelerometer and gyroscope. The data from the magnetometer supports to generate exact results result of accelerometer and gyroscope for finding the angles as shown in Fig. 7.

Based on the sensitivity, the magnetometer produces the deviation more than 0.4 micro tesla (μT) on each axis as shown in Table 4.

5 Conclusion

This chapter is not so much about GPS and all that. But it is about how IoT applications built with IoT sensors, which can assist aid in solving a very large problem

Table 4 Magnetometer statistical values

AXES (rad/s)	Mean value of sampling rate	Upper threshold	Lower threshold	Square root of variance
x	0.4285	2.368	-0.8647	0.488565
y	22.7521	24.598	21.5896	0.435798
z	-44.8589	43.8569	-41.2567	0.475686

of localization [12] of sensor nodes without GPS, or the absence of GPS [3–5], in the chapter RFID localization [13, 14] pressure-based Elevation algorithm using the Barometer Sensor is proposed, and the LIDAR is used for state estimation and localization [13, 14] for self-driving cars is in the conclusion.

6 Summary

In this chapter, we are summarizing the concepts of both indoor and outdoor localization. The indoor localization can be possible by using IMUs, and the outdoor localization [13, 14] can be possible by using gyroscope, accelerometer, hall-effect sensor, barometer and LIDAR. Among these RFID, IMUs are common in all the localization [13, 14] methods, so if these methods are used to implement the localization, then we can easily implement hybrid localization [13, 14] of both indoor and outdoor in the absence of GPS with high performance and accuracy.

References

1. Jyothirmai V, Ramesh NVK, Navyateja Y, Vijay R (2018) Localization of mobile using RSS and Hilbert-Huang transform. *Int J Eng Technol (UAE)* 7(2.7):253–256
2. Gundu RP, Pardhasaradhi P, Koteswara Rao S, Gopi Tilak V (2018) TOA-based source localization using ML estimation. *Int J Eng Technol (UAE)* 7(2.7):742–745
3. Sirdhara AL, Ratnam DV (2019) Multipath mitigation in GPS receiver using Taylor integrated bidirectional least mean square algorithm. *Trans Emerg Telecommun Technol* 30(12). <https://doi.org/10.1002/ett.3760>
4. Srivani I, Siva Vara Prasad G, Venkata Ratnam D (2019) A deep learning-based approach to forecast ionospheric delays for GPS signals. *IEEE Geosci Remote Sens Lett* 16(8):1180–1184
5. Vishnu Vardhan Reddy VC, Manohar Reddy M, Akshitha G, Ramesh KS, Revathi R (2019) Evaluation of tropospheric and clock errors for precision GPS positioning and navigation. *Int J Recent Technol Eng* 8(2):882–887
6. Safeea M, Neto P (2019) Minimum distance calculation using laser scanner and IMUs for safe human-robot interaction. *Robot Comput-Integr Manuf*. <https://doi.org/10.1016/j.rcim.2019.01.008>
7. Cheerla S, Venkata Ratnam D, Teja Sri KS, Sahithi PS, Sowdamini G (2018) Neural network based indoor localization using Wi-Fi received signal strength. *J Adv Res Dynam Control Syst* 10(4):374–379

8. Cheerla S, Venkata Ratnam D, Saivamsi Y, Kundana PV, Vaishnavi S (2018) Analysis of RSS based path loss model for cooperative localization. *J Adv Res Dynam Control Syst* 10(4):369–373
9. Zafari F, Gkelias A, Leung KK (2019) A survey of indoor localization systems and technologies. *IEEE Commun Surv Tutor*. <https://doi.org/10.1109/COMST.2019.2911558>
10. Fu Y, Liu R, Zhang H, Liang G, ur Rehman S, Liu L (2019) Continuously tracking of moving object by a combination of ultra-high frequency radio-frequency identification and laser range finder. *Int J Distrib Sensor Netw*. <https://doi.org/10.1177/1550147719860990>
11. Rehman SU, Liu R, Zhang H, Liang G, Fu Y, Qayoom A (2019) Localization of moving objects based on RFID tag array and laser ranging information. *Electronics* (Switzerland). <https://doi.org/10.3390/electronics8080887>
12. Gottapu SK, Appalaraju V (2018) Cognitive radio wireless sensor network localization in an open field. In: 2018 conference on signal processing and communication engineering systems (SPACES). Vijayawada, pp 45–48. <https://doi.org/10.1109/SPACES.2018.8316313>
13. Rajendra Prasad C, Bojja P (2017) A review on bio-inspired algorithms for routing and localization of wireless sensor networks. *J Adv Res Dynam Control Syst* 9(18):1366–1374
14. Akram PS, Ramana TV (2017) Mobile aided improved trilateral localization by adopting random way point pattern. *ARNP J Eng Appl Sci* 12(21):6080–6086

Localization Algorithms and Approaches for Navigation in Advanced IoT Applications



P. Kanakaraja, Sarat K. Kotamraju, L. S. P. Sairam Nadipalli,
S. V. Aswin Kumer, and K. Ch. Sri Kavya

Abstract This chapter mainly deals with the current trends and innovations of the localization like inertial measurement units (IMUs) and radio frequency identification (RFID). The above techniques include both indoor and outdoor localization for the Internet of Things applications. The IMUs mainly focus on three algorithms, namely step detection algorithm, stride length detection algorithm, and peak detection algorithm. This chapter also addresses the current scenario of the implementations of both indoor and outdoor localization. The indoor and outdoor localization is widely using mobile phones having inertial measurement units like accelerometer, gyroscope, barometer, and magnetometer. Everybody having their gadgets like mobile phones, tablets, etc., that is measured the sensor data and combined to evaluate the location of a particular person or object with improved accuracy. The tracing and tracking of culprits and missing persons are done by these localization techniques, which is very helpful and suitable for investigation teams in the absence of GPS.

Keywords Indoor localization · Outdoor localization · Internet of Things (IoT) · Inertial measurement units (IMUs) · Step detection algorithm · Stride length detection algorithm · Peak detection algorithm · Global positioning system (GPS)

P. Kanakaraja · S. K. Kotamraju (✉) · L. S. P. S. Nadipalli · S. V. A. Kumer · K. Ch. Sri Kavya
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
e-mail: kksarat@gmail.com

P. Kanakaraja
e-mail: pamarthikanakaraja407@gmail.com

L. S. P. S. Nadipalli
e-mail: sai.nadipalli@gmail.com

S. V. A. Kumer
e-mail: svaswin@gmail.com

K. Ch. Sri Kavya
e-mail: kavyakorada@gmail.com

1 Introduction

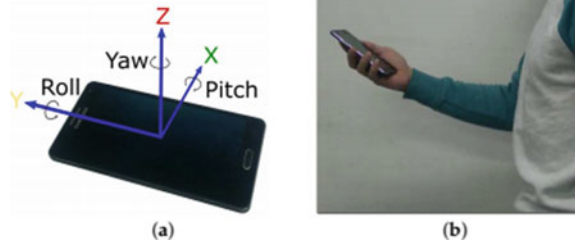
IoT sensors RFID, IMU, hall sensors, pressure sensors are the main sensors which are going to work with localization [1] it appears strange to implement perhaps that now here GPS actually shown and that is the key which are not really looking at anything with respect to localization [2] with GPS Systems because GPS [3] has its own problems and it has its own advantages it may have to use an indirect way of measuring distance and from distance try to arrive at localization [4] of these sensors. Let us consider RFID and look at what are these applications that will require RFID, to start putting some of applications in forward to you take a large warehouse, when you say warehouse it could be a warehouse where there are garments there can be jewelry or it could be a departmental store where everybody goes and buys groceries. So, all these essentials have a lot of items, and keeping inventory of these items is a tedious task; today, there is what is known as a barcode, each of these items has a barcode and they are barcoded and placed in a particular rack, and each time a customer picks an item, the item is put into a basket and brought in front of the point of sale system; there the barcode is read, the amount is deducted, and there is a billing which is all familiar with; this is fine as far as barcodes are concerned; supposing you want to do inventory, this barcode is not going to help you at all. It is only for billing; it is only meant to know the item and so on so forth. It cannot tell you where and how many of these items are available, which means anything with respect to inventorying them are not going to work. So, for that, what people have done is a dumb way of doing barcode is good, but it is dumb; let us now try and replace it with something much more useful something more interactive. So, they gradually started replacing the barcodes with RFID tags.

2 Indoor Localization Using IMUs

Now take the topic of again it is a same thing about Localization [5], let us look into an indoor application and you are trying to reach a position or a location, not position a location in inside an indoor, large industry or a large building complex. There are multiple floors, and in these multiple floors, there are certain designated locations where there is a cafeteria, there is a library, there is a particular laboratory, and there is a testing center thing like that. These are designated locations within not only within a floor, but also within the complete building. To scale down this problem and just look at one floor where such locations known locations are there, plus there are also office rooms and your own sitting cabins which are there and so on. Let us explore the use of these IMU sensors, inertial measurement unit systems to build an application of this nature. The generic and full-fledged applications can build on top of it. There are critical steps in building such an IOT application. Any application can be designed; for example, A. trying to locate a person inside a building B. how to come out of the building in the event of a catastrophe, both are good applications, and

those applications are based on these IMU systems. Next, the IMU-based localization [6] as an IoT application has problem with GPS. The working with the GPS is not necessary anymore, and the designer should concentrate on how one can use these IMU sensors to detect people inside a building and also find out where a person is inside a building and so on. Then, let us see what these IMU sensors are all about. When we talk about IMU sensors, we refer to them as three possible systems which are a combination of three sensors having different principles of operation; one is an accelerometer, and the other is a gyroscope, and the other is a magnetometer. Now, what are these basic sensors and how what is their working principle? The key is that inertial measurement units IMUs as it is shown here consist of one or more of the sensors as I mentioned, measuring change in kinetic, kinematic sorry kinematic energy of a moving body. It is change of kinematic energy of moving bodies that is being sensed and you basically look at if you leave magnetometer out for a moment because it has other applications, you just look at these two, accelerometer and gyroscope sensors; gyro essentially gives rotation rate of the body, that is the key here and accelerometer provides information about linear acceleration of the body. So, very clearly, they have different requirements and if you talk about 3D motion of a body, you need three axis accelerometers and you need three axis gyroscopes. The IMU's embedded in phones, they will be talking about 3 axis accelerometer because we talked about 3D motion essentially. Usually, the axis of the gyro and the axis of the accelerometer are coinciding for starting reference point; both the axes are usually together. Normally they coincide; that means, example is in a 3D orthogonal coordinate system, there are sensors to measure linear accelerations on each of these axis and rotation rate of the same axis. Then it becomes major advantage of these systems, but you will also have to note that how are these sensors and how do you relate these sensors to what is embedded in your phones. The accelerometer sensor must have three axes. When a person holds a phone in his hand and move forward, there will be a dominant axis which will give you some data which are processed to find out "where the person is?" using some dead reckoning-based mechanism. In dead reckoning mechanism, it must essentially estimate the current position from a known position that is the person is in some position. The system takes that known position and estimates the step length, the stride length and so on and then finds out the current position using the previously known position [7] is the commonly known as the dead reckoning-based method. The simple dead reckoning-based methods is used to actually estimate the current location. But the trouble with these sensors which are IMU sensors these are MEMS based sensors inside are if you do any measurement using these sensors, the key that you have to keep in your mind is which is with respect to these sensors is the accumulated errors. So, this is a nightmare, and there are several papers which talk about how to come out with smart algorithms which will sort of compensate for the accumulation errors that are associated with these sensors. One way is do not do integration at all with the accelerometer data that you have. Suppose you are interested in estimating velocity and distance, you will have to do one integral for estimating the velocity and you have to do one more integral for estimating the final distance. If you have to do finding out current position using

Fig. 1 Testing setup. **a** Axis notation. **b** Holding in hand mode



accelerometer and using this double integral method, an algorithm which will sort of compensates the accumulated errors.

Another approach is the raw accelerometer value and find a signature from it, that signature is used in a smart way and then it will estimate the current location. One is just using the data do integration for distance; integration is for velocity and double integration for getting the distance. Otherwise, the other way is taking the raw values and do some amount of processing on them [8]. It is very simple for you even to imagine what this, what kind of signature will actually come from an accelerometer. It is quite simple for you to imagine. Take a phone and find the axis X , Y and Z . If a person holds in his hand and start walking, then one can find the direction of phone moves like that. It makes a very fine movement. So, if a person catches hold of the Z -axis and looks at the signature coming from a Z -axis of system, then Z -axis of the accelerometer has a good possibility that may actually estimate the step length quite well. Then, it provides the result that where the person is keeping the phone like in trouser pocket or chest pocket or in his hand and many people doing this inside building, when they are walking, they are all the time looking at WhatsApp.

That is they are holding the phone like this as shown in Fig. 1 and looking up their WhatsApp messages or texting or doing whatever, this is simply the texting mode [9]. Then, you have the walking mode that is hold the phone either in the right hand or in the left hand. So, whatever algorithm is coming out should be very clear. It can be the phone can be clasped either in the right hand or in the left hand depending on whether the person is a left hander or a right hander and that person could be walking, or a person could be running, because in the event of fire or the gas leak or a partial building collapse or whatever needs an app to evacuate the building at the earliest, sometimes a person may have to walk briskly leading to small sort of equivalent of running almost. So, in that situation the phone so, then if you start imagining what is actually happening to these sensors inside the X , Y and Z direction, 3 axis accelerometer sensors; what kind of signature you get, it is hardly imagine. So, the person will receive some funny waveforms essentially. If a person holds the phone in his hand and start moving, so that is essentially the suspense toward showing something very useful. Also, this is one part; the other part is what if a person holds this phone and for some reason that the person heard that there is a partial building collapse, the fire and all that. The person actually might shake hands and legs. Then,

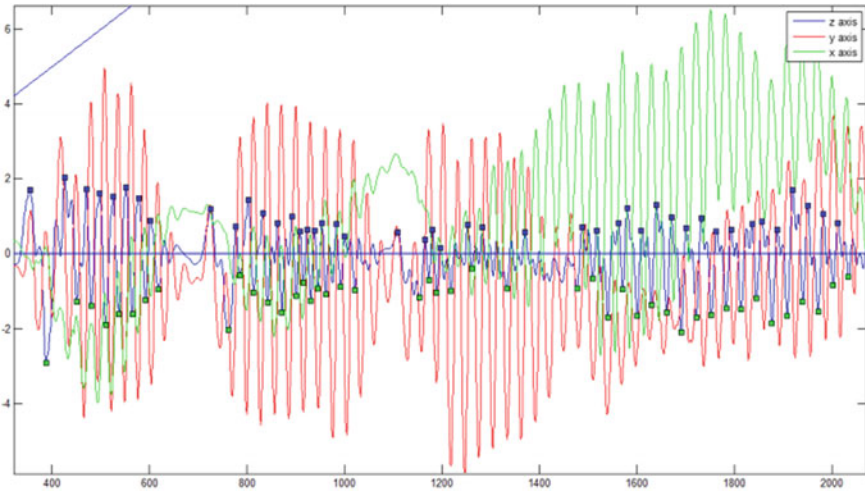


Fig. 2 Three axis accelerometer sensor data

the entire three axes can have a shaky waveform and the waveform that is associated with it is because of the shaking of the phone in his hand. So, the person may have to look at algorithms to eliminate the shaking of the waveforms.

So, let us look at some nice pictures, a view graphs and then build a story and then start looking at what are the issues accumulated errors and if it is not accumulated error how do you have the toughen the problem of a stride length estimation. Look at the results in Fig. 2, there are three axes results that you see here; three different colors. This is directly from a phone. The Z-axis is shown; the Y-axis, the Y- and X-axis are also shown.

This is a phone shake of all axis accelerometer data. In that, there are points where these locations here are no accelerometer data. Similarly, here, the person has stopped moving. So, here it indicates there are peaks and troughs in all three axes with some periods. Clearly there is a person has stopped moving here. Again, the person starts moving here. Again, the person that is not moving here then, some of the axis have moved away and the waveform is oscillating quite a bit. So, this is filtered accelerometer value. To use peak detection algorithm [10], one can easily calculate the number of peaks as show in Fig. 3.

2.1 Peak Detection Algorithm

This is essentially the three axes that you have you can see that the Z axis is the most prominent. This is clearly phone is held in the hand and a person is moving forward and you can see there is a peak, there is a maximum value and the minimum value call it and a_max accelerometer max value and then a_min value, max min, max min

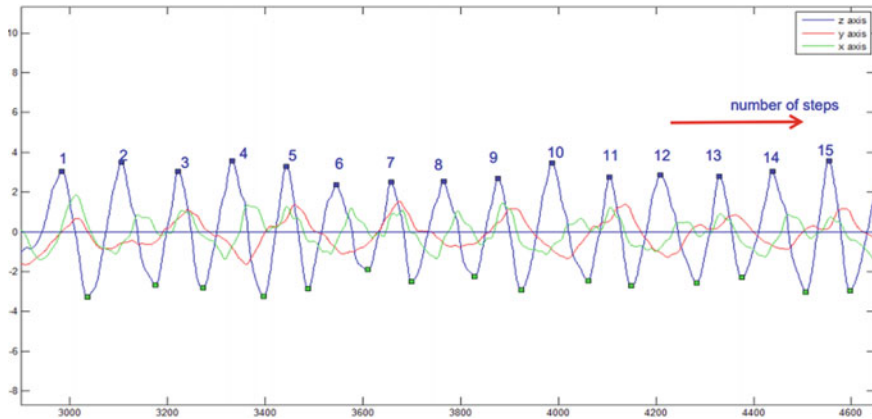


Fig. 3 Filtered accelerometer value

and so on. Now, it is quite easy. If you take from this peak to this peak, that must be one single step. It must be one step, and this is another step. This is another step. So, starts counting the peaks to find out the number of steps. Very prominent Z-axis is telling you that there are fifteen steps taken by an individual person, holding the phone perhaps in his or her hand. So, this is equivalent of a texting mode. So, this graph is for a person holding in the hand maybe right hand maybe left hand should not matter. And moving forward, the Z is moving up down up down up down and you can see beautifully that this up down movement of the Z axis is nicely captured. But, the stride length between these two peaks is not known. There are number of peaks which are the step count, step detection [11], but you do not know what the stride length is. This chapter already mentioned about motion mode recognition is an important thing whether you are holding the phone in the left hand, right hand. People can hold phone in any position; can be texting mode, can be in swimming mode, swinging mode, running mode, carrying phone in a bag walking while walking over while talking over a phone; that means, the phone is held against the year and you can do many things. For simplicity, just group all of them, but at least a few of them into 3 groups for major part of the work one can say that it is in either texting mode; that means, you are holding the phone like this or your walk you are holding the phone in your hand and you are moving forward either left hand or right hand, again a very important or you are running with this phone. These are three things. The signatures coming from these IMU sensors are observed to process them and accordingly and estimate the current position from known previous positions and all that. These are the three modes [12]. Then, the mode detection essentially looking at humans, who can hold phone; who can hold the phone in any 3 possible you know orientations, but there are other orient possibilities, but we just limited our discussion to 3 possible ways by which human can hold a phone. Human can hold a phone in texting mode, usually that is what you do when you are doing when you are looking up when you are walking and checking your email and WhatsApp messages and SMS's and all

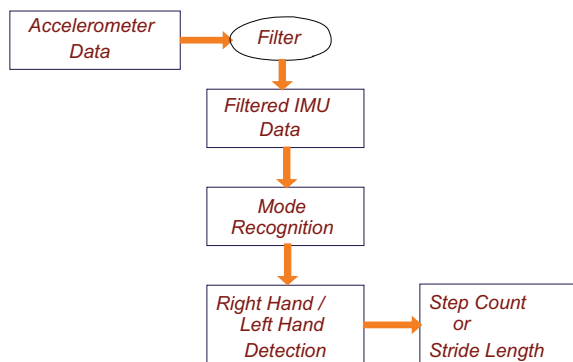
that with the risk that you may trip and fall. Never use any of these modes when you are walking, but anyway that is reality; people walk with this mode. So, let us stick to this. So, the human has texting mode, swinging mode and running mode. Swinging mode is holding the phone in human palm and moving forward in a basically sort of marching forward. So, that is essentially in human hands are swinging all the time and running mode is the human hold during running and then the elbow is bent for running to catch a bus or trying to reach out to nearest transport or house because the human has some urgency to get in or whatever. So, those are the modes which you can think of. The designer needs to detect this mode. To do that, one easy way is to look at the gravity vector which is easily available from the accelerometer.

2.2 Step Detection Algorithm

Now, if you look at texting mode the gravity vector will be very high on the Z-axis. Similarly on swinging and running mode the magnitude of gravity vector will be a high on the; high means what? Its high essentially means it will be 9.8 m/s^2 . It will be as close to 9.8 m whichever is the highest means it is in that particular mode depending on whether it is Z, X or Y you essentially can differentiate. So, essentially just look at the gravity vector; find out which is the one that is closest to 9.8 m/s^2 and then decide that it is indeed that mode under which the phone is being held. You can capture all of that into this kind of a nice mode into a nice algorithm as shown in Fig. 4.

To filter the accelerometer data from the smart phone [13], several filters can be used like Chebyshev filter, Butterworth filter with different orders of filters. At first, Butterworth filter is applied for instance. Butterworth may work well for your application and now the question will be, “what is the order of the filter that you may want to apply on the raw data?” So, anyway that is one bullet again. So, you may have to look at the type of filter. Easiest is collect this raw accelerometer data and read little bit about filters and try them in MATLAB. MATLAB will help you

Fig. 4 Step detection algorithm



to apply different filters on the raw data directly. But, you can try anything else you could look at any other toolbox not necessarily elevated to MATLAB because MATLAB is licensed software. So, anyway you end up with filtered IMU data, and then, you do mode recognition. In what mode is the phone being held; is it any one of the three that mentioned earlier and in what hand is it in which hand is it not what hand? Which hand; right hand, left hand? Then, you estimate the Step count and this hard problem Stride length. This is hard because it is an unsolved problem even today. Stride length is the hard problem. Step count you can get very accurately, but stride length estimation is perhaps the hardest problem. So, you can try MATLAB for processing as mentioned; you can also use the open-source tools like Octave. Octave, try Octave, it is the open source you can try that and then you will have to estimate the stride length alright. Here, you have raw accelerometer, you have filters, you have lot of noisy data, you have to pass it through high pass, low pass, band pass filters, what type of filter to choose, and this is an exercise for you. Then, you have to come up with human activity detection. Then, you have to do the filter data to detect the peak detection, to obtain the peaks; then, individual peaks have to be detected; gravity vectors peak above a threshold is used to detect the activity. Next is quite simple. You do not know whether you are holding your phone in the left hand or right hand, but it is pretty straight forward for texting mode. But for swinging mode, you can look at phone on the; if you are holding the phone on in the left in the right hand, the magnitude of acceleration due to g along x -axis is quite high as compared to the axis are high to other axis and is always positive that is very important. The values that you read off essentially are positive as far as the phone held in the left hand is concerned. When in swinging mode, the so both are with respect to swinging mode. The magnitude of acceleration due to g along x -axis peaks is high compared to other axis, the same as what you have seen in the right hand, except that the difference is the values are now negative. So, that is the key difference between the swinging mode, right hand or left hand. Now, what is it for running mode? Well, I have not put a slide on that and I expect you with the basic code that is available for gathering data available to you. You should be able to sort of find out just run and collect data and do a little bit of processing and find out what is the way the method that you can adopt for detecting left hand and right hand when you are in running mode. Next is step detection. Once you perform peak detection algorithm and you apply the necessary threshold-based analysis to each of the peak, then you considered a peak as a step [14].

2.3 Stride Length Detection Algorithm

At first, start from location A; that is, here, you can see in Fig. 5, there is no accelerometer values here. What is the x -axis? X -axis you can see is time. Y -axis is acceleration, and there is no acceleration means that person is stationary. Let us say this is the point where you got the last GPS [3] coordinate, you can use that. You are interested in hybridizing between GPS and IMU systems you can store, the last known GPS

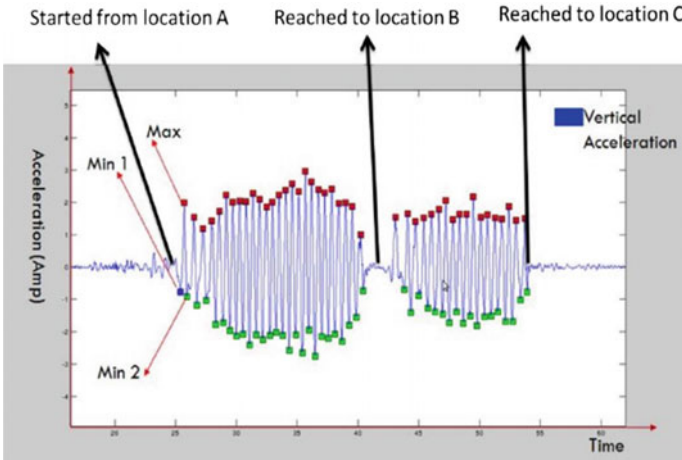


Fig. 5 Stride length detection algorithm

here and then let the IMU stake over; you can see that the max is indicated here, the min is indicated here. There is a second min as well. So, there is a difference between the min of 1 and min of 2; somehow, you have to estimate the minimum. Again, you have to try; this is a very important thing. You have to keep trying and see which suits you most. Then, person moved quite a few steps. If you count the number of peaks, you essentially do the number of steps that the person has taken [15]. Then, what has happened; after they walked, he or she walked, the acceleration is 0 which means that person has reached location B. Then again, the person has walked a few steps and he has reached this location C. So, all of this is nicely captured, and use this minimum and maximum that you see in these peaks to estimate the stride length. So, what is a simple expression? Well, let us see how that goes about and you can see that stride length estimation, you can use this simple expression.

$$\text{Stride Length} = K * \sqrt{\sqrt{(a_{\text{max}} - a_{\text{min}})}} \tag{1}$$

where K is a constant.

Orientation is another issue because you talk about a building you are not just all the time walking straight, but you are also turning. Inside buildings it can be sometimes a 45° turn, but often it is a 90° turn. So, orientation you need to get and you could use magnetometer quite effectively. This capable of detecting fluctuations and you could combine the accelerometer data with magnetometer data in order to arrive at the orientation of the of a person how as and when he is trying to reach the designated location or trying to evacuate from within a building. So, it can be very effectively used [16]. You can see that each mode has its own signatures that you will get. The unit algorithms for more recognition in any case. These are step detected in

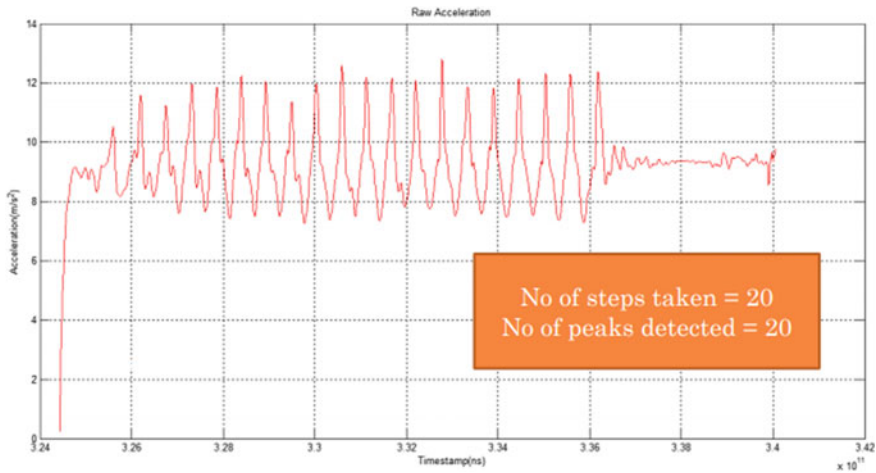


Fig. 6 Steps detected in texting mode

texting mode. Again, the x -axis is time; y -axis is acceleration; the number of steps taken is 20; the number of peaks detected is also 20 as shown in Fig. 6.

3 Conclusion

This chapter is discussed not so much about GPS and all that, but it is about how IoT applications built with IoT sensors, which can assist aid in solving a very large problem of Indoor localization of sensor nodes without GPS, or absence of GPS [3, 18, 19]. The Inertial Measurement Units (IMUs) having three important algorithms are implemented for Indoor localization [4] such as mainly step detection algorithm, stride length detection algorithm, and peak detection algorithm are discussed.

4 Summary

In this chapter, we are summarizing the main concentration on the concept of indoor localization [17]. The indoor localization [17] can be possible by using RFID, pressure based localization [2] and IMUs like gyroscope, accelerometer, hall-effect sensor, and barometer. Among these RFID, IMUs are common in all the localization methods, so if these methods are used to implement the localization, then we can easily implement hybrid localization of both indoor and outdoor in the absence of GPS [3, 18, 19] with high performance and accuracy.

References

1. Jyothirmai V, Ramesh NVK, Navyateja Y, Vijay R (2018) Localization of mobile using RSS and Hilbert-Huang transform. *Int J Eng Technol (UAE)* 7(2.7):253–256
2. Gundu RP, Pardhasaradhi P, Koteswara Rao S, Gopi Tilak V (2018) TOA-based source localization using ML estimation. *Int J Eng Technol (UAE)* 7(27):742–745
3. Sirdhara AL, Ratnam DV (2019) Multipath mitigation in GPS receiver using Taylor integrated bidirectional least mean square algorithm. *Trans Emerg Telecommun Technol* 30(12). <https://doi.org/10.1002/ett.3760>
4. Rajendra Prasad C, Bojja P (2017) A review on bio-inspired algorithms for routing and localization of wireless sensor networks. *J Adv Res Dynam Control Syst* 9(18):1366–1374
5. Cheerla S, Venkata Ratnam D, Teja Sri KS, Sahithi PS, Sowdamini G (2018) Neural network based indoor localization using Wi-Fi received signal strength. *J Adv Res Dynam Control Syst* 10(4):374–379
6. Gottapu SK, Appalaraju V (2018) Cognitive radio wireless sensor network localization in an open field. In: 2018 Conference on signal processing and communication engineering systems (SPACES). Vijayawada, pp 45–48. <https://doi.org/10.1109/SPACES.2018.8316313>
7. Wang Q, Ye L, Luo H, Men A, Zhao F, Huang Y (2019) Pedestrian stride-length estimation based on LSTM and denoising autoencoders. *Sensors (Switzerland)*. <https://doi.org/10.3390/s19040840>
8. Xing H, Li J, Hou B, Zhang Y, Guo M (2017) Pedestrian stride length estimation from IMU measurements and ANN based algorithm. *J Sensors*. <https://doi.org/10.1155/2017/6091261>
9. Ho NH, Truong, PH, Jeong GM (2016) Step-detection and adaptive step-length estimation for pedestrian dead-reckoning at various walking speeds using a smartphone. *Sensors (Switzerland)*. <https://doi.org/10.3390/s16091423>
10. Tran TV, Chung WY (2017) A robust peak detection algorithm for photoplethysmographic waveforms in mobile devices. *J Med Imag Health Inf*. <https://doi.org/10.1166/jmih.2017.2175>
11. Diaz EM, Gonzalez ALM (2014) Step detector and step length estimator for an inertial pocket navigation system. In: IPIN 2014—2014 international conference on indoor positioning and indoor navigation. <https://doi.org/10.1109/IPIN.2014.7275473>
12. Zafari F, Gkelias A, Leung KK (2019) A survey of indoor localization systems and technologies. *IEEE Commun Surv Tutor*. <https://doi.org/10.1109/COMST.2019.2911558>
13. Berkelman P, Abdul-Ghani H (2020) Electromagnetic haptic feedback system for use with a graphical display using flat coils and sensor array. *IEEE Robot Autom Lett*. <https://doi.org/10.1109/LRA.2020.2969913>
14. Ying H, Silex C, Schnitzer A, Leonhardt S, Schiek M (2007) Automatic step detection in the accelerometer signal. *IFMBE Proc*. https://doi.org/10.1007/978-3-540-70994-7_14
15. Brajdic A, Harle R (2013) Walk detection and step counting on unconstrained smartphones. In: *UbiComp 2013—Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing*. <https://doi.org/10.1145/2493432.2493449>
16. Sayeed T, Sama A, Catala A, Cabestany J (2013) Comparative and adaptation of step detection and step length estimators to a lateral belt worn accelerometer. In: 2013 IEEE 15th international conference on e-health networking, applications and services, Healthcom 2013. <https://doi.org/10.1109/HealthCom.2013.6720648>
17. Akram PS, Ramana TV (2017) Mobile aided improved trilateral localization by adopting random way point pattern. *ARNP J Eng Appl Sci* 12(21):6080–6086
18. Sijobert B, Benoussaad M, Denys J, Pissard-Gibollet R, Geny C, Coste CA (2015). Implementation and validation of a stride length estimation algorithm, using a single basic inertial sensor on healthy subjects and patients suffering from Parkinson's disease. *Health*. <https://doi.org/10.4236/health.2015.76084>
19. Cheerla S, Venkata Ratnam D, Saivamsi Y, Kundana PV, Vaishnavi S (2018) Analysis of RSS based path loss model for cooperative localization. *J Adv Res Dynam Control Syst* 10(4):369–373

A LabVIEW-Based Resistor–Capacitor Bank Controller for Automatic Signal Generation with a Multifrequency Voltage-Controlled Oscillator (VCO)



Tushar Kanti Bera

Abstract Automatic multifrequency signal generation is essential for a number of practical biomedical applications such as electrical impedance spectroscopy (EIS) and multifrequency electrical impedance tomography (mfEIT). In EIS or mfEIT, the voltage signal is used to apply an electrical signal to the object under study either converting it into a current signal or even applied as a voltage signal itself. The frequency of the signal generated by the function generator or oscillator circuit depends on the values of the electronic components used or the components connected to the function generator IC used. In this paper, a LabVIEW-based resistor–capacitor bank controller (RCBC) has been developed for automatic multifrequency signal generation with a voltage-controlled oscillator (VCO) circuit. The multifrequency signal generation of the VCO is controlled by a resistor–capacitor bank (RCB) connected with a multiplexer board. The RCB is developed with high precision passive components which are switched through the multiplexer board controlled by a LabVIEW-based graphical user interface (GUI). The RCBC is tested for a voltage controlled oscillator (VCO) circuit and found efficiently sweeping the VCO output frequency suitable for EIS and mfEIT applications.

Keywords Multifrequency signal generation · Voltage controlled oscillator (VCO) · Resistor–capacitor bank (RCB) · Electrical impedance spectroscopy (EIS) · Multifrequency electrical impedance tomography (mfEIT)

T. K. Bera (✉)

Department of Electrical Engineering, National Institute of Technology Durgapur (NITDgp),
Durgapur, West Bengal 713209, India

e-mail: tkbera77@gmail.com

Department of Instrumentation and Applied Physics, Indian Institute of Science (IISc), Bangalore,
Karnataka 560012, India

1 Introduction

The function generators (FG) [1–3] are electronic equipments used to generate different types of electrical voltage waveforms. In many areas of engineering and applied sciences, the multifrequency function generators are found essential for a number of practical applications [4]. A multifrequency function generator [2, 5, 6] generates electrical signals of constant amplitude over a wide range of frequency which is required to be applied for several electrical and electronic measurement and signal processing. The function generators or the oscillator circuits [2, 3] could be developed with active or passive electronic circuit components or using an integrated circuit (ICs) connected with some other circuit components. The frequency of the generated signal obtained from the function generator or oscillators output depends on the values of some electronic components used or the components connected to the function generator IC. For many specific applications and research purposes, portable, compact and miniaturized function generator circuits [2, 5–7] are preferably required to be developed as the standalone devices. Automatic generation of sinusoidal signals at different frequency points is essential for bioimpedance measurement techniques [8] such as electrical impedance spectroscopy (EIS) [9–14] or multifrequency electrical impedance tomography (mFEIT) [15–17]. As the electrical impedance of biological tissues produces capacitive reactance due to the effect of cell membrane capacitance, the bioimpedance varies with the change in signal frequency [8]. As the response of the bioimpedance changes with applied signal frequency, measuring impedance at different frequency points, one can collect a lot of information about tissue health. Thus, bioimpedance measurement in multifrequency domain [2, 18–26] has many advantages compared to the single frequency measurement. Therefore, the multifrequency bioimpedance measurement techniques, such as EIS and mFEIT, are found effective in solving many real-life problems. As the EIS and mFEIT are fast, low-cost and noninvasive material characterization technologies, a number of practical applications are found for EIS [18–24] and mFEIT [2, 25, 26].

In EIS and mFEIT, the automatic generation of multifrequency sinusoidal signals is found essential to measure the bioimpedance at different frequency points. In mFEIT, generally, the impedance measurement is conducted at lesser number of frequency points compared to the EIS technique. Though the frequency changing/sweeping process can be, sometimes, operated manually in mFEIT imaging, but in EIS an automatic frequency variation technique is always employed because in EIS the impedance is measured at a large number of frequencies. Therefore, to measure the impedance at large number of frequency points, the automatic generation of multifrequency signals using avoltage controlled oscillator (VCO) or similar circuit/device is essential. If the impedance measurement is conducted by applying a constant current signal, the VCO output is required to be converted to a constant amplitude current signal by using a voltage to current converter circuit [2, 27–29].

The frequency of the signal generated by a function generator or VCO can be changed by changing the values of some particular components used in the function generator circuit or the electronic components connected to the function generator ICs

if used. To automatically generate the multifrequency sinusoidal signal, the values of the specific circuit components are required to be changed automatically in a sequential manner. The generating signal at fewer numbers of frequencies can be handled by replacing the components manually or manually varying the component values when components with variable properties are used. But manual operation or manual circuit modification for huge number of frequency generation process will be time consuming and quite impractical. In such cases, automatic control of VCO is always sought for reducing the manual effort and the data collection time. In this paper, a Laboratory Virtual Instrument Engineering Workbench (LabVIEW) [10, 30]-based resistor–capacitor bank controller (RCBC) has been developed for automatic multifrequency signal generation using a portable VCO circuit suitable for bioimpedance measurements. A resistor capacitor bank (RCB) is developed and connected to a VCO circuit where the values of one resistor and one capacitor are required to be changed to generate a sinusoidal signal of a particular frequency. All the resistors and capacitors are connected through an analog multiplexer board (AMB) which sequentially switches a particular resistor and a particular capacitor to connect them with the VCO circuit to generate a signal with a required frequency. A LabVIEW-based graphical user interface (GUI) is developed to control the multiplexer board by feeding digital data generated by an electronic hardware directly connected to the PC. The voltage generation instruction to the VCO is provided from the PC through the LabVIEW-based GUI, and the digital data are generated to feed it to the AMB. Under the control of GUI, the digital data generating hardware feeds the digital data to the AMB and direct it to switch the resistors and capacitors in the RCB to connect a particular set of a resistor and a capacitor with the VCO circuit for multifrequency signal generation. The digital data generation and multiplexer switching are studied with LED-based circuits. The generated sinusoidal multifrequency voltage signals are studied with digital storage oscilloscope (DSO) and found suitable for EIS and mFEIT studies.

2 Materials and Methods

2.1 MAX038 IC Based VCO Circuit

The voltage controlled oscillator (VCO) [2, 3] is an electronic oscillator which generates the electrical signal at a frequency which can be controlled by the amplitude of an input voltage and/or circuit component values. The VCO circuits either can be developed with the combinations of electronic circuit components or a function generator IC. Function generator ICs can be utilized to generate sinusoidal or other waveforms by connecting with passive electronic components. In this paper, a MAX038 IC [31]-based VCO is used to generate the sinusoidal signal at different frequency levels. Block diagram of the MAX038 IC is shown in Fig. 1 which indicates its basic

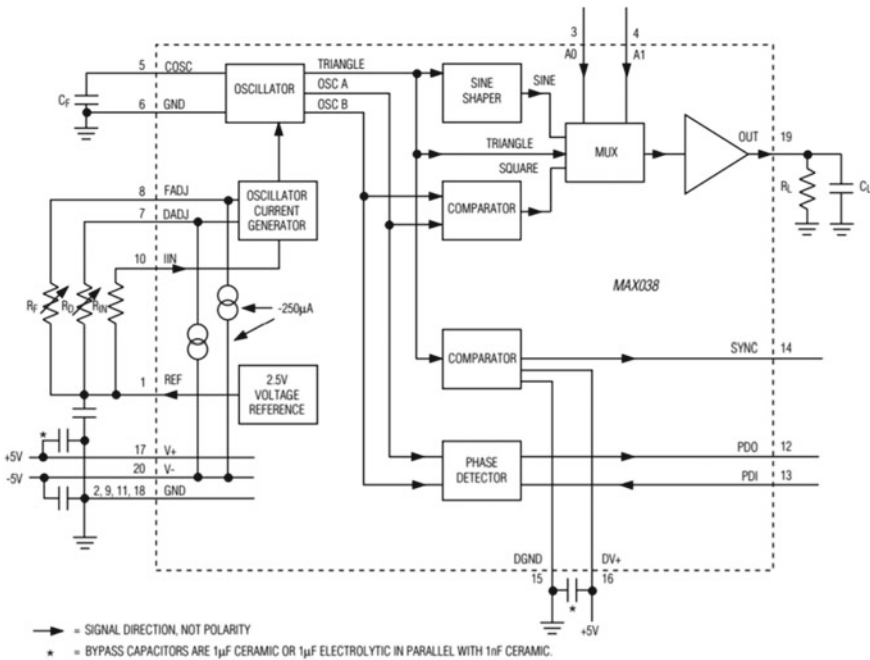


Fig. 1 Block diagram of the MAX038 IC indicating its associated circuit components

operating circuit components. The MAX038 (Fig. 1) IC is a high-frequency, function generator IC which produces accurate, sinusoidal, triangular, sawtooth, square, and pulse waveforms from 0.1 Hz to 20 MHz. The MAX038 IC is required to be connected with a fewer number of external components to make it capable of generating analog signals. The VCO circuit developed with MAX038 IC along with its associated circuit components is shown in Fig. 2a. The frequency of the generated voltage signal available at the VCO output (Pin no. 19) is controlled by changing the value of R_{IN} and C_F with an internal 2.5 V bandgap voltage reference. The output signal of the VCO is an analog sinusoidal signal with an amplitude of 2 V peak to peak signal (symmetrical around ground). The voltage available on the pin named DADJ (Fig. 2b) controls the waveform duty cycle. The output duty cycle can be varied over a wide range (15–85%) by changing this voltage from +2.3 to -2.3 V. When $V_{DADJ} = 0$ V, the duty cycle is 50% (Fig. 2b), and beyond ± 2.3 V, the output frequency can be shifted and circuit may be unstable. Providing suitable DC voltage levels at the Pins A0 and A1, their digital signal levels (0 or 1) can be changed, and accordingly sine, square, or triangular waveforms can be generated at the output.

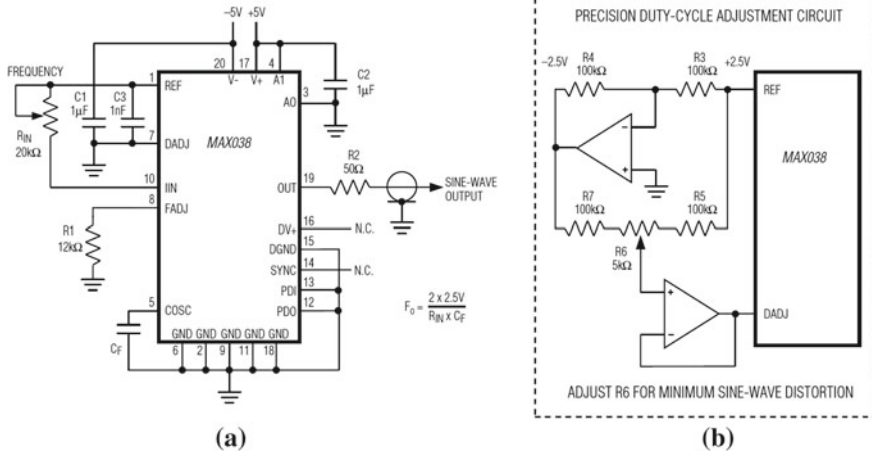


Fig. 2 Circuit connection of MAX038 IC-based VCO **a** MAX038 IC-based VCO circuit, **b** 50% duty cycle; SYNC and FADJ disabled

2.2 Resistor–Capacitance Bank (RCB)

A resistor–capacitance bank (RCB) is developed for providing the different values of the resistors and capacitors in the VCO circuit for multifrequency signal generation. In the RCB, 64 resistors and 64 capacitors are connected through the multiplexer board to generate the sinusoidal voltage signals at 64 distinct frequency points. As shown in Eq. (1), the generated signal frequency (F_o) is a function of the input voltage and the values of R_{IN} and C_F . Keeping the applied voltage fixed, the frequency generation becomes fully dependent on the values of R_{IN} and C_F . In the present study, to change the values of R_{IN} and C_F , a resistor–capacitor bank has been developed after calculating the frequencies from Eq. (1). In the proposed work, 64 frequency points are generated, and hence, the RCB is developed with 64 resistors (Fig. 3a) and 64 capacitors (Fig. 3b) all controlled by a 64:1 multiplexer board. As per the calculation, all the resistor values are not available, and hence, to get the exact value of the resistances, variable resistors (trimmer potentiometers) are used in the RCB.

$$F_o = \frac{2 \times 2.5}{R_{IN} \times C_F} \tag{1}$$

2.3 Analog Multiplexer Board (AMB)

An analog multiplexer board (AMB) is developed (Fig. 4a) to switch the resistors and capacitor to connect the required values of R_{IN} and C_F in the VCO circuit.

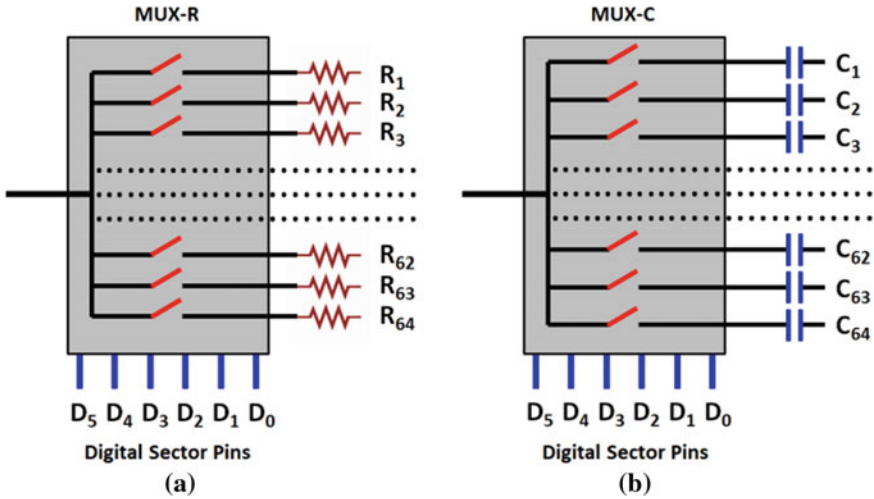


Fig. 3 Schematic diagram of resistor–capacitor bank (RCB) connected with AMB **a** resistor bank connected with 1:64 MUX-R, **b** capacitor bank connected with 1:64 MUX-C

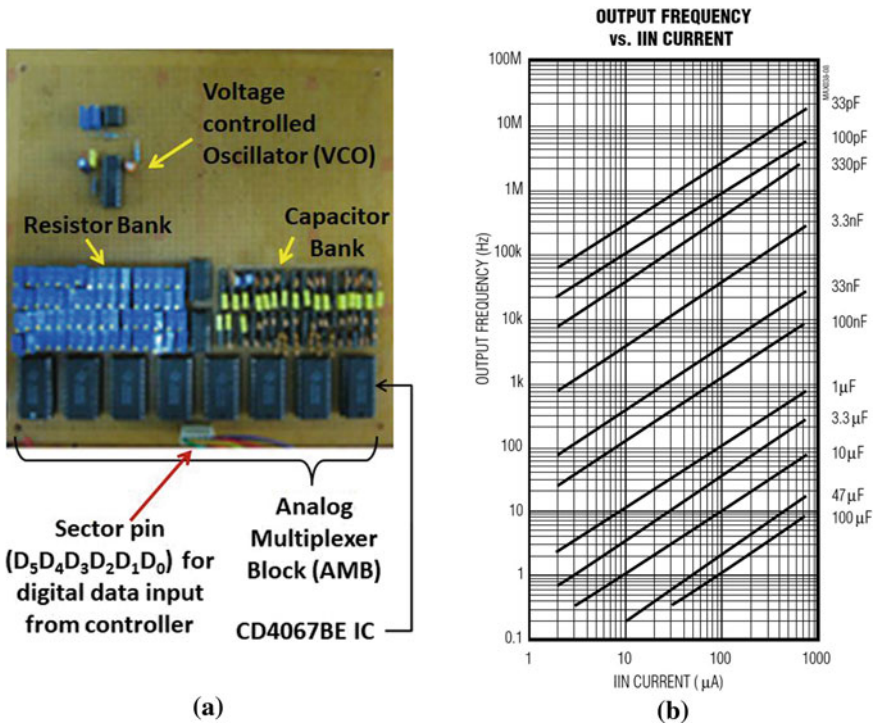


Fig. 4 **a** Resistor–capacitor bank controller (RCBC) board connected with VCO circuit showing AMB and resistor–capacitor bank and VCO circuit. **b** Major frequencies bands obtained at the VCO output for different capacitance values

For generating signals with 64 different frequencies, 64 resistors and 64 capacitors are required to be switched. Therefore, two 64:1 multiplexer modules are required to switch the 64 resistors and 64 capacitors. The AMB is developed with two 64:1 multiplexer/demultiplexer modules named as MUX-R and MUX-C which are controlled by the digital data fed through the selector or control pins ($D_5D_4D_3D_2D_1D_0$). MUX-R switches the resistor bank ($R_1, R_2, R_3 \dots R_{64}$), whereas the capacitor bank ($C_1, C_2, C_3 \dots C_{64}$) switching is done by MUX-C. Each of the 64:1 multiplexer/demultiplexer modules is developed with four CD4067BE ICs from Texas Instruments Inc. USA [32]. The CD4067BE ICs are 16:1 CMOS analog multiplexers/demultiplexers which have low ON impedance and low OFF leakage current and could be used as the digitally controlled switches. It is observed that the MAX038-based VCO generates a particular frequency band for a certain value of the capacitor C_F (Fig. 4b). Within a particular frequency band obtained for a specific capacitance value (C_F), the intermediate frequency points are obtained by varying the resistor (R_{IN}).

2.4 LabVIEW-Based GUI

Laboratory Virtual Instrument Engineering Workbench (LabVIEW) is a graphical programming tool developed by National Instrument Inc., USA [10, 30]. A graphical user interface (GUI) is developed with LabVIEW to interface and control the hardware of the RCBC. The hardware of the RCBC consisted of the resistor–capacitor bank (RCB), multiplexer board (AMB), and a digital data generating hardware. NI USB 6251 DAQ card (National Instruments Inc., USA) [33] is used to generate the digital signals required to operate and control the MUX-R and MUX-C. As the MUX-R and MUX-C both are 64:1 multiplexer, a set of 6-bit parallel digital data ($d_5d_4d_3d_2d_1d_0$) are generated using the GUI. The GUI is written in such a way that it sequentially generates the $d_5d_4d_3d_2d_1d_0$ data sets as the combinations of digital signals 0 (0 V analog) and 1 (5 V analog). The digital data $d_5d_4d_3d_2d_1d_0$ are fed to MUX-R and MUX-C to connect a particular resistor and capacitor respectively with the VCO circuit, and a particular frequency is generated. The GUI is written with an input delay time which will set a particular speed of switching. The time delay is important not only for troubleshooting the circuit, but also a certain delay is required to allow the impedance measurement system for an efficient data acquisition process.

2.5 Multifrequency Signal Generation

The instruction for multifrequency voltage generation is provided from the PC to the VCO through the LabVIEW-based GUI. The command given from PC is used for digital data generation by the NIUSB6251 DAQ card which is connected with the PC through USB port. The digital signal output of the NI USB6251 DAQ is fed to the multiplexer board through SCB68 board (National Instruments Inc., USA) [34]. The

Table 1 Major frequencies generated at the VCO output and the associated resistance and capacitance values as calculated by Eq. (1)

Sl. No.	Frequency (Hz)	C_F (F)	R_N (Ω)
1	10	10 μ F	50.4 k
3	100	1 μ F	56.6 k
6	1k	0.1 μ F	55.7 k
15	10k	0.1 μ F	4.7 k
32	50k	10 nF	10.2 k
42	100k	330 pF	178 k
52	200k	330 pF	87.6 k
53	300k	330 pF	56.9 k
54	400k	11 pF	490 k
55	500k	11 pF	390 k
56	600k	11 pF	324 k
57	700k	11 pF	276 k
58	800k	11 pF	240 k
59	900k	11 pF	213 k
60	1M	1.65 pF	280 k
62	5.0M	1.65 pF	61.8 k
64	10M	1.65 pF	34.2 k

SCB68 board is the connection terminal box of the NI USB6251 card. The digital data generated by the NI USB6251 card are available at the screw terminals of the SCB68 board and are fed to the multiplexer board. Six digital I/O pins of the NI USB6251 card are used to generate the six-bit parallel digital data ($d_5d_4d_3d_2d_1d_0$). The $d_5d_4d_3d_2d_1d_0$ data sets are supplied to the selector pins ($D_5D_4D_3D_2D_1D_0$) of MUX-R and MUX-C for resistor–capacitor switching. A particular R_N and C_F are connected with the VCO circuit as per the switching process in MUX-R and MUX-C with the digital data fed, and a sinusoidal voltage signal with a specific frequency is generated. Few of the major frequencies generated at the VCO output for the associated resistance and capacitance values are given in the Table 1.

3 Results

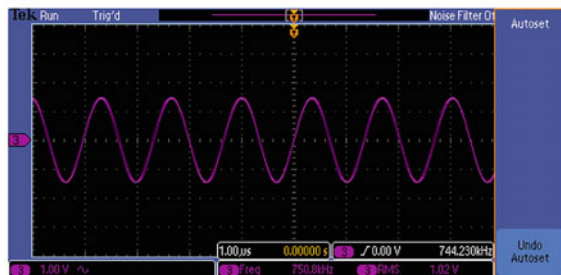
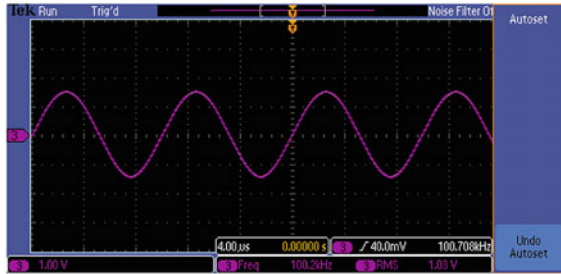
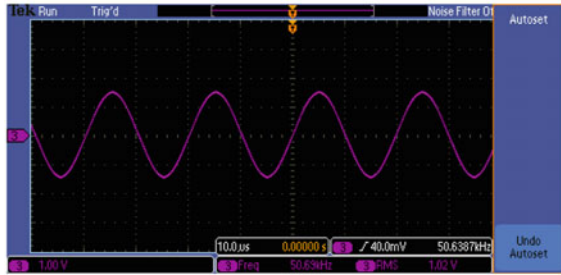
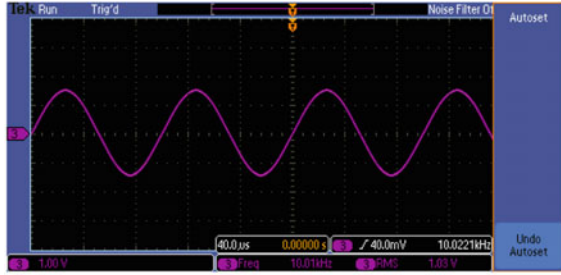
The digital data generated by the NIUSB6251 card under the control of the GUI are tested with light-emitting diode (LED)-based digital data testing kit which is a bank of LEDs connected with resistors in series to limit the current through the LEDs. The generated 6-bit parallel digital data sets ($d_5d_4d_3d_2d_1d_0$) are found sequentially generated from 000000 (1st set of digital data) to 111111 (64th set of digital data) by the DAQ card. After feeding these data sets to the AMB, switching of both the 64:1 multiplexer is studied with the LED-based digital data tester which was used to assess the switching process with the help of a DC powered LED array. The LEDs in the

LED array is connected in series with the $1\text{ k}\Omega$ resistors and the 64 pins of MUX-R and MUX-C. After testing the digital data generation and multiplexer switching, the multiplexer board is connected with the RCB and the voltage signals are generated by the VCO. The voltage signals generated by the VCO from 10 Hz to 10 MHz are studied with CRO. Figure 5 shows the analog voltage signals generated at 10, 50, 100 and 750 kHz frequencies.

4 Conclusions

A LabVIEW-based resistor–capacitor bank controller (RCBC) has been developed for automatic signal generation with a multifrequency voltage controlled oscillator (VCO). The MAX038-based VCO is connected with an automatic resistor–capacitor bank which is controlled by an analog multiplexer board (AMB). The AMB is developed with two 64:1 multiplexer/demultiplexer modules called MUX-R and MUX-C to switch 64 resistors and 64 capacitors of the RCB to generate the sinusoidal voltage signals with 64 different frequencies. Both the MUX-R and MUX-C are developed with four 16:1 CD4067BE multiplexers cascaded to develop the 64:1 multiplexers. All the six selector pins ($D_5D_4D_3D_2D_1D_0$) of the MUX-R and MUX-C are connected with the supply of 6-bit parallel digital data ($d_5d_4d_3d_2d_1d_0$) generated by the controller. Six-bit parallel digital data are generated sequentially by the NI USB 6251 DAQ card controlled by a LabVIEW-based GUI and fed to the AMB to switch the resistors and capacitors. With the connection of the resistors and capacitors with VCO circuit, sinusoidal voltage signals are generated from 10 Hz to 10 MHz. The digital data generated by NI USB 6251 DAQ card are studied with LED-based digital data testing kit and found suitable for operating the multiplexers. Digital data testing also confirms the operation of AMB and the GUI. The VCO output signals are tested in DSO and found as the high quality sine waves suitable for multifrequency impedance measurement techniques such as EIS and mfeIT or other multifrequency electronics or instrumentation applications.

Fig. 5 Analog voltage signals generated by the VCO controlled by the LabVIEW-based resistor–capacitor bank controller (RCBC) **a** 10 kHz, **b** 50 kHz, **c** 100 kHz and **d** 750 kHz



References

1. Graf RF (1997) Oscillator circuits. Newnes
2. Bera TK, Jampana N (2010). A multifrequency constant current source suitable for electrical impedance tomography (EIT). In: 2010 International conference on systems in medicine and biology. IEEE, pp 278–283
3. Razavi B (1997). A 1.8 GHz CMOS voltage-controlled oscillator. In: 1997 IEEE international solids-state circuits conference. Digest of Technical Papers, IEEE, pp 388–389
4. Godfrey K (1991) Design and application of multifrequency signals. *Comput Control Eng J* 2(4):187–195
5. Zhang X, Chen S, Wang H, Zhang Y (2010). FPGA-based multi-frequency excitation and modulation technology in EIT system. In: 2010 3rd international conference on biomedical engineering and informatics, vol 2. IEEE, pp 907–911
6. Kassanos P, Triantis IF (2014). A CMOS multi-sine signal generator for multi-frequency bioimpedance measurements. In: 2014 IEEE international symposium on circuits and systems (ISCAS). IEEE, pp 249–252
7. Nasir NAAM, Mohtar MN, Yunus NAM (2018). Development of signal generator for lab on a chip application. In: 2018 IEEE 5th international conference on smart instrumentation, measurement and application (ICSIMA). IEEE, pp 1–4
8. Bera TK (2014) Bioelectrical impedance methods for noninvasive health monitoring: a review. *J Med Eng* 2014
9. Macdonald JR (1992) Impedance spectroscopy. *Ann Biomed Eng* 20(3):289–305
10. Bera TK, Jampana N, Lubineau G (2019) A LabVIEW-based electrical bioimpedance spectroscopic data interpreter (LEBISDI) for biological tissue impedance analysis and equivalent circuit modelling. *J Electr Bioimpedance* 7(1):35–54
11. Orazem ME (2020) Electrochemical impedance spectroscopy: the journey to physical understanding. *J Solid State Electrochem* 1–3
12. Retter U, Lohse H (2010) Electrochemical impedance spectroscopy. *Electroanalytical methods*. Springer, Berlin, Heidelberg, pp 159–177
13. Orazem ME, Tribollet B (2020) A tutorial on electrochemical impedance spectroscopy. *ChemTexts* 6(2):1–9
14. Azzarello E, Masi E, Mancuso S (2012) Electrochemical impedance spectroscopy. *Plant electrophysiology*. Springer, Berlin, Heidelberg, pp 205–223
15. Record PM, Gadd R, Vinther F (1992) Multifrequency electrical impedance tomography. *Clin Phys Physiol Meas* 13(A):67
16. Bera TK, Nagaraju J, Lubineau G (2016) Electrical impedance spectroscopy (EIS)-based evaluation of biological tissue phantoms to study multifrequency electrical impedance tomography (Mf-EIT) systems. *J Vis* 19(4):691–713
17. Brandstätter B, Scharfetter H, Magele C (2001) Multi frequency electrical impedance tomography. *COMPEL Int J Comput Math Electr Electron Eng*
18. Bera TK, Mohamadou Y, Lee K, Wi H, Oh TI, Woo EJ, Seo JK (2014) Electrical impedance spectroscopy for electro-mechanical characterization of conductive fabrics. *Sensors* 14(6):9738–9754
19. Almuhammadi K, Bera TK, Lubineau G (2017) Electrical impedance spectroscopy for measuring the impedance response of carbon-fiber-reinforced polymer composite laminates. *Compos Struct* 168:510–521
20. Han A, Yang L, Frazier AB (2007) Quantification of the heterogeneity in breast cancer cell lines using whole-cell impedance spectroscopy. *Clin Cancer Res* 13(1):139–143
21. Shah P, Narayanan TN, Li CZ, Alwarappan S (2015) Probing the biocompatibility of MoS₂ nanosheets by cytotoxicity assay and electrical impedance spectroscopy. *Nanotechnology* 26(31):315102
22. Olmo A, Hernández M, Chicardi E, Torres Y (2020) Characterization and monitoring of titanium bone implants with impedance spectroscopy. *Sensors* 20(16):4358

23. Bera TK, Nagaraju J (2019) Electrical impedance spectroscopic studies on broiler chicken tissue suitable for the development of practical phantoms in multifrequency EIT. *J Electr Bioimpedance* 2(1):48–63
24. Dean DA, Ramanathan T, Machado D, Sundararajan R (2008) Electrical impedance spectroscopy study of biological tissues. *J Electrostat* 66(3–4):165–177
25. Soni NK, Hartov A, Kogel C, Poplack SP, Paulsen KD (2004) Multi-frequency electrical impedance tomography of the breast: new clinical results. *Physiol Meas* 25(1):301
26. Romsauerova A, McEwan A, Horesh L, Yerworth R, Bayford RH, Holder DS (2006) Multi-frequency electrical impedance tomography (EIT) of the adult human head: initial findings in brain tumours, arteriovenous malformations and chronic stroke, development of an analysis method and calibration. *Physiol Meas* 27(5):S147
27. Surakamponorn W, Riewruja V, Kumwachara K, Surawatpunya C, Anuntahirunrat K (1999) Temperature-insensitive voltage-to-current converter and its applications. *IEEE Trans Instrum Meas* 48(6):1270–1277
28. Fotouhi B (2001) All-MOS voltage-to-current converter. *IEEE J Solid-State Circ* 36(1):147–151
29. Bera TK, Nagaraju J (2009) A study of practical biological phantoms with simple instrumentation for electrical impedance tomography (EIT). In: 2009 IEEE instrumentation and measurement technology conference. IEEE, pp 511–516
30. Travis J (2009) LabVIEW for everyone. Pearson Education India
31. Data Sheet, MAX038, MAXIM Inc., USA
32. Data Sheet, CD4067BE, Texas Instruments Inc., USA
33. Data Sheet, NI USB6251, National Instruments Inc., USA
34. Data Sheet, NI SCB68, National Instruments Inc., USA

Machine Learning

Detecting Negative Emotions to Counter Depression Using CNN



Pooja Pathak, Himanshu Gangwar, and Aakash Agarwal

Abstract The ever-growing stress and the depression cases have recently caught the eyes of many researchers and medical practitioners. From online consultation to wearable devices, detection of depression became easier. The prolonged effect of negativity due to stress or depression appears on the face of the subject, and this fact can be utilized to detect the depression and stress using Convolution Neural Network (CNN) from a video sequence of the subject recorded while answering a questionnaire remotely.

Keywords Depression · Convolution neural network · Stress

1 Introduction

The rising level of stress in this high demanding world is significantly high. Prolong stress due to any of the reasons be it family issues, extensive workload, relationship issues or aspirations may lead to depression. Depression can be defined as a mood disorder that interferes with the person's conventional routine. This problem is even more prevalent in students because of parental pressure, bleak prospect of jobs, complicated relationships, etc. These were some of the reasons for the dominance of this disorder among students. Many different methods have been rendered by researchers to detect mood disorders like stress, anxiety, and depression. The two prominent approaches are contact based (intrusive) and contact less (non-intrusive). Intrusive methods comprises of sensors providing vital information of the subject under consideration. Bitkina et al. [1] used electrodermal activity (EDA) signals to detect stress in drivers. Peripheral physiological signals have also been used to create a transductive model to detect stress [2]. Electroencephalography (EEG) signals from

P. Pathak (✉)

Department of Mathematics, Institute of Applied Sciences, GLA University, Mathura, India
e-mail: pooja.pathak@gla.ac.in

H. Gangwar · A. Agarwal

Department of Computer Application and Engineering, GLA University, Mathura, India

scalp were used to detect depression [3]. Ahn et al. [4] further used electrocardiograms (ECG) along with EEG to assess stress in subjects. All of the aforementioned methodologies require skilled health workers and maintenance of expensive sensors.

On a contrary, non-intrusive approaches are rather simpler than intrusive approach because it involves approaches of assessment of facial expressions or speech or both which can be conducted via a questionnaire with the patient. With the advents of technology and rise of artificial intelligence and image processing, the non-intrusive methods can now be used in remote areas due to its cheapness and accessibility. For detecting stress and classifying it into three classes low, high, and medium, facial features have been extracted using Histogram of Gradients (HOG), Discrete Wavelet Transform (DWT), and Difference of Gradients (DOG) from frontal faces of the Face Recognition Technology (FERET) dataset. All these histogram features are extracted using Convolution Neural Network (CNN) to detect stress with accuracy of 82% [5]. Local Curvelet Binary Patterns have also been devised in three orthogonal planes to predict the severity of depression with accuracy of 73% [6]. To counter the limitations of State-Trait Anxiety questionnaire, the conventional method for measuring anxiety, recognition rates of facial emotion have been used to automatically predict the anxiety trait [7]. Motion History Histogram (MHH) and Partial Least Square (PLS) regression combined can predict depression using framework trained on Audio Video Emotion Challenge (AVEC2013). MHH detects the subtle changes in the facial and vocal expression, whereas PLS combines these subtle details with depression scales to predict the severity of depression [8].

Since stress is associated with increase in negative emotions like anger, disgust, etc. [9], the questionnaire sessions alone may not always be precise facial features which can assist in increasing the precision and accuracy [10]. Depression Anxiety Stress Scale (DASS), a metric to measure the severity of this disorder, can be predicted using a multi-layered neural network that analyzes the Facial Action Coding System (FACS) of the frontal faces [11]. Different frequencies and orientations of Gabor filters around 40 have been used on the faces detected using Voila Jones face detection on the Japanese Female Facial Expression (JAFPE) dataset and then fed to the Support Vector Machine (SVM) classifier. The trained classifier is then provided the frames from the video of the subject while answering a questionnaire [12].

To add to the research work for detecting depression, a simple CNN model trained to detect the emotions from the subject's frontal face has been used. The trained model is then used to count the negative emotions from the live streaming video, while a questionnaire consisting of both pen and closed is answered by the subject. The proposed work aims to assist the people to reach to professionals if the early symptoms of stress or depression have been encountered, since most of the cases of depression are reported when the case reaches to clinical depression which sometimes leads to the death. This proposed work will monitor the subjects for some sessions of the questionnaire, and if the person shows prolonged negative emotions continuously, then the psychological assistance can be provided to the person which can prevent the unseen danger.

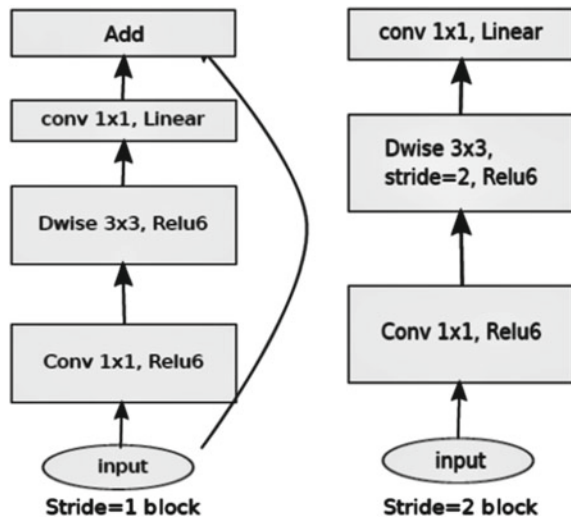
2 Methodology

To train the CNN model, Transfer Learning has been used which takes only the architecture of already proposed model without the last layer. The model is retrained on the Real-world Affective Faces Database (RAF-DB) [13] comprising of seven emotions namely surprise, happy, fear, disgust, sad, anger, and neutral. The architecture chosen for this task is MobilenetV2 as it consumes less number of FLOPS and is lightweight which can easily be used on any type of devices [14]. The architecture for the same is shown in Fig. 1. The parameters of the MobileNetV2 which have been trained on ImageNet [15] are re-updated to satisfy the new requirements. Stochastic Gradient Descent (SGD) method has been used to update the parameters with momentum value = 0.95, this smoothens the learning, and when doing classification, sometimes outperforms its counterparts like Adam which are more prominently used.

The last layer for this classification task is global average pooling [16] which gives a column vector of dimension equal to number of emotions. Categorical cross-entropy loss has been used to penalize and monitor the learning of parameters.

After the parameters are finalized after completion of epochs, the model is saved which can now be used to count the emotions to detect the depression among the subjects from the videos. To perform this task, videos are converted to frames, and these frames are then sent to model for prediction. The whole methodology can be summed in Fig. 2.

Fig. 1 Architecture of MobilenetV2



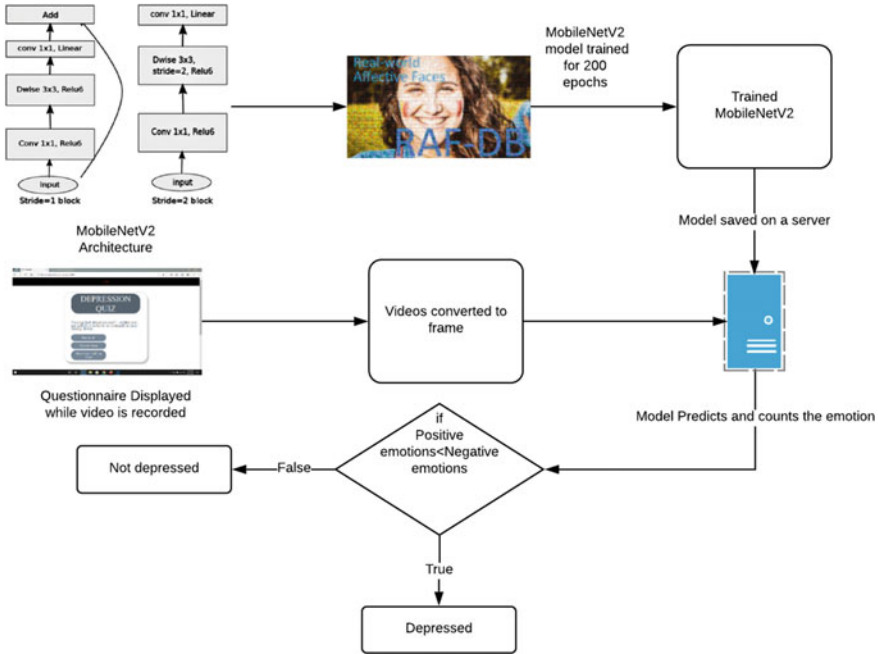


Fig. 2 Methodology

3 Datasets

The chosen dataset for training the MobilenetV2 is RAF, which comprises of simple and compound emotions. Only simple emotions are considered which has 15,340 images of frontal faces from wild. The dataset for cross-validation has been divided into test and train set in the ratio of 80:20. The choice of this dataset can be attributed to the reason that training the model on images of laboratory subjects and then testing it in wild leads to loss in accuracy of the model, but if reverse is done, then this shortcoming can be handled [13].

To prove the effectiveness of the proposed methodology, a dataset with around 18 subjects have been created. Out of these 18 subjects, five were suffering from depression as confirmed by psychologist.

4 Results and Discussion

The training of the model was done for 200 epochs resulting in training accuracy of 99.2% and testing accuracy of 73.827%. The training and testing graph for accuracy is shown in Fig. 3.

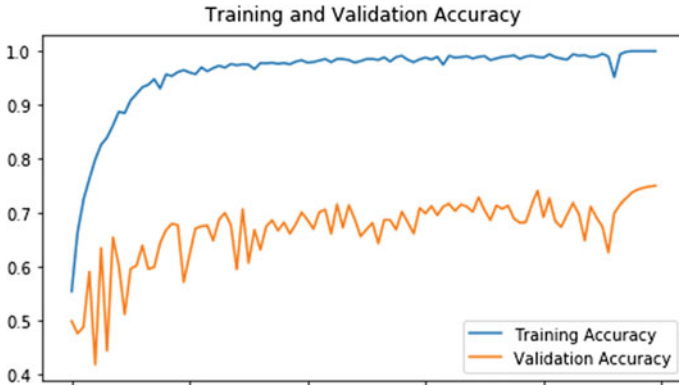


Fig. 3 Training and validation curve

The classification matrix for the classification task is shown in Fig. 4. The trained MobilenetV2 model is then tested on the video sequence of the 18 subjects. The video sequences were converted to frames, then, each frame was sent for emotion recognition, and the count for positive and negative emotions was kept. As the sequence ends, based on the count, the result is displayed. Out of the 18 subjects, eight were deemed as depressive. This constituted all the five clinically proven depressed students and three others who were just sad while answering the questionnaire. The results have been satisfactory since no depressed patient was predicted negative. The classification matrix and other metrics for the same is provided in Fig. 5 and Table 1.

Most of the scholars have used the audio and the visual data for detecting depression by giving more privilege to audio data because it contains the results of the questionnaire session. On the other hand, some other researchers used the facial features to predict the stress or anxiety level using DASS [13]. Therefore, rendering the relative study is difficult.

		Truth Data							Classification overall
		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	
Classifier Results	Class 1	269	12	10	28	21	12	86	438
	Class 2	17	46	10	23	16	11	19	142
	Class 3	4	3	83	20	33	10	46	199
	Class 4	16	3	12	1054	41	7	60	1193
	Class 5	4	5	15	17	318	8	69	436
	Class 6	7	4	13	15	13	103	8	163
	Class 7	12	1	17	28	36	11	392	497
	Truth overall	329	74	160	1185	478	162	680	3068

Fig. 4 Classification matrix of test dataset

Truth Data				
Classifier Result		Class 1	Class 2	Classification overall
	Class 1	5	3	8
	Class 2	0	10	10
	Truth overall	5	13	18

Fig. 5 Classification matrix of subjects

Table 1 Metrics of classification of the model on subjects

Measure	Value
Sensitivity	1.0000
Specificity	0.7692
Precision	0.6250
Negative predictive value	1.0000
False positive rate	0.2308
False discovery rate	0.3750
False negative rate	0.0000
Accuracy	0.8333
F1 score	0.7692
Matthews correlation coefficient	0.6934

5 Conclusion


This work uses CNN to tackle the increasing stress and depression cases. To achieve this, a questionnaire and the facial expressions while answering it are considered. For facial emotion recognition, MobileNetV2 trained on RAF dataset is used. The frames from the video are then provided to the classifier which counts the emotion recognized and provides the apt result. If negativity is more dominant over positive emotions, then the person is deemed depressed and referred to a clinical psychologist. This work as the results prove can assist in detecting early symptoms by monitoring the subjects for three to four sessions, and if they are deemed depressed, help them out. In the future work, the weightage of questionnaire will be improved since responses from questionnaire could provide insight about the mental stress, especially open questionnaire which allows the subjects to speak their heart out. To achieve this, Natural Language Processing (NLP) will be used which will increase the effectiveness of this methodology.

References

1. Bitkina OV et al (2019) Identifying traffic context using driving stress: a longitudinal preliminary case study. *Sensors* 19(9):2152
2. Li M, Lun X, Wang Z (2019) A transductive model-based stress recognition method using peripheral physiological signals. *Sensors* 19(2):429
3. Liao S-C et al (2017) Major depression detection from EEG signals using kernel Eigen-filter-bank common spatial patterns. *Sensors* 17(6):1385
4. Ahn JW, Ku Y, Kim HC (2019) A novel wearable EEG and ECG recording system for stress assessment. *Sensors* 19(9):1991
5. Prasetyo BH, Tamura H, Tanno K (2018) The facial stress recognition based on multi-histogram features and convolutional neural network. In: 2018 IEEE international conference on systems, man, and cybernetics (SMC). IEEE
6. Pampouchidou A et al (2015) Designing a framework for assisting depression severity assessment from facial image analysis. In: 2015 IEEE international conference on signal and image processing applications (ICSIPA). IEEE
7. Huang X et al (2013) Automatic prediction of trait anxiety degree using recognition rates of facial emotions. In: 2013 Sixth international conference on advanced computational intelligence (ICACI). IEEE
8. Meng H et al (2013) Depression recognition based on dynamic facial and vocal expression features using partial least square regression. In: Proceedings of the 3rd ACM international workshop on audio/visual emotion challenge
9. Feldman PJ et al (1999) Negative emotions and acute physiological responses to stress. *Ann Behav Med* 21(3):216–222
10. Pediaditis M et al (2015) Extraction of facial features as indicators of stress and anxiety. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE
11. Venkataraman D, Parameswaran NS (2018) Extraction of facial features for depression detection among students. *Int J Pure Appl Math*
12. Gavrilescu M, Vizireanu N (2019) Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors* 19(17):3693
13. Li S, Deng W, Du J (2017) Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition
14. Sandler M et al (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
15. Deng J et al (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE
16. Lin M, Chen Q, Yan S (2013) Network in network. *arXiv preprint [arXiv:1312.4400](https://arxiv.org/abs/1312.4400)*

Computer Security Based Question Answering System with IR and Google BERT



Pragya Agrawal , Priyanka Askani , Ranjitha Nayak ,
and H. R. Mamatha 

Abstract Machine learning perception and question answering is a fundamental errand in natural language processing. As of late, pre-prepared contextual embeddings (PCE) model, bidirectional encoder representations from transformers (BERT) has pulled in loads of consideration because of its incredible execution in a wide scope of natural language processing (NLP) undertakings. In this venture, the BERT model is fine-tuned with extra undertaking question–answer specific layers to improve its exhibition on Stanford Question Answering Dataset (SQuAD 2.0). A closed domain question answering system is developed, which is ‘computer security’ domain-specific for the use of interactive learning which makes learning very exciting. The system has also been extended by adding document retrieval and cache for faster access.

Keywords BERT · SQuAD · Information retrieval · NLP

1 Introduction

As technology increases, there is a lot of human computer interaction and exponential rise in data in day-to-day life. The field of natural language processing made it possible with its methodologies to use the data to generate a variety of models that makes this interaction easy. One such model is question answering model which requires to understand the query given and respond with a concise answer.

P. Agrawal · P. Askani · R. Nayak (✉) · H. R. Mamatha
Department of Computer Science, PES University, Bengaluru, India
e-mail: ranjithanayak48@gmail.com

P. Agrawal
e-mail: pragya0698@gmail.com

P. Askani
e-mail: priya.askani@gmail.com

H. R. Mamatha
e-mail: mamathahr@pes.edu

There are different types of QA systems based on the question type like fact, list, definition, how, why, hypothetical, semantically constrained, and cross-lingual questions. In general, there are two kinds of question answering systems: closed domain and open domain QA. Closed domain QA is very context specific. It is simpler to build a closed domain QA when compared to open domain QA because of its restricted context. QA systems are better than search engines because they cut a manual effort of humans to look for an answer in the pool of documents returned by the search engines. These systems give the feel of talking to another human. The focus here is on restricted domain QA.

QA frameworks expect to return highlight point answers instead of flooding with documents, coordinating sections as the data retrieval frameworks do. For e.g., “what does the ACL stand for?” The specific answer expected by the client for this question is (Access Control List). The major testing issues in the question answering framework is to give exact answers from colossal information accessible on the Web.

In any case, with ongoing advancements in profound learning, neural system models have demonstrated guarantee for QA. Despite that these frameworks for the most part include a smaller learning pipeline, they need a huge measure of preparing. GRU and long short-term memory (LSTM) units permit recurrent neural systems (RNNs) to deal with the longer messages required for QA. Such systems give the best in class execution for profound learning-based QA. Question answering systems offer a computerized approach to procuring answers for inquiries communicated in characteristic language.

A ton of QA surveys have classified question answering frameworks dependent on various criteria such as inquiries asked by clients, highlights of information bases used, nature of generated answers, question answering approaches, and methods. To fully understand QA frameworks, how it has developed into its current QA needs, and the need to scale up to meet future desires, a more extensive study of QA systems becomes basic.

The main aim of this system described in this paper is to answer the questions asked. The most common question answering system is selecting an answer from the context. At the point when the client offers a question to a framework sitting on a gigantic database of unstructured information, the principal request of business is to diminish that heap to maybe a bunch of records where the answer is probably found. This implies utilizing quick yet non-precise choice techniques regularly for document retrieval. Questions are tokened and sent to a document retrieval phase. Applications of these systems are vastly seen in customer service bots, chat-bot, search engines, and domain-specific bots for learning.

2 Background Work

2.1 BERT

BERT [1] stands for bidirectional encoder representations from transformers. BERT is the first deeply bidirectional, unsupervised language representation, pre-trained using only a plain text corpus. It is a pre-prepared model created by Google which can be tweaked on new information and can be utilized to make NLP frameworks. BERT is a huge model, with 24 transformer blocks, 1024 hidden units in each layer, and 340 M parameters. The model is pre-trained on 40 epochs over a 3.3 billion word corpus, including English Wikipedia (2.5 billion words), and Book Corpus, a dataset containing +10,000 books of various kinds. BERT is a contextual model, and it catches the connections between the contexts in a bidirectional manner. Therefore, the pre-prepared BERT model can be calibrated with only one extra output layer to make cutting edge models for a wide scope of NLP errands.

2.2 BERT Versus RNN, Convolutional Neural Networks (CNN), and LSTM

BERT, as a contextual model, produces a portrayal of each word that depends on different words in the sentence, in contrast to the RNN and CNN models that take in words consecutively, making it hard for long sentence conditions, with no contexts and to accomplish the intensity of parallel processing.

RNN models have momentary memory issues making it hard to recall the more extended data sources [2]. The transformer system is preferred, which has a more straightforward architecture and can train quicker than an LSTM-based model. It is additionally ready to learn complex examples in the information by utilizing the attention mechanism [3].

BERT catches the connections in a bidirectional manner [2], unlike the transformer models that are just unidirectional.

2.3 Research in Question Answering

Question replying (QA) is an all-around explored issue in NLP. Regardless of being one of the oldest research zones, QA has application in many tasks. Some examples could be entity extraction and information retrieval. As of late, QA has additionally been utilized to create chat-bots planned to recreate human discussion and for dialog systems. Customarily, the majority of the research in this space utilized traditional NLP methods, for example, co-reference resolution, parsing, and part-of-speech tagging. IBM Watson, a state-of-the-art QA system, uses these strategies.

But with late advancements in deep learning, neural system models have demonstrated guarantee for QA. GRU and LSTM units permit recurrent neural networks (RNNs) to deal with the lengthier texts required for QA. Further enhancements—for example, memory networks and attention mechanisms—permit the system to concentrate on the most important facts. Such systems give the current state-of-the-art execution for deep-learning-based QA.

2.4 Using BERT for Question and Answering

BERT model is characterized in understanding the given text synopsis and responding to the inquiry from that rundown. To comprehend the question-related data, BERT has trained on SQUAD dataset and other labeled question and answer dataset.

Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, comprising of inquiries presented by crowd workers on a lot of Wikipedia articles, where the response to each address is a portion, or span, from the content, or the inquiry may be unanswerable (Fig. 2).

SQuAD2.0 joins the 100,000 inquiries in SQuAD1.1 with more than 50,000 new, unanswerable inquiries composed adversarially by crowd workers to appear to be like answerable ones.

Training on such dataset makes BERT the optimal choice to build a question Answering model.

3 Proposed Solution

The question answering framework can be portrayed as a four-stage process: question definition, document retrieval, answer extraction, and answer caching. The initial 2 stages is achieved by the document retrieval strategies, answer extraction is achieved by utilizing a pre-prepared BERT model, and the caching stage improves the response time of the model.

Figure 1 shows the various modules in the design of the system. Each module is itself a complete entity, it has its significance and design constraints. Each module is independently built and integrated. The user interface is designed in such a way that any person with little knowledge on Web sites can use it without any problem. The document retrieval phase is tried out with different information retrieval (IR) methodologies, and based on the results, obtained latent semantic indexing (LSI) was chosen as IR methodology. The QA model is built on pre-trained BERT because BERT is better when compared to some older CNN and RNN models.

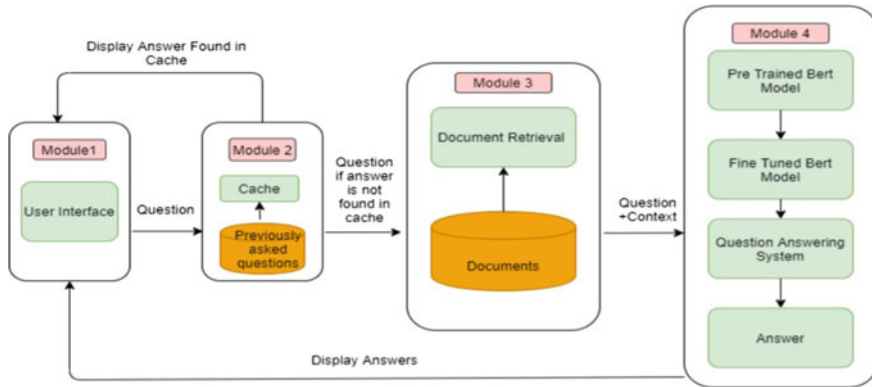


Fig. 1 System architecture

3.1 Collecting and Preparing Data

The dataset used in training the question answering system is called SQuAD. The Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset that was developed at Stanford. The dataset comprises inquiries presented by crowd workers on a lot of Wikipedia articles. The response to each address is a section of content, or range, from the respective paragraph. There are more than 100,000 question–answer pairs on more than 500 articles.

There are two files to get started with the dataset and evaluate and train the models:

- train-v1.1.json—41.13 MB
- dev-v1.1.json—4.2 MB

For preparing domain-specific data, the dev-v1.1.json file was used. The dev-v1.1.json file of SQuAD 2.0 contains a variety of topics for its contexts. They include Normans, computational complexity theory, packet switching, etc. For making the project domain-specific, only the contexts from this file under the topic “computer security” are chosen. These contexts are split into various documents to give an idea of how the documents are indexed in search engines. The length of each document can vary.

3.2 Training the Model

There are two stages of preparing the BERT model for use. They are pre-training and fine-tuning.

The pre-preparing method to a great extent follows the current writing on language model pre-training. For the pre-training corpus, the BooksCorpus (800M words) and English Wikipedia (2500M words) were utilized. The whole process is fairly

In the context of information security, and especially network security, **a spoofing attack is a situation in which a person or program successfully identifies as another by falsifying data, to gain an illegitimate advantage.**

Many of the protocols in the TCP/IP suite do not provide mechanisms for authenticating the source or destination of a message, and are thus vulnerable to spoofing attacks when extra precautions are not taken by applications to verify the identity of the sending or receiving host.

What is a spoofing attack?

A situation in which a person or program successfully identifies as another by falsifying data, to gain an illegitimate advantage.

A cookie is a small amount of data generated by a website and saved by your web browser. Its purpose is to remember information about you, similar to a preference file created by a software application.

While cookies serve many functions, their most common purpose is to store login information for a specific site.

What is a cookie?

A small amount of data generated by a website and saved by your web browser.

Fig. 2 Question–answer pairs for a sample passage in the SQuAD dataset. Each of the answers is a segment of text from the passage

expensive and takes 4 days on 4 to 16 Cloud Tensor Processing Units (TPUs). Hence, the pre-trained model released by Google is used.

Fine-tuning is more straightforward than pre-training. It is inexpensive, takes at most 1 h on a single Cloud TPU, or a few hours on a Graphics Processing Unit (GPU). Time was significantly cut down using the TPU environment in Colab. Two models were explored for their properties and to examine which suits better for the purpose. Table 1 shows a comparison between the two.

Table 1 Comparison of the models used

Feature	BERT-base	BERT-large
No. of transformer blocks	12	24
No. of hidden layers	768	1024
Attention heads	12	16
Exact match	75.9538	75.9875
F1 score obtained	79.1827	79.2559

3.3 Picking the Relevant Document

The technique followed to calculate the similarity between a query and the pool of documents is the latent semantic index that uses singular value decomposition to form LSI vectors. A comparison of LSI vectors of query and documents is done with cosine similarity. The LSI model is prepared with 800 topics. The Genism library in Python was used to implement the same.

If the entered question is not directly found in the documents, or the given paragraph, the system tries to understand it using the bidirectional transformers and matches it with the documents or paragraphs. If the sentence does not relate semantically, the system returns “Not Found”.

3.4 Finding the Right Answer

Once the right context segment is determined, this is then passed, along with the input question, as input to the model. The model sends the context and question, and the following process takes place:

- Pre-processing - Conversion of input passage and question into a SQuAD Example by removal of whitespaces and tokenizing it.
- Addition of SEP (separator) and CLS (clear screen) tokens:
 - SEP is for separating sentences
 - CLS is placed at the beginning of the input example sentence/sentence pair.
- Collecting the following parameters essential for making input Tensor dataset to the model:
 - input ids
 - input masks
 - segment ids
 - example indices.
- Sampling the elements using Sequential Sampler and loading the data using Dataloader.
- Evaluating the data batch wise and obtaining the answer.

4 Results

4.1 Model Evaluation

Fine-tuned the model with one output layer and accomplished accuracy of 80 percent. The model evaluation is based on two measures:

- Exact match: A double proportion of whether the framework yield matches the ground truth answer precisely
- F1 score: Harmonic mean of precision and recall.

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{F1} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

4.2 Document Retrieval for Query

Attempted various ways like TF-IDF and latent semantic index for document retrieval. It was found that LSI works the best for the required system.

The strategy followed to find out the similarity among question and the pool of records is latent semantic analysis which uses particular worth disintegration to shape LSI vectors. Focus is on the LSI vectors of question and documents with cosine likeness. Set up of the LSI model with 800 topics utilized Python Gensim library to build LSI model for document retrieval and accomplished accuracy of about 90%.

The model is evaluated using a script with questions and manually checked if it matches the expected result. Answer prediction for given query from the fetched document using the functions provided by the pre-trained BERT model. BERT has a length limit on the context provided. The lengthier contexts are taken care of by breaking down the context into smaller paragraphs and choosing the most relevant part. Model returns the correct answer 85–90.

4.3 Increase in Response Time by Caching

By caching previously asked questions, assuming a question is most likely posed again, avoids going through other phases which are computation intensive. It also decreases the load on the server which performs some intensive tasks.

QA frameworks intend to deliver precise answers; however, the current QA built framework is succeeded just somewhat. As guaranteed before, all the predefined expectations have been met, and the model is stretched out to deal with a pool of documents and a lengthier context. These expectations are accomplished with a decent exactness and likewise, and a user interface (UI) is made for simple utilization and introduction. In this manner, an attempt to expand the present use of the BERT model from a basic question answering model to incorporate document retrieval and quick access utilizing store has been made.

5 Conclusions

Question answering framework is one of the rising territories of research in NLP. QA frameworks mean to create precise answers, yet the current QA frameworks distinctly somewhat explored various available models and frameworks for question answering systems, understood their functioning and their likening to the users. A search is for the various enhancements done and how recent that they were in terms of development. Finally, chose the BERT model by Google, since at present, it is the best answer for this need. There are a lot of resources and developments done in this model, and it also is fully supported by the Google community.

BERT is a pre-trained model used by a ton of specialists for developing state-of-the-art models. The BERT is utilized and calibrated to build question answering systems explicitly. Integrating BERT with IR tackled the issue of a pool of documents. Storing the outcomes in the database improved the time unpredictability radically.

6 Future Enhancements

The model fabricated for this project is in an area explicit to computer security. The long-term aim is to make it a computer science question answering bot which could be used much for one-on-one learning. This will help in expanding the domain of the system from one sector of computer science to all sectors of it. This will also act as an encyclopedia of the computer science domain which will be very helpful for learning the domain.

Likewise, another feature could be to take the feedback from the client for the application and investigate the feedback taken for the present application and improve the model more by adjusting the model by differing parameters. This feedback could be used to fine-tune the model to better the latest user requirements.

The next enhancement to carry out in the future would be to explore and use better and more proficient information retrieval techniques. These techniques would help in finding a better-matched document with more accuracy for a given question. The IR approaches used here are the most fundamental strategies in light of the fact

that lesser information is available in the form of documents. Utilizing progressively refined IR strategies make the document retrieval process increasingly proficient.

References

1. Zhang Y, Xu Z (2019) BERT for question answering on SQuAD 2.0
2. Devlin, Chang M (2020) Open sourcing BERT: state-of-the-art pre-training for natural language processing
3. Rizvi MSZ (2020), Demystifying BERT: a comprehensive guide to the groundbreaking NLP framework

Digit Dataset Generation Using DCGAN: A ResNet Experimentation



Anurag Wani, Aditya Sarwate, Sourav Sahoo, Sahil Sehgal,
and Manisha Thakkar

Abstract GANs or generative adversarial networks are deep generative models that have the capability to imitate a given data distribution and generate realistic data, that can be in the form of images, music, and even outcomes of scientific experiments. An important application of GANs is data creation and augmentation in cases where there is not enough data or the variety in the data is less for training a model to perform well in real-world testing. A lot of applications ranging from license plate recognition, bank check processing, auto filling form details, etc., make use of digit recognition. The digit recognition and classification accuracy of machine learning models, generally if not always, can be improved when more data (images of digits in this case) is provided. It can help the model to learn distinct features from the wide variety of data for better predictions and generalization. In this study, we use GANs and the SVHN dataset of 73 thousand colored, house number images to generate new and unseen, similar, realistic looking images of digits and then evaluate our GAN model using the Inception Score as the metric for evaluating quality and diversity of newly generated images. Later with an intent to further improve the quality of the newly generated images, we study the effects of modifying the discriminator (CNN) architecture by adding residual blocks to make it deeper, along with some other minor experimentation with hyperparameters, with an expectation of improving the classification accuracy of the discriminator, thus helping the generator improvise in generating more realistic images and finally discuss the results.

Keywords Deep learning · GANs · CNN · Image processing · Residual networks

1 Introduction

GANs [1] were invented by Ian Goodfellow in 2014 which is a category of generative models whose architecture consists of two different neural networks called the generator and the discriminator. The job of generator is to produce fake data and trick

A. Wani (✉) · A. Sarwate · S. Sahoo · S. Sehgal · M. Thakkar
Department of Information Technology, MIT World Peace University, Pune, India
e-mail: wanianurag3@gmail.com

the discriminator, whereas the discriminator classifies if the data fed to it is coming from the real distribution or the fake one. The word adversarial in GAN refers to the fact that the two networks—generator and discriminator—are in competition with each other. Yann LeCun (Director AI, Facebook) has called it as “the most interesting idea in the last ten years in machine learning”. GAN, since its inception in 2014, has continuously gained popularity and is unleashing its potential in various domains in exceptional ways. It has been explored for various use cases like generating videos [2], high resolution images from low resolution images [3], text-to-image translations [4], pose generation [5], music generation [6], medical image analysis [7], image colorization [8], etc. There is also research going on in the field of video streaming in which GANs can help generate additional video frames to increase the frame rates, trying to generate videos from just textual descriptions, etc.

With the increase of computational capacity and the advances in machine learning and deep learning architectures, the task of classification and recognition has become increasingly popular and necessary to automate a lot of manual tasks. To our specific case, i.e., digit recognition, it is used for many types of real-world tasks. Some of them include postal mail sorting, bank check processing, vehicle registration plate recognition, labeling of addresses in textual form from their image counterparts, etc. Generating more realistic and diverse images will help in data augmentation and building larger datasets containing more variety of data samples. This will be further helpful to build classifiers that are more robust and better at generalizing in real-world testing. The later can be achieved using semi-supervised training of the GAN wherein the discriminator also acts as a class label classifier rather than just classifying if the input data is real or fake. In this project, we tested our model firstly on the MNIST digit dataset (1-channel) to ensure that the generator and discriminator were learning correctly with some fine tunings. Later, we used it to generate more complex three-channel new images of digits by learning from the SVHN dataset using the same DCGAN architecture. Next, with an intent of further improvement of the image quality, we experimented by adding residual blocks to the discriminator and then evaluated the Inception Scores of the images generated from these two different model architectures.

2 Literature Survey

GAN training is unstable, time-taking, and the outputs are difficult to interpret owing to the fact that two (generator and discriminator) different neural networks are getting trained and optimized simultaneously to get better themselves, while trying to increase the loss of the other one. This adversary between the two networks can cause different types of failure modes [9]. Various experimentations have been done to the GAN architecture to carry out different types of tasks by the research community. Some examples include Vanilla GAN [1], SRGAN [3], FCC-GAN [10], etc. however, we used the DCGAN [11] architecture for our study.

Deep convolutional GAN or DCGAN [11] is a type of GAN architecture that was proposed by Soumith Chintala and Alec Radford in 2015 which is a relatively more stable GAN. It takes the advantages of (convolutional neural networks) CNNs which help in processing multi-channel images more effectively rather than plain ANNs. The generator is a multi-layer deconvolutional neural network that maps a given uniform random input noise vector Z , to a specified tensor of dimensions, equal to that of the images from the real data. The discriminator is a deep convolutional neural network which acts as a classifier with sigmoid activation function to identify and output the probabilities of the input image being real or fake. Some specifications of it are the use of strided convolutions instead of pooling layers for learning spatial upsampling and downsampling, avoiding hidden fully connected layers, batch normalization (mean zero with unit variance), and some others which will be discussed in the later sections. In this study, we used the DCGANs architecture along with the optimizations later released by Soumith Chintala, the author of DCGANs himself for improving the performance and stable training of the traditionally proposed DCGAN.

One of the challenges in GAN optimizations and interpretations is the evaluation of the outputs of the models. In their paper, Salimans et al. [12] proposed the Inception Score (IS) as a metric that uses Kullback–Leibler (KL) divergence (a measure of how similar/different two probability distributions are). Inception Score utilizes the Inception pretrained model and runs the generated images through it. The score captures two characteristics of the generated images—(a) Diversity (b) Quality. The higher the score, better the quality and variations of the generated images.

Some other techniques for improved GAN training from the paper include virtual batch normalization (VBN) and feature matching. VBN avoids high dependence of output of a specific input from the other inputs of current minibatch and is also more computationally demanding than batch normalization.

In our study, we have experimented modifying the discriminator by making it deeper with the use of residual blocks [13] with skip connections, to help it become better at classification thereby helping the generator learn to generate better images over time. The skip connections are helpful in increasing the depth of the CNN, helping it learn more features from the given input data while reducing the problem caused by vanishing and exploding gradients.

3 Implementation

3.1 DCGAN Architecture:

Generator. The generator takes in a 100-dimensional input vector Z sampled from a random uniform distribution. We used four deconvolution layers for upsampling using fractional strided convolutions and 5×5 filters. Batch normalization followed by Leaky-ReLU [14] was used. For the output layer, tanh activation is used as it

makes the outputs of generator compatible with the inputs of discriminator. A fake sample of dimensions $28 \times 28 \times 3$ is produced (generated image) as shown in Fig. 1.

Discriminator. As shown in Fig. 2, an input image of shape $28 \times 28 \times 3$ is fed to the discriminator. 5×5 filters, strided convolutions, and batch normalization are used in all layers except the first layer. The Leaky-ReLU activations are used in the hidden layers, and the output layer uses the sigmoid activation for predicting the probability, if the input to the discriminator is a fake sample generated from the generator or a real image from the real dataset.

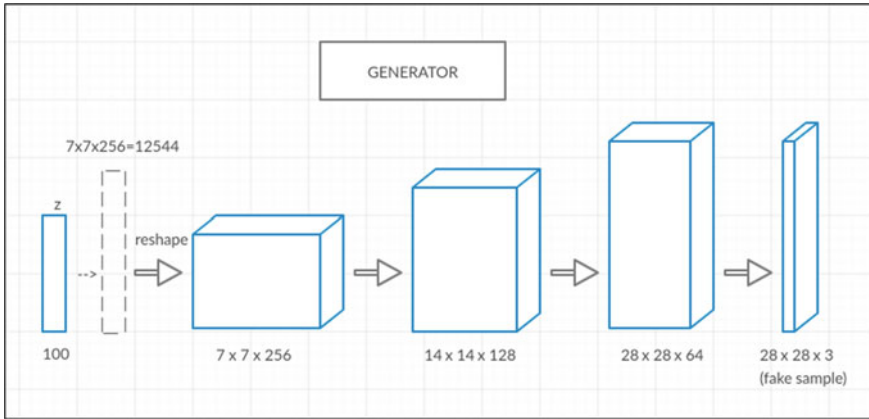


Fig. 1 DCGAN generator

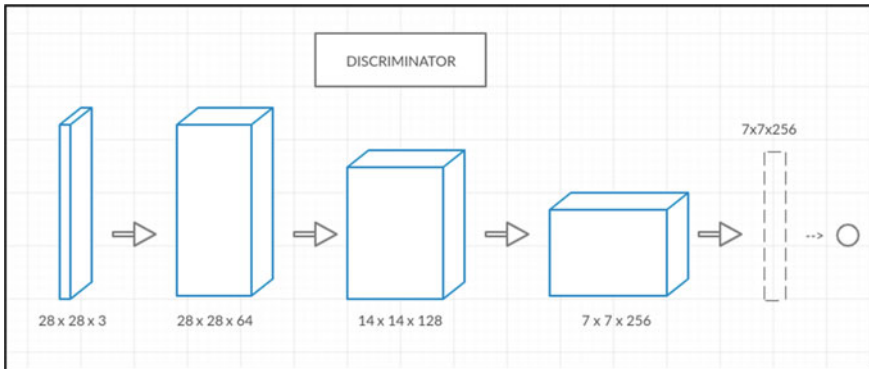


Fig. 2 DCGAN discriminator

3.2 Cost Functions and Hyperparameters

We used the binary cross-entropy loss for both the networks. Following is the discriminator loss:

$$J_D = -\frac{1}{m_D} \sum_{i=1}^{m_D} y_D^{(i)} \log(D(x^{(i)})) - \frac{1}{m_G} \sum_{i=1}^{m_G} (1 - y_G^{(i)}) \log(1 - D(G(z^{(i)})))$$

here, m_D is number of real images, m_G is number of fake images, $y_G^{(i)}$ is the label for fake images which is equal to zero, $y_D^{(i)}$ is label for real images which is equal to one, $D(x^{(i)})$ represents the output of discriminator for real input $x^{(i)}$, and $D(G(z^{(i)}))$ is the output of discriminator for the fake data input $G(z^{(i)})$.

For the generator, we use the non-saturated loss:

$$J_G = -\frac{1}{m_G} \sum_{i=1}^{m_G} \log(D(G(z^{(i)})))$$

Learning rate = 0.0002

Minibatch size = 128

Adam optimizer $\beta_1 = 0.5$

Epochs = 20

3.3 Method

The SVHN train dataset by default contains more than 73 thousand, $32 \times 32 \times 3$ dimensional images. Due to the time and GPU constraints, we reduced the size of the images and resized them to be $28 \times 28 \times 3$ as our training standard. We then first implemented our DCGAN architecture on the simpler MNIST digit dataset containing $28 \times 28 \times 1$ single channel images to ensure that the networks were training correctly. Most of the hyperparameters were kept as suggested in the literature except for some changes in the weight initializers, learning rate, and experiments with dropout [14]. We also utilized the optimizations like: label smoothing which helps reduce overconfidence and better resistant to adversarial examples, Leaky-ReLU in both generator and discriminator is scaling the real images to $(-1$ to $1)$ to match the generated images from tanh activations of the generator before giving

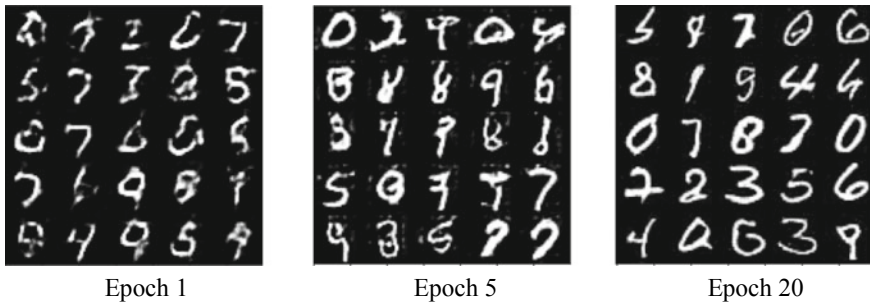


Fig. 3 MNIST 1-channel generated images

them as input to discriminator, forming separate minibatches of real and fake input images to the discriminator rather than mixing them.

Once the model was generating satisfactory results on MNIST dataset, we used the untrained DCGAN model on the SVHN dataset. Following figures show that the model after 20 epochs was able to produce realistic images that were recognizable to human eyes. Refer Fig. 3.

3.4 *Modified Discriminator*

We experimented further by increasing the depth of the discriminator using few residual blocks [13] with skip connections with a view of increasing the accuracy of the discriminator as a classifier to detect fake images. It would help the discriminator learn complex feature map representations, thereby improving the generator and producing better images. The purpose of using residual blocks instead of just stacking new layers to the discriminator was that the residual blocks are observed to be helpful in making multilayer deeper CNN models by avoiding the problems of vanishing and exploding gradients caused in deeper networks. By using it, we were able to add an additional seven layers to the network over the pre-existing four layers, and a few newer minor changes in generator were tried. Following is the modified discriminator in Fig. 4.

3.5 *Inception Score (IS)*

It is not always feasible to practically have humans to annotate, evaluate, and assess the generated images. One of the quantitative metrics which is able to correlate with human judgment is the Inception Score introduced by Salimans et.al [12]. The Inception Score uses a statistical formula called the Kullback–Leibler divergence which is helpful in evaluating how similar/different two probability distributions are. Firstly,

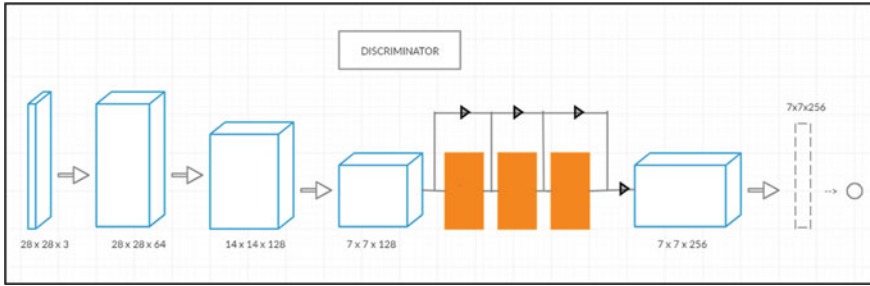


Fig. 4 Modified discriminator with residual blocks

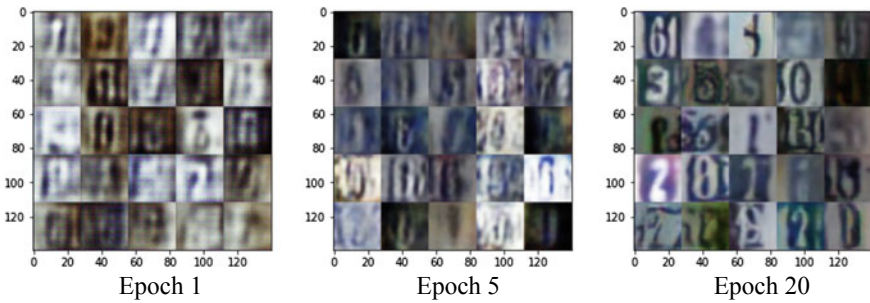


Fig. 5 SVHN 3-channel generated images from DCGAN architecture

the generated images are fed to the Inception model [12], and then, the KL divergence between the conditional label distribution of each image $p(y|x)$ and the marginal distributions $\int p(y|x = G(z))dz$ are calculated [15]. Images with meaningful objects should have conditional distribution with low entropy, and distributions with high variety of images should have marginal distribution with high entropy.

$$IS = \exp(E_x \text{KL}(p(y|2)||p(y)))$$

The two properties, quality and diversity, are captured by Inception Score. Higher Inception Score implies more diverse and realistic images. In our experiment, we generated 200 images after every epoch and gave these images as input to the inception model and calculated the mean Inception Scores. Results are discussed in the later section.

4 Results

The DCGAN architecture was beginning to generate realistic and human recognizable images from the SVHN dataset by the completion of 18–20 epochs. Figure 5 shows some samples taken from training process.

It can be seen from Fig. 6 (left) that the discriminator and generator are competing with each other since the beginning of the training. Whenever the cost of generator got reduced, the cost of discriminator could be seen increasing which meant that it became difficult for the discriminator to differentiate. Conversely whenever the discriminator loss decreased, it meant that it was getting better at classifying the images as real or fake, so the generator loss increased. The generator loss was seen to be decreasing over time which means that the generator was generating better realistic looking outputs with time. Also, it was observed that the discriminator cost did not go to zero (implying that the model did not encounter failure mode [9]) but rather approached toward 0.5 steadily along with a decreasing generator cost. Here, generator cost reduced marginally at around 4000th step from around 7–3.5 (which implied there had not been any convergence failure [9] yet). In our experiments, we generated and stored 200 new images after completion of every epoch. The mean Inception Scores plotted in the graph in Fig. 6 (right) for these newly generated and stored images show that the Inception Score was increasing, which meant that the generation of fake images was getting more realistic with time.

Next, we experimented with modified discriminator of increased layers containing skip connections along with a few changes to generator layers. From Fig. 8 (left), it can be observed that the average discriminator loss was a little lower than 0.5, about 0.25–0.35, but not completely zero and the trend remained the same for the number of epochs under consideration. On the other hand, the generator losses kept fluctuating as seen in the figure. One thing to note was that, though the losses were fluctuating heavily in case of generator, the images were getting better as time passed and were diverse (Fig. 7). The same is evident from the Inception Score plot in Fig. 8 (right) as well. The discriminator probably became much better at

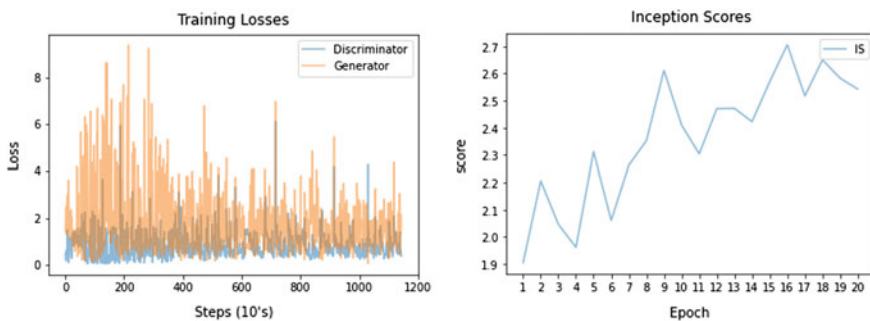


Fig. 6 Left: Loss function for DCGAN architecture. Right: Mean inception score of generated images at every epoch

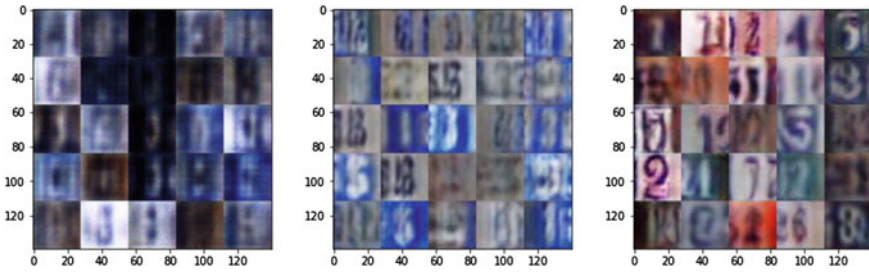


Fig. 7 SVHN 3-channel generated images from the modified discriminator architecture

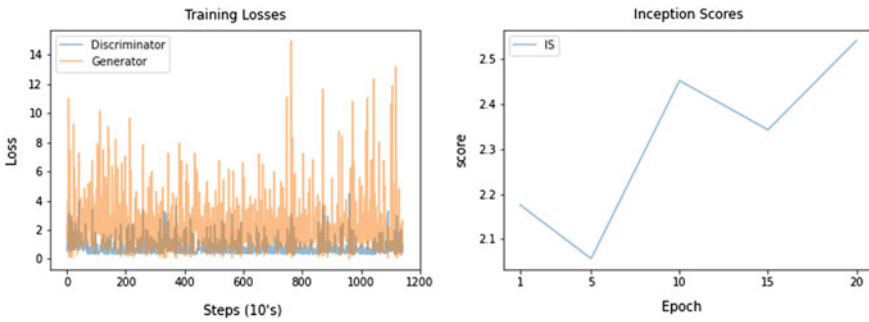


Fig. 8 Left: Loss function for the modified architecture. Right: Mean inception score of generated images after every five epochs

classifying fake images thereby having reduced losses and encouraged the generator to generate counterfeits with time that was more challenging and realistic for the discriminator. Increase in these layers conspicuously increases the computation time and can also cause overfitting, so the dropout regularization was used for few layers along with batch-normalization and some minor hyperparameter tuning, but no significant improvements were found.

5 Conclusion

In this study, we used the DCGAN, a more stable GAN architecture to generate a training dataset of images of digits. The images appeared to be more realistic as the training progressed. We then experimented with the discriminator by increasing its depth by adding skip connections and residual blocks. We were able to train the DCGAN without failure modes by using the suggested optimizations. Mode collapse and convergence failure were avoided as the images had diversity, and the generator and discriminator errors as seen in Fig. 6 (left) were steadily converging.

In the future, we wish to explore the effects of virtual batch normalization and feature extraction and training for longer time as well as using conditional GANs for generating pre-specified types of images. Along with this, we can use semi-supervised learning [12] to increase the accuracy of digit classifiers using generative models by using the discriminator as a digit classifier and training it on a dataset formed by the combination of limited labeled real images, unlabeled real images, and realistic looking, fake images generated from the generator.

References

1. Goodfellow I, Pouget-Abadie J, Mirza M et al (2014) Generative adversarial nets. In: NIPS
2. Michael C, Kang S, Kim S (2017) (Re)live photos: generating videos with GANs. Stanford CS231N course project
3. Ledig C et al (2016) Photo-realistic single image super-resolution using a generative adversarial network. [arXiv:1609.04802](https://arxiv.org/abs/1609.04802)
4. Reed S et al (2016) Generative adversarial text to image synthesis. [arXiv:1605.05396](https://arxiv.org/abs/1605.05396)
5. Ma et al (2018) Pose guided person image generation. [arXiv:1705.09368](https://arxiv.org/abs/1705.09368)
6. Dong et al (2017) MuseGAN: multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. [arXiv:1709.06298](https://arxiv.org/abs/1709.06298)
7. Baur C et al (2018) GANs for medical image analysis. [arXiv:1809.06222](https://arxiv.org/abs/1809.06222)
8. Ng E et al (2018) Image colorization using generative adversarial networks. [arXiv:1803.05400](https://arxiv.org/abs/1803.05400)
9. <https://machinelearningmastery.com/practical-guide-to-gan-failure-modes/>
10. Barua S et al (2019). FCC-GAN: a fully connected and convolutional net architecture for GANs. [arXiv:1905.02417](https://arxiv.org/abs/1905.02417)
11. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. [arXiv preprint arXiv:1511.06434](https://arxiv.org/abs/1511.06434)
12. Salimans T et al (2016) Improved techniques for training gans. In: Advances in neural information processing systems. pp 2234–2242
13. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. [arXiv:1512.03385](https://arxiv.org/abs/1512.03385)
14. <https://github.com/soumith/ganhacks>
15. <https://medium.com/octavian-ai/a-simple-explanation-of-the-inception-score-372dff6a8c7a>

A Novel Proof of Concept for Twitter Analytics Using Popular Hashtags: Experimentation and Evaluation



Kiran Ahire, Manali Bagul, Swapnil Dhanawate, and Suja Sreejith Panicker

Abstract Twitter analytics is a classic research area especially with the widespread presence of Big Data in various online media such as—social network sites, online portals for shopping, e-commerce, forums, chats, recommendation systems, and online services. Ascertaining the sentiment behind, the various types of tweets by different persons can provide great insights on various aspects including behavioral patterns. Besides highlighting the newest trends in the field, we retrieved real-time twitter data pertaining to three currently popular hashtags in the Indian context and carried out extensive experimentation analysis about the prevailing sentiment of a strata of population. Inclusion of current challenges, future trends and applications of sentiment analysis from Twitter data makes this novel work very useful for fellow researchers.

Keywords Machine learning · Natural language processing · Sentiment analysis · Twitter data analytics · Opinion mining · Data visualization · Recommendation systems · Attitude analysis · Polarity determination · Sentiment classification

K. Ahire · S. Dhanawate
Department of Computer Engineering, Maharashtra Institute of Technology, Pune, Maharashtra, India
e-mail: Ahirekiran930@gmail.com

S. Dhanawate
e-mail: swapnildhanawate@gmail.com

M. Bagul · S. S. Panicker (✉)
School of Computer Engineering and Technology, MIT World Peace University, Pune, Maharashtra, India
e-mail: suja.panickar@mitwpu.edu.in

M. Bagul
e-mail: manalibagul123@gmail.com

1 Introduction

Analytics of twitter data is predominantly conducted to ascertain the underlying sentiment behind the tweets/images. Hence, this effectively narrows to sentiment analysis. Sentiment analysis is computational study of the various opinions, emotions, sentiments, and attitude which is expressed by different users in the form of texts pertaining to an entity of interest. Sentiment analysis is also called as review mining, opinion mining, or attitude analysis [1–5].

Motivation for the surge in voluminous user content globally is attributed to technological advancements as also increased Internet activities like—discussion forums, conferencing, online transactions, e-commerce, chatting, surveillances, ticket booking, websites of merchants, widespread and continual communications on various social media, and the variety of other online activities [1, 3, 6, 7].

Current work is organized as follows: Sect. 2 covers motivation, Sect. 3 covers literature survey, Sect. 4 covers experimentation and results, Sect. 5 presents observations, Sect. 6 highlights the novelty, Sect. 7 presents the various applications, Sect. 8 presents challenges, Sect. 9 presents research contribution, and Sect. 10 covers conclusion.

2 Motivation

This novel technique will help people to analyze various data from Twitter and help understand the public opinion or sentiment of people behind the specific keywords, and this will be useful in various sectors like business, marketing, forecasting, politics, and tourism.

3 Literature Survey

There are mainly two approaches found in existing literature [1, 7–18] for performing sentiment analysis—lexicon based and machine learning based. Concept of polarity is used in the former while suitable classification models are developed in the latter.

Detailed survey of recent work is presented in Table 1, and research gaps are highlighted.

4 Experimentation and Results

We used Tweepy to fetch the tweets in real time for three currently popular hashtags in India: #MakeInIndia, #AtmNirbharIndia, #VocalforLocal. The `tweepy.Cursor()`

Table 1 Survey of sentiment analysis in recent works

Ref. No.	Year	Machine learning algorithm	Datasets used	Result	Research gap
[18]	2020	NLP, Decision Tree, SVM, Random Forest	Collecting datasets (tweets—40,000)	Highest accuracy 99.4 obtained with Random Forest	SVM technique having less accuracy for analysis
[19]	2020	Deep recurrent neural network classifier	Real data	Highest accuracy of 93% obtained with Hadoop-based deep RNN	Adding more features in the feature extraction process
[20]	2020	Valence Aware Dictionary and Sentiment Reasoner	Geo tweet Harvard Center for Geographic Analysis	Overall posted positive tweets are increasing from Monday to Sunday and peaks on Sunday	Data noise, Advanced bots detection methods should be employed to refine dataset, Negative sentiment analysis
[21]	2020	NLP, Map visualization	Real data Kumamoto Earthquake	Map visualization with statistical relevancy index	To gather more exhaustive requirements for each area and the information become more reliable
[22]	2020	SVM	Real data	Negative and positive comments	Improve the precision in feeling mining in python structure
[23]	2020	Feature ensemble model, GloVe Model	Real data	Proposed method had F1 score of 0.81	Apply Parsimonious Extreme Machine Learning
[24]	2020	Logistic Regression Classifier algorithm, SVM	Stanford dataset, Sanders Twitter sentiment corpus dataset	Logistic Regression with highest accuracy of 91%	Combine emergency detection and polar sentiment analysis. Detect public health emergencies such as 2019 nCov
[25]	2020	DNN, Naïve Bayes	Kaggle dataset, Twitter streaming API	11.2% positive for females, 12.65% positive for males	Producing new database that will contribute to further studies in this subject

How people are reacting on MakeInIndia by analyzing 1000 Tweets.

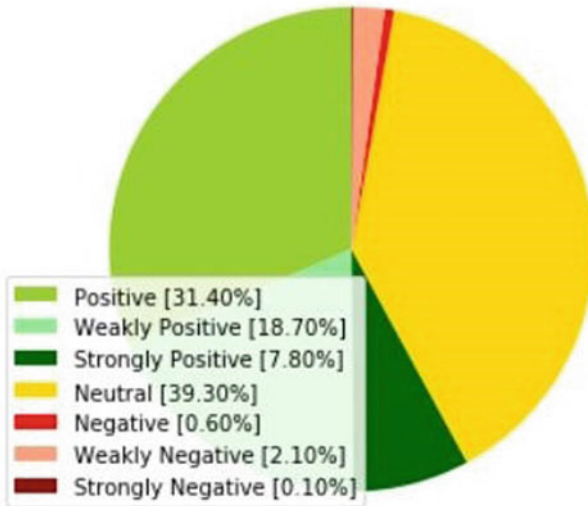


Fig. 1 Analysis of 1000 tweets for #MakeInIndia

function was used to fetch all latest tweets. Preprocessing was performed using the ‘re’ library of python. TextBlob was used for polarity determination. We wrote a python program to encode the seven class labels as follows: -1 negative, -0.6 to -1 strongly negative, 0 to -0.3 weakly negative, 0 neutral, 0 to 0.3 weakly positive, 0.6 to 1 strongly positive, and 1 for positive and performed three experiments as under.

4.1 Experiment 1: #MakeInIndia

We fetched 1000 tweets in real time and have analyzed the same for ascertaining the sentiment. Visualization results for seven sentiment classes are as illustrated in Fig. 1

4.2 Experiment 2: #AtmNirbhar

We fetched 1000 tweets in real time and have analyzed the same for ascertaining the sentiment. Visualization results are as illustrated in Fig. 2

How people are reacting on Atmnirbhar by analyzing 1000 Tweets.

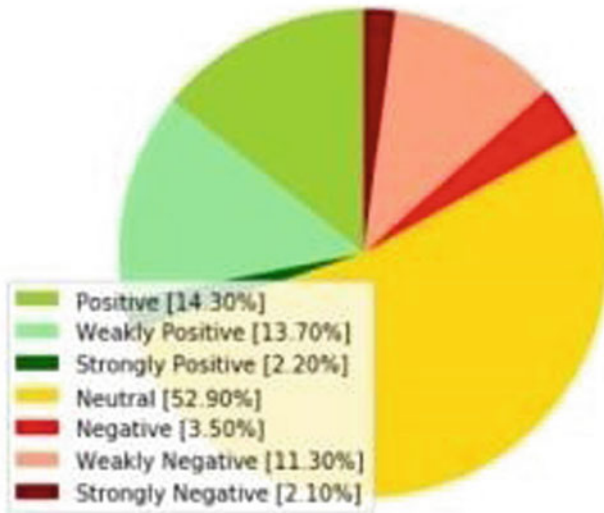


Fig. 2 Analysis of 1000 tweets for #AtmNirbhar

4.3 Experiment 3: #VocalforLocal

Figure 3 illustrates the outcome of analyzing 200 tweets.

We performed comparative analysis of the two hashtags with respect to seven sentiment classes as illustrated in the stacked bar chart in Fig. 4

To validate the obtained results, we assigned the task of annotation to two human experts and noted the findings. Figures 5 and 6 illustrate the differences in annotation between the two experts using RMSE and standard deviation, respectively.

5 Observations

- From Figs. 1, 2, 3 and 4, we infer that the highest positive percentage of tweets was for #MakeInIndia while the highest negative tweets were for #Atmnirbhar
- From Table 1, it is observed that although some standard datasets do exist, most researchers prefer to gather tweets in real time. Tweepy was observed to be the predominant choice. Also, SVM and Random Forests have frequently yielded high accuracy of over 95%

How people are reacting on VocalForLocal by analyzing 200 Tweets.

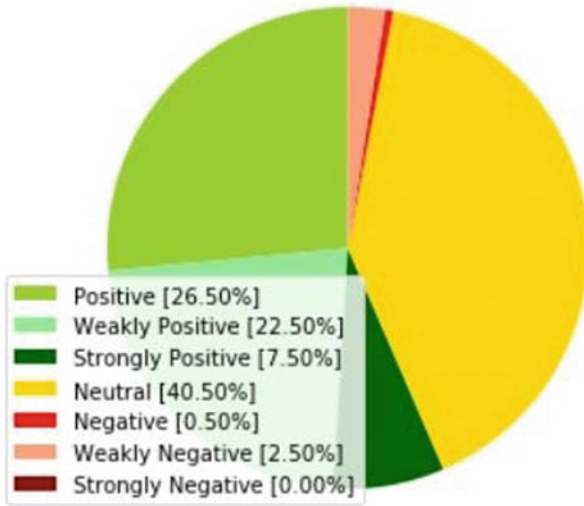


Fig. 3 Analysis of 200 tweets for #MakeInIndia

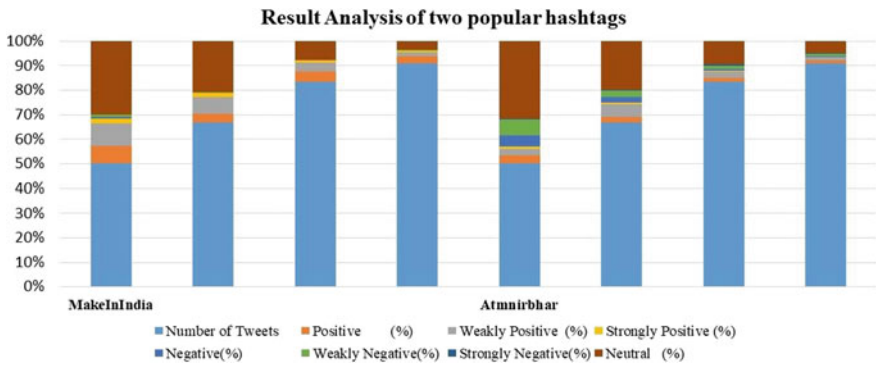


Fig. 4 Comparative analysis of two hashtags with respect to seven sentiment classes

6 Novelty

This technique gives the result visualization in the form of pie-chart along with seven classes which gives the clearer idea about the sentiment behind keyword, and this novel approach of result visualization helps people to understand result in detail.

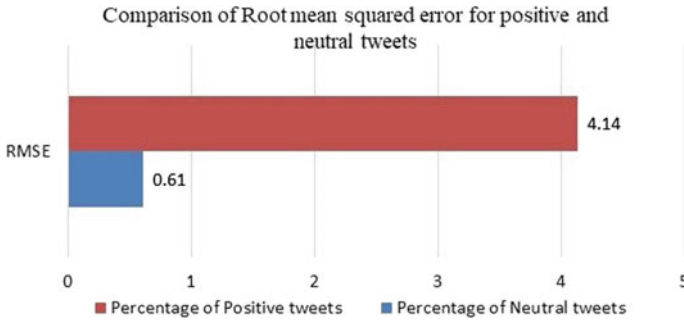


Fig. 5 Differences in RMSE values for the two human experts

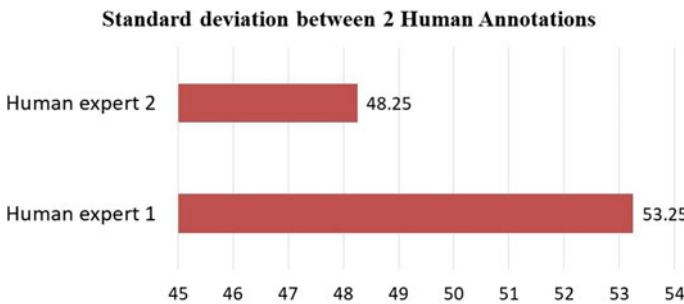


Fig. 6 Inter-individual differences in annotation by two human experts

7 Applications

Twitter data analytics has variety of applications such as

- For generating reputation for brands or products [26–28],
- For increasing the customer engagement, having better informed decisions toward risk analysis, efficient credit ratings for various customers, and performing competitive analysis [29],
- Increasing productivity and efficiency of restaurants [30],
- For better market intelligence and improve customer satisfaction [3, 36],
- Increased tourism [37],
- Monitoring and analyzing public opinions concerning political issues [3],
- To forecast the price changes as per news sentiments [1],
- To develop new products, services and promote products as per the customers reviews [1] and social advertising [38, 39].

8 Challenges in Twitter Data Analytics

- i. Determining the contextual information for sentiments and forming a generalized foundation globally is difficult [30].
- ii. There is increased difficulty due to the widespread use of onomatopoeias, idioms, homophones, alliterations, and acronyms [30]. Hence, complex NLP techniques are required to decipher the correct context and meaning of various words.
- iii. Aspect-based sentiment analysis is an important challenge [36].
- iv. Opinion summarization, subjectivity classification, and opinion retrieval [36]
- v. Lack of large annotated data to train models across various domains [40]

9 Research Contribution

- Current work is a novel approach of visualizing and analyzing the three currently popular hashtags in India. Our extensive experimentation and analysis about the prevailing sentiment shall be greatly beneficial for fellow researchers.
- We have also covered important aspects such as—current challenges, future trends, and applications of sentiment analysis.

10 Conclusion

We have successfully implemented the proof of concept toward gathering tweets in real time and attempting to analyze the sentiment of a part of population using the lexicon-based technique. We have performed extensive experimentation and analyzed the sentiments for 100, 200, 500, and 1000 different sets of tweets for three most currently most popular hashtags. Ample data visualization performed in this work would be great asset to fellow researchers thereby carving the path for future research.

References

1. Kumar R, Vadlamani R (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge Based Systems*
2. Liu B (2010) Sentiment analysis: a multi-faceted problem. *IEEE Intell Syst* 25(3):76–80
3. Tubishat M, Idris N, Abushariah MAM (2018) Implicit aspect extraction in sentiment analysis: review, taxonomy, opportunities, and open challenges. *Inf Process Manage* 54:545–563
4. Montoyo A, Martínez-Barco P, Balahur A (2012) Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments. *Decis Support Syst* 53:675–679
5. Kumar S, Morstatter F, Liu H (2013) *Twitter data analytics*. Springer

6. Liu J (2008) Opinion spam and analysis. In: Proceedings of the international conference on web search and web data mining, ACM
7. Tan LKW, Na JC, Theng YL et al (2012) Phrase-level sentiment polarity classification using rule-based typed dependencies and additional phrases consideration. *J Comput Sci Technol* 27(3):650–666
8. Wang T et al (2014) Product aspect extraction supervised with online domain knowledge. *Knowledge-Based Syst* 71:86–100
9. Kunte AV, Panicker S (2020) Analysis of machine learning algorithms for predicting personality: brief survey and experimentation. In: 2019 global conference for advancement in technology (GCAT)
10. Kunte A, Panicker S (2020) Personality prediction of social network users using ensemble and XGBoost. In: Das H, Pattnaik P, Rautaray S, Li KC (eds) *Progress in computing, analytics and networking. Advances in intelligent systems and computing*, vol 1119. Springer, Singapore
11. Kunte AV, Panicker SS (2019) Using textual data for personality prediction: a machine learning approach. In: 2019 4th international conference on information systems and computer networks (ISCON)
12. Panicker S, Kunte A (2019) Personality prediction using social media. In: 2019 5th international conference for convergence in technology (I2CT), Pune (in Press)
13. Mane VL, Panicker SS (2015) Knowledge discovery from user health posts. In: IEEE 9th international conference on intelligent systems and control (ISCO)
14. Mane VL, Panicker SS (2015) Summarization and sentiment analysis from user health posts. In: 2015 international conference on pervasive computing (ICPC). IEEE
15. Salunke V, Panicker SS (2021) Image sentiment analysis using deep learning. In: Ranganathan G, Chen J, Rocha A (eds) *Inventive communication and computational technologies. Lecture notes in networks and systems*, vol 145. Springer, Singapore
16. Dangra BS, Rajput D, Bedekar MV, Panicker SS (2015) Profiling of automobile drivers using car games. In: International conference on pervasive computing (ICPC). IEEE
17. Bedekar MV, Atote B, Zahoor S, Panicker S (2016) Proposed used of information DisPersal Algorithm in user profiling. In: International conference on ICT for sustainable development, Goa, India
18. Khan M, Malviya A (2020) Big data approach for sentiment analysis of twitter data using Hadoop framework and deep learning. In: 2020 international conference on emerging trends in information technology and engineering (ic-ETITE), Vellore, India
19. Hu T, She B, Duan L, Yue H, Clunis J (2020) A systematic spatial and temporal sentiment analysis on geo-tweets. *IEEE Access* 8, 8658–8667
20. Murakami A, Nasukawa T, Watanabe K, Hatayama M (2020) Understanding requirements and issues in disaster area using geotemporal visualization of Twitter analysis. *IBM J Res Develop*
21. Kumar TS, Nabeem PM, Manoj CK, Jeyachandran K (2020) Sentimental analysis (opinion mining) in social network by using SVM algorithm. In: 2020 fourth international conference on computing methodologies and communication (ICCMC), Erode, India
22. Phan HT, Tran VC, Nguyen NT, Hwang D (2020) Improving the performance of sentiment analysis of Tweets containing fuzzy sentiment using the feature ensemble model. In: *IEEE Access*, vol 8, pp 14630–14641
23. Bhatnagar D, SubaLakshmi RJ, Vanmathi C (2020) Twitter Sentiment Analysis Using Elasticsearch, LOGSTASH And KIBANA. In: 2020 international conference on emerging trends in information technology and engineering (ic-ETITE)
24. Oyasor J, Raborife M, Ranchod P (2020) Sentiment analysis as an indicator to evaluate gender disparity on sexual violence tweets in South Africa. In: 2020 international SAUPEC/RobMech/PRASA conference, Cape Town, South Africa
25. Joshi PA, Simon G, Murumkar YP (2018) Generation of brand/product reputation using Twitter data. In: 2018 international conference on information, communication, engineering and technology (ICICET), Pune
26. Wang W, Li B, Feng D, Zhang A, Wan S (2020) The OL-DAWE model: tweet polarity sentiment analysis with data augmentation. In: *IEEE Access*, vol 8, pp 40118–40128

27. Li YM, Shiu YL (2012) A diffusion mechanism for social advertising over microblogs. *Decis Support Syst* 54:9–22
28. Du J et al (2013) Box office prediction based on microblog. In: *Expert systems with applications*
29. Al-Moslmi T, Omar N, Abdullah S, Albared M (2017) Approaches to cross-domain sentiment analysis: a systematic literature review. *IEEE Access* 5:16173–16192
30. Li SK, Guan Z, Tang LY et al (2012) Exploiting consumer reviews for product feature ranking. *J Comput Sci Technol* 27(3):635–649

Investigating the Role of Machine Learning in Detecting Psychological Tension



Suja Sreejith Panicker and P. Gayathri

Abstract Psychological tension is a growing concern worldwide and has gradually victimized individuals across differing age groups, gender and nationality globally. Despite the omnipresence of technology, especially in the field of healthcare, psychological tension continues to be a powerful and widespread disorder, implications of which manifest in the form of varied physical and mental ailments. This paper presents a detailed review regarding the role and efficiency of popularly used machine learning algorithms such as Bayes, SVM, ANN, kNN, random forests in determining psychological tension. Systematic analysis of the physiological features, their thresholds and the scenario in question leads to successful classification of tension as low, medium or high. Knowledge-based systems that could effectively diagnose psychological tension with scientific quantification techniques shall be immensely useful in studying human affect and also successfully mitigate tension/strain by promoting its early automated/semi-automated detection, thus largely contributing to mankind.

Keywords Machine learning · Psychological tension · Data mining · Knowledge-based systems · Health care · Human mental stress · Anxiety detection

Abbreviations

ANN	Artificial neural network
CNN	Convolutional neural network
DBN	Deep belief network
ECG	Electrocardiogram
EDA	Electrodermal activity

S. S. Panicker (✉) · P. Gayathri
School of Computer Science and Engineering (SCOPE), VIT University, Tamil Nadu, Vellore
632014, India
e-mail: suja.panickar@mitwpu.edu.in

P. Gayathri
e-mail: pgayathri@vit.ac.in

ELM	Extreme learning machine
EMG	Electromyography
fNIRS	functional near-infrared spectroscopy
HRV	Heart rate variability
IoT	Internet of things
kNN	k nearest neighbour
PCA	Principal component analysis
PPG	Postprandial plasma glucose
SVM	Support vector machine

1 Introduction

Psychological tension is an important precursor to varied physical and mental ailments affecting the global population annually. With an evident impact on emotional health and physical health, it is crucial to minimize the effects of mental tension/strain with effective automated/semi-automated systems that can aid the medical fraternity with early detection. Needless to say, mental tension is a growing concern worldwide and has gradually victimized individuals across age groups, gender and nationality. Despite the omnipresence of technology, especially in the field of healthcare, it continues to be a powerful and widespread disorder implication of which manifest in the form of varied ailments ranging from milder ones, to the chronic and severe ones resulting in fatalities. Building knowledge-based systems for detecting and managing psychological tension is an amalgamation of health sciences and machine learning.

Mental stress is the clinical term analogous to terms such as strain, tension. Among the various definitions of stress reviewed in the classic work [1], a noteworthy one is—Stress is a relationship between an individual and the surroundings, and it is evaluated by the individual as tedious, thereby endangering either the life or safety of the concerned individual [2].

This paper is organized as under: Sect. 2 covers literature survey, Sect. 3 covers results and discussion, Sect. 4 covers gaps, Sect. 5 presents issues, Sect. 6 covers novel case studies, Sect. 7 highlights the various challenges and future scope in this field, Sect. 8 presents research contribution, Sect. 9 outlines proposed work, and Sect. 10 presents conclusion.

2 Literature Survey

This sections presents the survey in two sub-sections—survey of various methodologies in assessing psychological tension, followed by role of IoT in psychological tension detection systems, and lastly, we sum up role of machine learning in psychological tension detection systems.

2.1 *An Overview of Various Methodologies in Assessing Psychological Tension.*

As per Lazarus [2], the very perception and presence of stress is a pure subjective assessment performed by the individual himself. On surveying past works on stress detection, we observed the widespread use of conventional assessment methodology by questionnaires as opposed to the evidential physiological assessment of the person's biological features such as—heart rate, respiration rate, blood volume pressure, skin temperature, etc. Despite their popularity, an important limitation of such self-assessment techniques is that people with more awareness of mindfulness may tend to go for comparatively lower scores and the opposite is true for people with less awareness. This is a major challenge regarding the accuracy of self-reported questionnaires.

2.2 *Overview of IoT in Psychological Tension Detection Systems*

In a research carried out on six individuals, PPG signals from wrist and ECG signals from chest were acquired in real time. Promising results of 98% accuracy of detecting the mental stress within individuals and F1 score of 80 were obtained.

It was noted that the HRV parameters illustrated significant differences in the stress and non-stress states [3]. In a novel work [4] a personalized smartphone keyboard (android based) was developed to track the multiple human emotions of users in real time, by leveraging on the entry pattern of text. Results on 22 human subjects yielded 78% accuracy [4].

Various SVM models were tested in [5] by experimenting with varied ECG features and SVM kernels. Highest accuracy of 98.6% was obtained with gaussian function. This work emphasized the significance of respiratory information in detecting human stress [5]. In a study on 20 human subjects, CNN was compared with six traditional methods based on HRV. Novelty was use of super short window of 10 s. Results on CNN indicated an improvement of 7.2%. Results are promising and highlight CNNs contributions in this field [6]. fNIRS were used in a study to compare shallow models vs deep models. As against the 64.74% accuracy with SVM

and 71.13% accuracy with AdaBoost; CNN and DBN yielded much higher accuracies of 72.77 and 84.26% [7]. A wireless sensor model which recorded neural and physiological signals from human subjects in a real-time setup is proposed in [8]. It was noted that the changes in brain signals correlated with intense induced stress [8]. In [9], a wearable system was proposed with lower power Bluetooth for collecting the physiological data over longer time spans. A smartphone was also used to collect the sensors data (applied on body). An accuracy of 80% was noted [9].

In [10], an unique multimodal dataset was developed for stress management by combining different modalities such as physiological, expressions on the face, information about the body posture and behavioral information. Besides the conventional detection of stress by physiological parameters, there are studies that analyzed the spectral features and processed the images obtained from cameras [11] for detecting stress levels of individuals.

Advantages of wearables include small size, cost effectiveness, user friendliness [12] and uninterrupted monitoring for sufficiently long time span [13]. Apart from the evident advantages of wearables, there are some challenges such as they should be portable, consume less energy, have less weight and should not cause any injury/discomfort to the wearer [14] and privacy issues [11].

The use of smartphones as a part of daily life has increased enormously since the last decade and this growing pervasiveness is a motivating factor for research on effective stress detection [15]. It is now possible to assess the individual's stress by taking into consideration user-specific behavior between individual and the smartphone [11]. An important limitation is the restricted battery capacity which may directly hamper the long-term monitoring of subjects [16, 17].

We have summarized the findings of popular methodologies (with predominant use of questionnaires) coupled with statistical or machine learning techniques in Table 1.

We have summarized the findings of popular machine learning classifiers in Table 2.

It is noted [37–39] that presence of stress leads to the onset of several ailments, mental or physical. Hence, there is strong need to curb the effects of stress globally.

3 Results and Discussion

- The total number of subjects considered for experimentation is an important marker for generalization ability of the stress detection system. It is observed from Table 1 that considering a sufficiently large sample size is important, but this is a common impediment in many works focusing on real data. Majority works have experimented with less than 50 human subjects.
- It is noted that strict laboratory-controlled setups have been used for eliciting the required emotion/response from participants, which is a major setback toward

Table 1 Survey of existing literature depicting popular methodologies and research gap

Ref. No.	Methodology used	Number of human subjects in experiment	Dataset used	Results achieved	Research gap
[18]	Questionnaire and Support vector machine	11	Real data	Highest recognition was 81.82% for the emotion label of joy	To consider more supervised learning algorithms and to include EMG signals
[19]	kNN, decision tree, random forest, naïve Bayes, SVM and multilayer perceptron	30	Real data	General model showed lower recognition than the group model	To use physiological and facial expression information
[20]	PCA, SVM and ELM	15	HAPPEI is collected from Flickr	Global and local attributes for theme expression analysis	To incorporate more parameters such as gender, age
[21]	Questionnaires and SVM	21	Real data	Physiological features showed gold performance (mean F1 score)	More number of subjects required for validation
[22]	Questionnaire	21	Real data	Accuracy of 88.1% with the automatic classifier	Cannot operate in real-time, should be tested on lie detection
[23]	Questionnaires, sparse Bayesian learning, PCA, kernel-based classifiers	14	Real data	Accuracy over 99%	More number of subjects required for validation
[11]	Questionnaires, naïve Bayes and Decision Tree	30	Real data	71% accuracy For user-specific models	Increased accuracy needed

(continued)

Table 1 (continued)

Ref. No.	Methodology used	Number of human subjects in experiment	Dataset used	Results achieved	Research gap
[24]	Empatica E4 wristband, median filter, Savitsky–Golay filter, butterworth filter	10	Real data	More than 87% overall accuracy	More number of subjects required for validation
[25]	Questionnaires and statistical analysis	39	Real data collection	Marked improvement with the use of clustering	Only two clusters were considered. Testing to be done with more number of clusters
[26]	Questionnaires and t tests	46	Real data	Decreased stress and improved coping strategies observed	Precise results needed
[27]	Questionnaires and feature selection	32	Real data	91.93% in Distinguishing strong stress and weak stress state	More validation needed with larger sample size
[28]	Naive Bayes, MLP, 1R rule, Nearest neighbour, Bagging, AdaBoost and SVM	166	Real data collection	F1 score—0.984 (for high level)	To implement for people with disabilities
[29]	Random forest, kNN, SVM, LDA, AdaBoost	39	Real time data	Highest accuracy of 94% with SVM	To provide mechanisms for users to be able to visualize stress patterns on the mobile devices
[30]	Questionnaires	82	Real data	Reduction in mind wandering	Small sample size, difficulty in accessing the proportion of mind wandering

Table 2 Summary of machine learning techniques used in the field of stress detection

Ref. No	Machine learning technique	Attributes extracted	Class labels used	Results obtained	Gap found
[31]	Convolutional long short-term memory network, constrained independent component analysis algorithm	EEG, PPG	Positive, negative	15.96% increase in accuracy	<ul style="list-style-type: none"> To learn the time-based physiological patterns in extracting emotions To extend the same To promote early detection of emotions in the real world scenario
[32]	Fuzzy logic, random forest, svm, decision-tree, expectation maximization	ECG, EEG, oxygen saturation, blood pressure and respirationrate; Environmental Parameters: loction, time, temperature, weather	normal, low, middle and high stress level	90% accuracy with SVM	<ul style="list-style-type: none"> Use of modern biosensors for obtaining increased accuracy Customized smart services aimed at promoting Mental wellness
[33]	SVM	Electrocardiogram, skin temperature variation and electro dermal activity	Sadness, anger, stress, surprise	78.4% and 61.8% respectively for recognition of three and four categories	<ul style="list-style-type: none"> Increasing sample size Improving classification results
[34]	logistic regression model	EDA, heart rate variability	Mental stress and relaxation	81%	<ul style="list-style-type: none"> To reduce the time spent in manual calibration phase
[35]	SVM, kNN, forward feature selection and PCA	GSR, ECG, Respiration rate, blood pressure, blood oximeter	No stress, stress	High accuracy of 95.8% by the individualized model	<ul style="list-style-type: none"> Extending the current work with more physiological parameters Promote user friendly interface Experiment with a larger sample size for better generalization of results Considering more events that induce stress To successfully detect the various progressive levels of stress

(continued)

Table 2 (continued)

Ref. No	Machine learning technique	Attributes extracted	Class labels used	Results obtained	Gap found
[36]	Principal component analysis, logistic regression	Neural activity, EDA	Concentrated, tension, tired and relaxed	3% rise as compared to previous work	<ul style="list-style-type: none"> • To increase the sample size • To validate the system with leave one out method • To detect more parameters such as driver's age, experience, medications • To use other data such as the exerted pressure on brakes/accelerator to detect the driver's emotion

real-time implementation. Also, important aspects such as gender ration, demographics and other such details about participants are not readily shred in existing works.

- From Table 1, it is noted that validating the proposed systems with more machine learning algorithms is desirable for robustness. Several works have focused on only a small number of classifiers.
- From the survey, it is observed that stress assessment has been predominantly performed with self-reported questionnaires. This can be enhanced by verification from a teacher signal for better reliability of results.
- From Table 2, it is noted that SVM has been popularly used, and the highest classification results of over 90% have been obtained with SVM.

4 Gaps

- Predominant use of self-assessment by getting questionnaires filled by the participants for evaluating their mental stress has important lacunae of ground truth.
- Since several researchers use real-time data, which is not public, there is a prevalent issue of benchmarking in this field.
- Lack of a holistic system that considers various modalities such as physiology, questionnaires, biometric features review, environmental features. Such a model would be more robust and generalized in nature.
- A generalized study that is validated on comparatively larger number of samples.
- These can be addressed in future works, preferably with multimodal holistic frameworks.

5 Issues

We summarize some key contemporary issues that need to be addressed.

- Environmental reasons such as erratic traffic, high levels of noise, bad weather, etc. [40].
- Social reasons such as family issues or problems pertaining to friends/relatives/colleagues, financial problems) [40].
- From mega events such as planning a wedding, presenting to a large audience etc. [40].
- High accessibility as well as heightened expectations of duty, especially for information technology employees [40, 41].
- Personal satisfaction, employment stress and administrative impacts [42].
- Irregular cycles of sleep–wake are common reasons for stress among students/academicians [42].
- Daytime sleepiness, negative attitude, inappropriate sleep duration, anxiety, depression [42].

6 Case Study

- i. As a case study of detecting psychological tension with machine learning in real time, we highlight a novel study [43] conducted on students. As opposed to the situation existing a decade back, today's youth face additional stressful situations / triggers such as depression, suicidal tendencies, physical ailments such as heart attack, stroke [43], diabetes. Increased competition and ever changing lifestyles, with decline in quality of life, could be underlying reasons for the same.
- ii. With a research objective to analyze the mental stress in these students, experiment was conducted wherein mental stress was calculated and compared at two events—7 days before exams commenced and during internet usage. Besides this real data, experiment was conducted on standard dataset—Jaypee Institute of Information Technology. Linear regression, random forest, naïve Bayes, SVM were used for classification of data, and results were observed. The highest accuracy of 85.71% was achieved with SVM. Lower accuracy could be attributed to presence of noise and outliers in real data.
- iii. Another interesting case study [44] emphasizes on real-life driving, which gets very taxing owing to overuse/strain of sensory, visual, motor inputs which normally work harmoniously for a rewarding driving experience. Negative emotions such as frustration, anxiety, anger can be potentially dangerous as it may endanger the precious lives of several innocent persons on the road. With a research objective to reduce the risk of accidents, EEG was extracted in real-time and valence-arousal model was developed for classifying the driver's

- emotion. Additionally, the control of emotions is obtained via the concept of music therapy.
- iv. Other novel applications of detecting psychological stress / negative traits from wide strata of population via real-time twitter analytics and machine learning has been performed in [45–48].

7 Challenges and Future Scope [49]

Challenges in developing a real-time knowledge-based system for determination of human stress include

- Influence of any physical activity (intentional or involuntary) by the human subject during data acquisition phase.
- The presence of motion artifacts contributing to lowered signal quality.
- Lack of a global generalized stress reference.

Future scope in this field includes:

- Use of larger sample size for generalization ability of classifiers.
- Use of longitudinal trials from across the various population groups.
- Proposing timely interventions for successful disease prevention as well as interception.

8 Research Contribution

- i. A comprehensive survey on automated/semi-automated knowledge-based systems for detecting mental stress has been conducted in this paper, with insights on traditional questionnaire based methods as well as the efficient machine learning techniques.
- ii. This work shall be tremendously useful to fellow and future researchers to achieve thorough insights of the field and also motivate to address the current research gaps in literature.

9 Proposed Work

Design and development of a generic and scientific system for monitoring and evaluating human stress states for effective knowledge discovery. Experimentation shall be performed on larger samples to reflect good representative power.

10 Conclusion

Mental stress negatively affects one's biological parameters and also impacts the state of emotional health thus directly affecting overall wellness of individuals. The need for systematic, efficient, reliable and robust stress detection systems is hence crucial for mankind. In this paper, we have carried out extensive literature review with respect to the various machine learning techniques used in stress detection, along with research gaps thereof. It was noted that SVM has been popularly used besides kNN and ANN, which have also yielded high accuracy of over 90%.

References

1. Panicker SS, Prakasam G (2019) A survey of machine learning techniques in physiology based mental stress detection systems. *Biocybern Biomed Eng* 39(2):444–469
2. Lazarus RS, Folkman S (1984) *Stress, appraisal, and coping*. Springer, New York
3. Chen C, Chunhung L, Tsai CW, Deng X (2019) Evaluation of mental stress and heart rate variability derived from wrist-based photoplethysmography. In: 2019 IEEE Eurasia conference on biomedical engineering, healthcare and sustainability (IEEE ECBIOS 2019)
4. Ghosh S, Sahu S, Ganguly N, Mitra B, De P (2019) EmoKey an emotion-aware smart-phone keyboard for mental health monitoring. In: 11th IEEE international conference on communication systems and networks
5. Rizwan MF, Farhad R, Mashuk F, Islam F, Imam MH (2019) Design of a biosignal based stress detection system using machine learning techniques. In: 2019 international conference on robotics, electrical and signal processing techniques (ICREST)
6. He J, Li K, Liao X, Zhang P, Jiang N (2019) Real-time detection of acute cognitive stress using a convolutional neural network from electrocardiographic signal. *IEEE Access* 7
7. Ho TKK, Gwak J, Park CM, Song JI (2019) Discrimination of mental work load levels from multi-channel fNIRS using deep leaning-based approaches. *IEEE Access* 7
8. Masood K, Alghamdi MA (2019) Modeling mental stress using a deep learning framework. *IEEE Access* 7
9. McWhorter TM, Ni Y, Nie H, Iarve J, Majumder AKM, Ucci DR (2019) sEmoD: A personalized emotion detection using a smart holistic embedded IoT system. In: 2019 IEEE 43rd annual computer software and applications conference (COMPSAC)
10. Koldijk S, Sappelli M, Verberne S, Neerinx MA, Kraaij W (2014) The SWELL knowledge work dataset for stress and user modeling research. *ACM*
11. Ceja EG, Osmani V, Mayora O (2016) Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE J Biomed Health Inform* 20
12. Baig MM, Hosseini HG, Moqem AA, Mirza F, Lindén M (2017) A systematic review of wearable patient monitoring systems—current challenges and opportunities for clinical adoption. *Mobile Wireless Health*
13. Murali S, Rincon F, Atienza D (2015) A wearable device for physical and emotional health monitoring. *Comput Cardiol*
14. Leu FY, Ko CY, You I, Choo KKR, Ho CL (2017) A smartphone-based wearable sensors for monitoring real-time physiological data. *Comput Electr Eng*
15. Ciman M, Wac K (2016) Individuals' stress assessment using human-smartphone interaction Analysis. *IEEE Trans Affect Comput*
16. Kandias M, Gritzalis D, Stavrou V, Nikoloulis K (2017) Stress level detection via OSN usage pattern and chronicity analysis: an OSINT threat. *Comput Secur*

17. Pandey P, Lee EK, Pompili D (2016) A distributed computing framework for real-time detection of stress and of its propagation in a team. *IEEE J Biomed Health Inform* 20(6)
18. He C, Yao Y, Ye X (2017) An emotion recognition system based on physiological signals obtained by wearable sensors. *Wearable Sens Robots*
19. Li C, Xu C, Feng Z (2016) Analysis of physiological for motion recognition with the IRS Model. *Neurocomputing* 103–111
20. Dhall A, Goecke R, Gedeon T (2015) Automatic group happiness intensity analysis. *IEEE Trans Affect Comput* 6
21. Aigrain J, Spodenkiewicz M, Dubuisson S, Detyniecki M, Cohen D, Chetouan M (2015) Multimodal stress detection from multiple assessments. *IEEE Trans Affect Comput* 14
22. Chen T, Yuen P, Richardson M, Liu G, She Z (2014) Detection of psychological stress using a hyperspectral imaging technique. *IEEE Trans Affect Comput* 5(4)
23. Chen L, Zhao Y, Ye P, Zhang J, Zou J (2017) Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Syst Appl*
24. Sevil M, Hajizadeh I, Samadi S, Feng J, Lazaro C, Frantz N, Yu X, Brandt R, Maloney Z, Cinar A (2017) Social and competition stress detection with wristband physiological signals. *IEEE* (2017)
25. Xu Q, New TL, Guan C (2015) Cluster-based analysis for personalized stress evaluation using physiological signals. *IEEE J Biomed Health Inform* 19
26. Freedenberg VA, Hinds PS, Friedmann E (2017) Mindfulness-based stress reduction and group support decrease stress in adolescents with cardiac diagnoses: a randomized two-group study. *Pediatr Cardiol*
27. Xie J, Wen W, Liu G, Chen C, Zhang J, Liu H (2016) Identifying strong stress and weak stress through blood volume pulse. *IEEE*
28. Martinez R, Irigoyen E, Arruti A, Martin JI, Mugerza J (2017) A real-time stress classification system based on arousal analysis of the nervous system by an F-state machine. *Comput Methods Programs Biomed* 81–90
29. Hana L, Zhanga Q, Chena X, Zhanc Q, Yangc T, Zhao Z (2017) Detecting work-related Stress with a wearable device. *Comput Ind*
30. Xua M, Purdona C, Selib P, Smileka D (2017) Mindfulness and mind wandering: the protective effects of brief meditation in anxious individuals. *Consciousness Cogn* 51
31. Zhanga Q, Chen X, Zhan Q, Yang T, Xia S (2017) Respiration-based emotion recognition with deep learning. *Comput Ind* 92:84–90
32. Jung Y, Yoon YI (2017) Multi-level assessment model for wellness service based on human mental stress level. *Multimedia Tools Appl* 76:1305–11317
33. Kim KH, Bang SW, Kim SR (2004) Emotion recognition system using short-term monitoring of physiological signals. *Med Biol Eng Comput* 42:419–427
34. Choi J, Ahmed B, Gutierrez-Osun R (2012) Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE Trans Inf Technol Biomed* 16(2):279–287
35. Rebolledo-Mendez G, Reyes A, Paszkowicz S, Domingo MC, Skrypchuk L (2014) Developing a body sensor network to detect emotions during driving. *IEEE Trans Intell Transport Syst* 15(4)
36. Lee BG, Chung WY (2017) Wearable glove-type driver stress detection using a motion sensor. *IEEE Trans Intell Transport Syst* 18:1835–1845
37. Shroff S, Pise S, Chalekar P, Panicker S (2015) Thyroid disease diagnosis: a survey. In: *IEEE sponsored 9th international conference on intelligent systems and control (ISCO)*
38. Mane VL, Panicker SS (2015) Summarization and sentiment analysis from user health posts. In: *2015 international conference on pervasive computing*
39. Mane VL, Panicker SS (2015) Knowledge discovery from user health posts. In: *2015 IEEE 9th international conference on intelligent systems and control (ISCO)*
40. Pathak P, Rathor S (2019) Human dynamic analysis and inference system on health problems and stress in IT industry by using soft computing techniques. In: *2019 4th international conference on information systems and computer networks (ISCON), Mathura, India, pp 230–233*

41. Panicker S, Bhujange S, Karne S, Kadam R (2000) Finding patterns in biological parameters. *Int J Recent Innov Trends Comput Commun* 4(12). ISSN: 2321-8169
42. Komarov O, Ko L, Jung T (2020) Associations among emotional state, sleep quality, and resting-state eeg spectra: a longitudinal study in graduate students. *IEEE Trans Neural Syst Rehabil Eng* 28(4):795–804
43. Ahuja R, Ban A (2019) Mental stress detection in university students using machine learning algorithms. In: *International conference on pervasive computing advances and applications (PerCAA 2019)*. *Proc Comput Sci* 152:349–353
44. Bankar C, Bhide A, Kulkarni A, Ghube C, Bedekar M (2018) Driving control using emotion analysis via EEG. In: *2018 IEEE Punecon, Pune, India*, pp 1–7
45. Kunte AV, Panicker S (2020) Analysis of machine learning algorithms for predicting personality: brief survey and experimentation. In: *2019 global conference for advancement*
46. Kunte A, Panicker S (2020) Personality prediction of social network users using ensemble and XGBoost. In: *Das H, Pattnaik P, Rautaray S, Li KC (eds) Progress in computing, analytics and networking. advances in intelligent systems and computing, vol 1119*. Springer, Singapore
47. Kunte AV, Panicker SS (2019) Using textual data for personality prediction: a machine learning approach. In: *2019 4th international conference on information systems and computer networks (ISCON)*
48. Elzeiny S, Qaraq M (2018) Machine learning approaches to automatic stress detection: a review. In: *2018 IEEE/ACS 15th international conference on computer systems and applications (AICCSA)*, Aqaba, pp 1–6
49. Smets E, Raedt WD, Hoof CV (2018) Into the wild: the challenges of physiological stress detection in laboratory and ambulatory settings. *IEEE*

Enhanced Movie Reviews Classification Using Precise Combination of PCA with LSTM Model



Kriti Bansal, Nancy Bansal, Aman Agrawal, and Bharat Mohan Thakur

Abstract As everything is becoming online day by day, a large number of comments or reviews are recorded for various types of products, movies, photographs, etc. Due to its large number, it is very difficult and time taking for someone to read all those comments and then classify each comment as positive or negative. So, in this work, my area of concern is to extract the sentiment from the text and to classify it into suitable categories. In this work, long short-term memory (LSTM) and recurrent neural network (RNN) in addition to hybrid model of (LSTM and PCA) are developed to analyze and classify reviews of the movie. Here, principle component analysis (PCA) is used to get the principle components or we can say for extracting the features which can be used as predictors. In this work, we are using IMDB movie survey dataset, so that it arranges a survey depends upon the neural system in LSTM with PCA. For movie description, it takes useful keywords from reviews of the movie. By this we could physically choose a film is positive reviews or negative. Utilizing the AI approach, we are arranging the film audits to such an extent it is easily decided that movie is positive or negative. Results have shown that the hybrid model of LSTM with PCA have outperformed the singular LSTM and RNN networks. LSTM with PCA have reported the accuracy of 88.2% as it gives higher accuracy as compared to traditional approaches.

Keywords Long short-term memory (LSTM) · Natural language processing (NLP) · Principle component analysis (PCA) · Recurrent neural network (RNN)

K. Bansal (✉) · A. Agrawal

Department of Computer Engineering and Application, GLA University, Mathura, India

e-mail: kriti.bansal@gla.ac.in

A. Agrawal

e-mail: aman.agrawal@gla.ac.in

N. Bansal

Department of Computer Engineering (Cyber Security), National Institute of Technology, Kurukshetra, India

e-mail: nencybansal42@gmail.com

B. M. Thakur

Alten India Private Limited, Bengaluru, India

e-mail: bharatmohanthakur1@gmail.com

1 Introduction

In the field of NLP, text characterization has been a great undertaking and warmed examination hotspot. As with the increase in use of Internet, data on the Internet is increasing day by day. As anything we select on Internet, we select on the basis of its review. As users are increasing, reviews of a product, movie, etc., are also increasing day by day. It is very difficult and time consuming to read all the reviews and classify them as positive and negative reviews. We have made this model, to classify the reviews as positive reviews and negative reviews of the movie. The problem with previous models is, that the output of classification depends only on the current state. Output was not able to consider all the previous data. LSTM stores previous state data and while generating the output, considers the previous state data. Hence, it produces more accurate results when compared with previous models like CNN, RNN, etc.

The aim is to define categories for given sequence of text to learn text introductions for arrangement, for example, convolutional neural system (CNN) models [1] and recurrent neural system (RNN) models [2]. In its initial stage, estimation order is performed utilizing techniques like SVM, Naive Bayes, and so on. As of late profound techniques of learning like utilizing these systems (CNN, RNN, ANN, and so on) are picked up notoriety by indicating amazing outcomes. Customary strategies when contrasted with profound learning techniques have their confinements in breaking down the crude information or in its normal structure. While in profound learning it learns the portrayal of information at various layers by consolidating the straightforward non-direct capacities which change the crude information into the higher dynamic level portrayal. This undertaking is on the film audit grouping dependent on assessment investigation utilizing the AI approach. There are numerous ways that I can use to characterize the survey depends on the feeling yet I am utilizing the trend-setting innovation which is LSTM [3] systems. The main problem of RNN is vanishing gradient.

We realize that for a customary feed-forward neural system, the weight refreshing that is applied on a specific layer is a variety of the learning rate, the blunder term from the past layer, and the contribution to that layer. In this way, the mistake term for a specific layer is someplace a result of every past layer's blunders. When managing actuation capacities like the sigmoid capacity, the little estimations of its subsidiaries (happening in the blunder work) get duplicated on numerous occasions as we move toward the beginning layers. Therefore, the slope nearly evaporates as we move toward the beginning layers, and it gets hard to prepare these layers. Because of this issue, in the event that we need the data after a little league, it might be reproducible, yet once plenty of words are taken care of in, this data gets lost someplace.

This issue can be settled by applying long short-term memory networks [4]. So as to include new data, RNN changes the current data completely. Along these lines, the whole data is adjusted. There is no classification for important data and not important data. LSTM cells make little changes to the current data by passing the data through various gates. These gates help to store only the important data of the previous states and current state. Data which is not so important is removed from the memory cells.

Because of this property of LSTMs [5], they do not change the whole data, rather tries to keep the required data from the previous state, so that it is able to keep history of long time.

With the rise of Web-based life, a lot of significant information opens up on the Web and simple to get to. Online networking clients examine all that they care about through blog entries or tweets share their suppositions and show intrigue uninhibitedly, while they do not really do it face to face. We read about political discussions, social issues, inquiries regarding a specific item, and so forth. Organizations likewise utilize interpersonal organizations to advance their items and benefits and investigate individuals' sentiments to improve their items and administrations, in this way creating a tremendous measure of information.

In this specific circumstance, the requirement for a logical instrument that can procedure the client's information and orders them as far as estimation polarities is expanded and turn into a need. Subsequently, Sentiment examination is significant on the grounds that everybody needs their film being seen emphatically, or if nothing else more positive than the motion pictures of contenders. Assumption investigation, if precise, can be a truly significant instrument for this particular use case.

2 Related Work

Sentiment Analysis is a well-known field to discover experiences from content information from different sources like Facebook, Twitter, Amazon, and so on. Jagdale et al. [6] applied machine learning algorithms for preprocessing. In this paper, Amazon dataset is used which contains reviews of camera, laptops, mobile phones, tablets, etc. Reviews are both positive and negative. Both support vector machine and Naive Bayes are applied and 73.54 and 78.17% accuracies are achieved, respectively, for camera reviews. Erik et al. [7], In this author used a couple of symbolic and sub-symbolic AI to consequently find applied natives for sentiment analysis.

One of the interesting applications of sentiment analysis is to analyze Twitter data so Fouad et al. [8] proposed an efficient technique for Twitter sentiment analysis. An AI calculation is utilized to fabricate an application for identifying positive and negative tweets. This model uses different strategies to speak to named input tweets in the preparation stage where it utilizes different highlights sets. The following stage is the order stage; the classifier bunch is given particular base classifiers for continuously exact results. The proposed structure can be used for evaluating customers' appraisal from their tweets which is useful in various applications, for instance, showcasing, political furthest point recognizable proof, and inspecting things.

The next article is also based on machine learning algorithms and it uses Naive Bayes and support vector machine; so Hasan et al. [9] proposed a work for sentiment analysis during elections. This work provides the comparison between different lexicons like TextBlob, W-WSD, and SentiWordNet and selects best between them. These three lexicons are validated using two machine learning techniques, i.e., support vector machine and Naive Bayes. These two classifiers are used to test result

with TextBlob, W-WSD, and SentiWordNet. This paper concludes that results with TextBlob are better but with W-WSD gives the best result for this approach.

In past explores plainly LSTM model is effective with assorted data sources however it does not have a touch of exactness on account of short sentences. Many proposed models are not at all that fruitful in effectively grouping the long content information. Then again models consolidating LSTM systems are demonstrating noteworthy outcomes that are on the grounds that its ability to manage the long content information.

In the work of Li [10] has discussed improvement for classify the sentiment analysis using logistic regression. They have introduced the dataset origin, preprocessing steps, improvement of the features, analysis of error, and various parameters are tuned through CV parameter selection.

Sajeevan et al. [11] investigated two models of deep learning for exploring IMDB movie reviews. One is CNN-LSTM model and other is LSTM-CNN model. These approaches performed well on IMDB review dataset and get higher accuracy as compared to other traditional methods.

Rahman and Hossen [12] proposed the approach of machine learning for polarity classification on movie review dataset. This technique is divided into training and testing set. Initially, dataset is collected. Then they perform preprocessing on the dataset by using the tool of NLP. After that creating the feature vectors then dataset is trained using different classifiers of machine learning like Bernoulli NB, entropy, decision tree, multinomial NB, and SVM which are tested using review dataset.

Bansal and Jain [13] proposed machine learning approach based on the performance of collaborative filtering (CF). The selected text is analyzed on the basis of recommendation algorithm and similarity threshold. Here, the CF-based recommendation algorithm performs best for 80% similarity threshold. Further, higher is the value of the similarity threshold and less is the MAE and better is the performance of the recommendation system. So, it is better to say that the similarity threshold plays an important role in the performance of CF-based recommendation system. High value of similarity threshold gives better results.

Bharadwaj et al. [14] proposed a model of the micro-blogging site like Twitter to analyze the user behavior. It analyzed the tweets on a particular topic, and the users of those tweets are also analyzed. User's previous tweets are analyzed to identify their tweet pattern.

3 Proposed Work

In this work, firstly we have to download and retrieve the movie data from IMDB dataset. Then, process and prepare the data using preprocessing steps which are described further in the paper as shown in Fig. 1. After that train the chosen model. Then, test the trained model (typically using a batch transform job) for predicting the results and finding accuracy and other performance metrics.

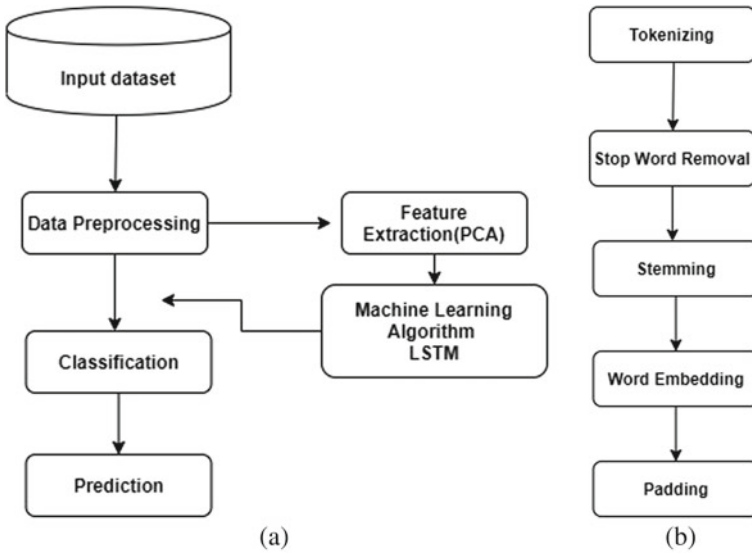


Fig. 1 Flowchart of **a** classification process and **b** data preprocessing

3.1 Dataset

IMDB movie review dataset [15] is used in this work, which is available by default in the Python module `keras.dataset`. It contains many datasets which are used in machine learning models for training. This dataset has a sum of 50,000 surveys which are text documents. 500 words is the most extreme length of a review. The whole dataset is isolated into two sets that are training and testing process. Both testing and training data sets contain positive as well as negative reviews. I have shuffled the reviews so that positive or negative reviews may not accumulate and are not continuous.

3.2 Preprocessing

In this work, we have processed and prepared the data for training after going through following processes:

Tokenization. This is the initial step to continue with NLP. Tokenization implies parting an expression, sentence, passage, or a whole book report into littler units, for example, singular words or terms.

Stop words. Stop words are removed from the dataset. These are the words which do not contribute to the sentiment of the review. For example, all the articles, punctuation marks, pronouns, etc., removing these words help in increasing accuracy.

Beautiful Soup. It is a library of Python for extracting the data from XML and HTML files. I have used Beautiful Soup in my work to remove all type of tags, hyperlinks, etc.

Stemming. It is used to convert similar metaphoric words into root word. For example, entertainment, entertaining, entertained will all be converted to the root word entertain. As we just want to analyze and understand the sentiment or emotion from the text, i.e., reviews of the person, we want to reduce all the similar words to the same stem.

Padding. It is applied for each review to convert them of the same size. Padding is done before training the model. While training the model, size of review must be same, so to make the size of the reviews same, padding is done on the reviews which have the smaller size.

3.3 Principle Component Analysis

Principle component analysis (PCA) [16] converts high-dimensional data into lower-dimensional data by which we get the maximum variance and minimum least square error. PCA is a process of decomposition of eigenvalues of the covariance matrix of the variables.

The steps of principle component analysis are as follows:

- Transform an $N \times d$ matrix X into an $N \times w$ matrix Y
- Centralized the data by subtracting the mean
- Then calculate the covariance matrix $d \times d$
- After that calculate the eigenvectors of covariance matrix
- Then select the w eigenvectors which is related to largest w eigenvalues of the covariance matrix.

3.4 Classification

In the wake of principle component analysis and preprocessing the information, information is prepared for preparing the model. We have utilized the Pytorch library in my task for preparing and testing the model. The model is instated utilizing different layers. In this stage, a model for grouping the surveys dependent on the notion is created utilizing the LSTM calculation. The model is prepared with the preparation dataset and afterward is tried in the coming stages. The model for assumption grouping is totally worked in this stage.

Here, we have used three layers to initialize the model: embedding layer, LSTM layer, and dense layer. The output of one layer is used as an input to the next layer. Final multidimensional output is squeezed to single dimension and passed as an input to the sigmoid function. It converts the output in the range of 0–1. We have

used binary cross-entropy loss function which calculates the loss or the error. We have used it as the decision boundary is large, so it performs better and provides more accurate results. Adam optimizer is used for optimization. It is also known as adaptive optimizer as it changes its value in every epoch. It is faster, more reliable to reach a global minimum and works better when minimizing the cost function while training the model. Finally, the model is tested using the test data and all the performance metrics are calculated to visualize the results.

LSTM works on the memory blocks known as cells. These cells can decide which information is less important that can be removed and which information is more important to store. It can also help to decide when to move the previous state's information to the next state. It uses three inputs: long-term memory, short-term memory, and current event: It is basically the current input provided for the prediction.

LSTM use four gates which are trained using backpropagation by adjusting the weights that are feed into them. These gates are: forget gate, learn gate, remember gate, and remember gate.

By applying LSTM we get good results as compared to recurrent neural network (RNN) model. After that we try to apply the combination of the models, i.e., LSTM with PCA and get the higher accuracy as compared to singular LSTM and RNN networks. Here, principle component analysis (PCA) is used to get the principle components or we can say for extracting the features which can be used as predictors.

The entire work is implemented on Jupyter Notebook using PyTorch backend. Adam optimizer is employed to the model with the default learning rate of 0.001. The performance of the model is measured across metrics of 'accuracy' and with loss function 'binary cross entropy' with number of epochs = 10. Vocabulary size is 5000. Hidden dimension is taken as 200 and embedding dimension is taken as for better results.

4 Results

Standard evaluation metrics is used for measuring the performance of classification model, i.e., precision, recall, and F-measure which is shown in Eqs. 1–3.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$\text{F-measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

ROC is the receiver operating characteristic curve. The area under this curve represents the accuracy of the model. For the ideal model, the area under the curve is

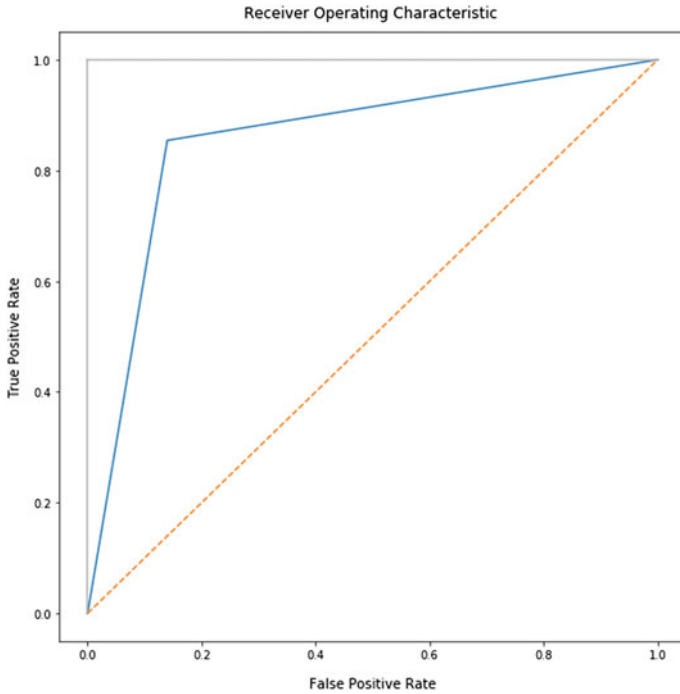


Fig. 2 ROC curve of proposed approach

1. As the accuracy of our model is 0.882, the area under the curve is 0.89 as shown in Fig. 2.

Table 1 shows the comparison of prediction results of all the three models which we have worked in this paper. Here, we calculate precision, recall, and F-measure and compared with all three models by it. Comparison of proposed LSTM+PCA approach with RNN and LSTM model w.r.t precision, recall, and F-measure is shown in Fig. 3a and accuracy comparison of proposed hybrid LSTM with PCA model and singular LSTM and RNN is shown in Fig. 3b. Here, we get 0.88 accuracy of the proposed model, i.e., LSTM with PCA.

5 Conclusion and Future Work

This model has been made to understand the feeling which a viewer feels, after watching a movie. LSTM classifies these reviews in positive and negative categories well overall. This system will be helpful to improve the motion pictures by becoming more acquainted with which substance has great gathering by the crowd and which substance is not without the necessity of any human asset. This framework is different and more efficient than earlier machine learning techniques as it also uses long-term

Table 1 Prediction results

	RNN				LSTM				Proposed model				Support
	Precision	Recall	F-measure		Precision	Recall	F-measure		Precision	Recall	F-measure		
Positive	0.82	0.79	0.83		0.86	0.83	0.87		0.89	0.86	0.9		12500
Negative	0.81	0.81	0.81		0.85	0.85	0.85		0.88	0.88	0.88		12500
Microaverage	0.81	0.83	0.83		0.85	0.87	0.87		0.88	0.9	0.9		25000
Macroaverage	0.82	0.83	0.84		0.86	0.87	0.88		0.89	0.9	0.91		25000
Weighted average	0.81	0.8	0.82		0.85	0.84	0.86		0.88	0.87	0.89		25000

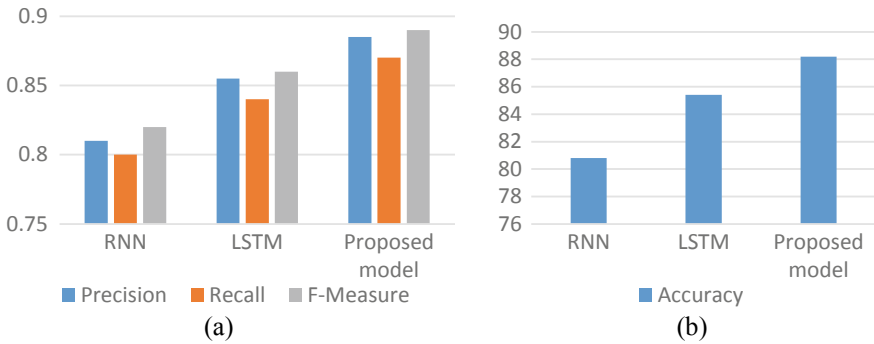


Fig. 3 **a** Comparison of proposed LSTM+PCA approach with RNN and LSTM model w.r.t precision, recall, and F-measure and **b** accuracy comparison of proposed hybrid LSTM with PCA model and singular LSTM and RNN

memory apart from short-term memory. It is therefore much more useful for long reviews apart from short reviews. LSTMs do not control the whole data identified with past encounters but instead adjust them somewhat as per the prerequisite. They can overlook and recall things specifically, which makes it progressively precise and helpful for long content and information.

By applying LSTM, we get good results as compared to recurrent neural network (RNN) model. After that we try to apply the combination of the models, i.e., LSTM with PCA and get the higher accuracy as compared to singular LSTM and RNN networks. By utilizing bigger datasets we can in any case improve the presentation of the model. A new algorithm known as BERT has been made for the text classification to further improve the accuracy of the model.

References

1. Conneau A, Schwenk H, Barrault L, Lecun Y (2016) Very deep convolutional networks for text classification. arXiv preprint [arXiv:1606.01781](https://arxiv.org/abs/1606.01781)
2. Kokkinos F, Potamianos A (2017) Structural attention neural networks for improved sentiment analysis. arXiv preprint [arXiv:1701.01811](https://arxiv.org/abs/1701.01811)
3. Priyantina RA, Sarno R (2019) Sentiment analysis of hotel reviews using latent dirichlet allocation, semantic similarity and lstm. *Int J Intell Eng Syst* 12(4):142–155
4. Wen S, Wei H, Yang Y, Guo Z, Zeng Z, Huang T, Chen Y (2019) Memristive LSTM network for sentiment analysis. *IEEE Trans Syst Man Cybern Syst*
5. Bodapati JD, Veeranjanyulu N, Shaik S (2019) Sentiment analysis from movie reviews using LSTMs. *Ingénierie des Systèmes d'Information* 24(1):125–129
6. Jagdale RS, Shirsat VS, Deshmukh SN (2019) Sentiment analysis on product reviews using machine learning techniques. In: *Cognitive informatics and soft computing*. Springer, Berlin, pp 639–647
7. Cambria E, Poria S, Hazarika D, Kwok K (2018) Senticnet 5: discovering conceptual primitives for sentiment analysis by means of context embeddings In: *Thirty-second AAAI conference on artificial intelligence*

8. Fouad MM, Gharib TF, Mashat AS (2018) Efficient twitter sentiment analysis system with feature selection and classifier ensemble. In: International conference on advanced machine learning technologies and applications. Springer, Berlin, pp 516–527
9. Hasan A, Moin S, Karim A, Shamshirband S (2018) Machine learningbased sentiment analysis for twitter accounts. *Math Comput Appl* 23(1):11
10. Li C (2019) Sentiment analysis for IMDB movie review
11. Sajeewan A, Lakshmi K (2019) An enhanced approach for movie review analysis using deep learning techniques. In: 2019 International conference on communication and electronics systems (ICCES). IEEE, pp 1788–1794
12. Rahman A, Hossen MS (2019) Sentiment analysis on movie review data using machine learning approach. In: 2019 international conference on bangla speech and language processing (ICBSLP). IEEE, pp 1–4
13. Bansal K, Jain V (2020) Performance analysis of collaborative filtering based recommendation system on similarity threshold. In: 2020 international conference on computing methodologies and communication (ICCMC). IEEE, pp 442–448
14. Bharadwaj D, Tripathi A, Agrawal A, Khan MA (2020) Analysis of users behavior on micro-blogging site using a topic. In: International conference on advanced machine learning technologies and applications. Springer, Berlin, pp 429–438
15. ai.stanford.edu/~amaas/data/sentiment/
16. Wang Y, Xie D, Wang X, Zhang Y (2018) Prediction of wind turbinegrid interaction based on a principal component analysis-long short term memory model. *Energies* 11(11):3221

Academic Analytics Through MOOCs



Arpana Rawal, Shruti Gupta, Nishant Chhattani, and Jaideep Sachdeva

Abstract The stepping milestone of achievements enjoyed by Mission NPTEL MOOCs under (Study Webs of Active–Learning for Young Aspiring Minds) SWAYAM has succeeded in diverting the teaching–learning environments to more amicable knowledge sharing platform and in enriching their learning experience in more promising directions using famous 4-quadrant pedagogy. This manuscript attempts to build a prototype model to capture necessary academic analytics (local chapterwise) by looking into enriched profiles of MOOC learners. In such a scholarly mission, such analytics is majorly necessary to tap learning skills and employability potential of youth actively engaged in MOOCs, nation-wide.

Keywords Academic analytics · NPTEL local chapters · Learners’ profiles · MOOCs · Recommendation system · Impact analyses

1 Introduction

SWAYAM NPTEL has been able to succeed through in breaking the geographical, psychological, and scholarly barriers of knowledge-pooling by offering enormously diversified self-paced online certification courses across engineering, humanities, and science streams for more than a decade; the pioneering effort began from March 2014 that has been gaining momentum with more than four lakh learners, both faculty and students, already been awarded with online certification in more than a thousand

A. Rawal (✉) · S. Gupta · N. Chhattani
Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, India
e-mail: arpana.rawal@gmail.com

S. Gupta
e-mail: shrutiguptakc@gmail.com

N. Chhattani
e-mail: nishant.chhattani@gmail.com

J. Sachdeva
Department of Computer Science and Engineering, Manipal University, Jaipur, India
e-mail: jaideepsachdev1@gmail.com

course run offered. As on date, the academic learners' profiles are an invaluable asset situated safe in consolidated formats inside National Server-stations of NPTEL organizers (IITs/IITs) of the Nation.

The local chapters have become the inevitable components behind milestone successes of MOOCs as they provide dedicated human resource chains of SPOCs—Mentors—Mentees. Mentees are the baseline operational machinery-cum-beneficiary who are consistently monitored and catered with all the needed learning stuff at high satisfaction levels. We would like to draw your attention to the fact that there exists a robust coordination hierarchy among the three mentioned components, and however, the monitoring of self-paced involvement by learners requires carefully formulated novel boosting mechanisms. This calls for exploring manifold queries at the end of local chapters as a part of academic analytics: *what percentage of learners are inclined to MOOCs courses in continuum? What are the learners' engagement statistics, their study yearwise? What are the popular MOOCs in demand among students, study yearwise? Do engagement in MOOCs help learners to select their directions of disciplinary specialization courses in higher study years?* and many more...

The rest of the paper details the content in three more sections. Section 2 outlines the motivational stepping stones behind undertaking learners' analytics. Section 3 identifies some of the analytical objectives necessary at levels of institutional and disciplinary representatives of local chapters. The learners' attributes contributing to analytics are assimilated and visualized. The encouraging conclusive remarks are provided in Sect. 4.

2 Mission SWAYAM NPTEL: Rising Wave

If the participation figures are analyzed for all local chapters, the academic organizations are seen, bound to get inclined toward overwhelming shift in imparting online mode of higher education by course organizers and revitalized online learning patterns of engaged learners, season after season across the whole country. The above analysis can be supported by the rising statistics observed in academic seasonwise, percentage increase in courses offered and also enrolled local chapter counts. This was perceived as an opportunity to explore learning patterns among various levels of learners and hidden correlations among crucial governing parameters for respective local chapters.

2.1 Engagement with MOOCs: The State-of-the-Art Figures

With the onset of any academic season, we begin analyzing the rising MOOC enrollment figures of SWAYAM NPTEL local chapter institutions vs. non-engaged higher education institutions of the country. The present case study focuses only to the

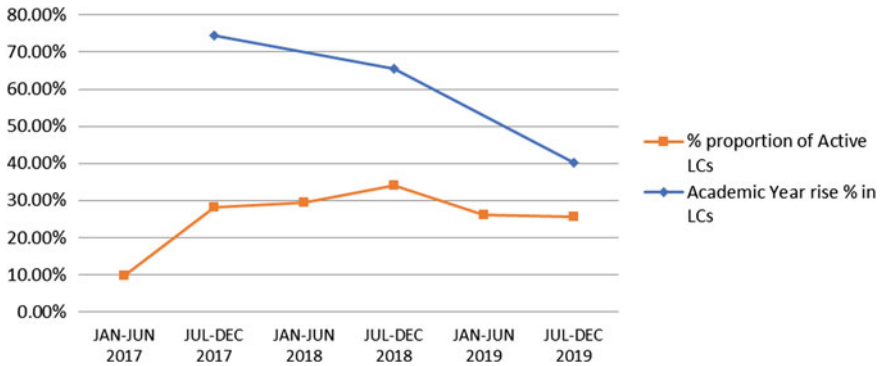


Fig. 1 Rising statistics of enrolled local chapters versus actively tagged local chapters

courses conducted by NPTEL forum under SWAYAM whose end user beneficiaries are engineering institutions across the nation. An academic institution needs no scholarly constraint to enroll itself as an upcoming local chapter; nevertheless, the count of consistently active local chapters requires the chapter to meet the scholarly criterion of satisfying two sets of conditions that a minimum count of learners appear for the end season certification exam as well as half of them get “certified”; this minimum threshold may be changed after few academic seasons, in order to be tagged as ACTIVE for three consecutive academic years. The graph shown in Fig. 1 reveals on an average, 60% rise in local chapter counts each academic year, while there is only less than half the rate (26%) increase in actively tagged local chapters, each academic season.

Another key observation of direct impact can be seen as tabulated in Fig. 2 for each academic season since the inception year (2014) of the mission. The statistics reveal an enormously disproportionate rise in learners’ enrollments, learners’ exam registrations, and learners’ certification counts in descending order of their dis-proportions. This too augmented the dearth of qualitative data analytics for MOOCs learners.

The above statistics reveal that MOOCs, though, have invaded as a strong academic supplement to higher education system of the Nation on quantitative scale, but yet to scale up on a quality productive note. To achieve more productivity, there is a need to keep constant vigilance on the instructional and managerial channels. Hence, a data analytic framework participating in 24 × 7 engagements with MOOCs in order to drill deep insights of learning levels is the appropriate solution to resolve the quality issue.

2.2 24 × 7 Engagement with MOOCs

The learner’s analytical framework requires a rigorous 24 × 7 scholarly evaluation of SWAYAM NPTEL MOOCs enrollment, registration, certification frequencies of each

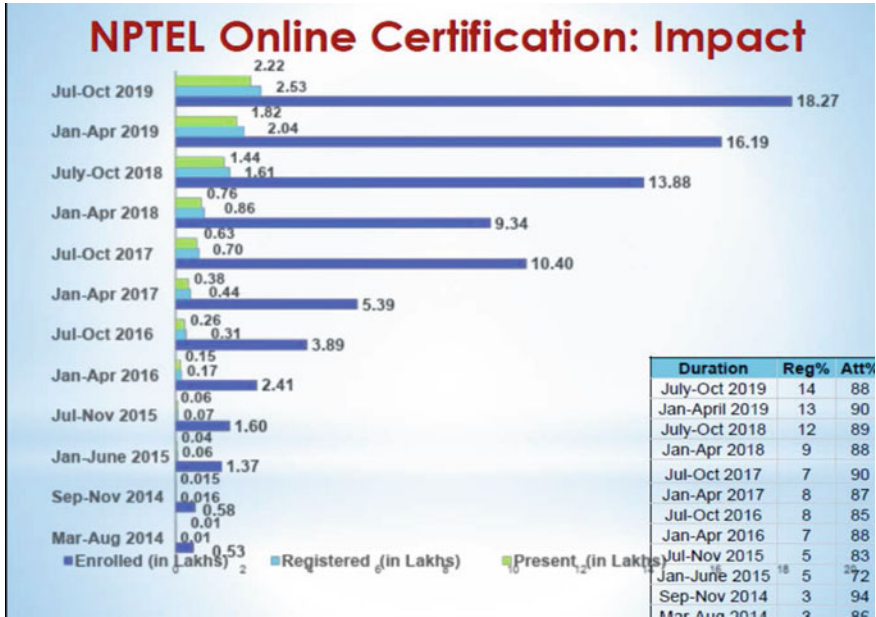


Fig. 2 Academic seasonwise impact metrics of MOOC learners (enrollment: registration: certification counts), courtesy. SWAYAM NPTEL, MHRD, Govt. of India

learner in past “n” consecutive academic seasons, past learners’ MOOCs profiles, their academic credentials, course-type taxonomies of the courses offered, assignment submission frequencies, discussion forums, live interactive sessions, mentees’ active engagement, feedback from mentors and mentees at a consistent pace, all ensured to be incorporated in a self-paced monitoring system. Such an analytical framework is sure to raise quality standards of a local chapter in appraising learners’ skills by being more punctual, more disciplined, more interactive with peers and course instructors.

2.3 Incentives, Awards, and Accolades

Tagging of “Active” local chapters, top 100 rating of local chapters, appreciably meritorious certification categories, top performing mentoring awards and NPTEL Stars are some of the incentives that keep the zeal and passion of learning in high flagging spirits. All such accolades designed to encourage the learning spirits of SWAYAM NPTEL learners shall only receive increasing attention, if there exists a robust qualitative assessment framework for each local chapter unit to provide predictive statistics on mentees, mentors, and the learning quotient of the chapter as a whole.

3 Learners' Analytics

Initially, the analytics focused on preliminary quantitative case study from a local chapter X which joined SWAYAM NPTEL mission with the commencement of its academic year 2016–2017. Several motivational factors were deployed for mass campaigning of NPTEL courses. The most prominent factor behind encouraging student communities toward NPTEL was made to realize that their learners' engagement is in line with their learning objectives set by the curriculum designers in all their academic years of study in higher engineering education. This was the juncture, the management component of academia decided to drill down into certification and learning behaviors of both, their certified and non-certified (but enrolled) learners (both students and faculty participants).

Datasets. The paper attempts to undertake analytical study for a span of 6–7 academic seasons. The authors have attempted to publish true findings based on the data analytic experiments using empirical datasets captured during mentioned academic study span.

3.1 Recommender Systems

The most crucial of the academic analytical objectives was, how to judge the most rationally justified scholarship recommendation nominations, if the student scholarship requests exceed the set threshold. Such situations are to be handled by SPOCs of respective local chapters by selecting nominations manually, trying to be unbiased though. The experimented framework took datasets of participated students in past six academic seasons of SWAYAM NPTEL. These students were uniquely identified by attributes: {email_Id (Primary Key), name, Enr_Cnt (The number of times a student has enrolled for any course), Reg_Cnt (number of times a student has registered for the exam for any course), Sch_Nom_Cnt (number of times a student has been nominated for a scholarship), Sch_App_Cnt: number of times a student has received the scholarship, Exam_Cnt: number of times a student has appeared for the exam he/she registered for, Final_Avg_Score: The mean of the final score of the student in all exams in past academic seasons & Year_Of_Passing: (graduation year of a student)}. It was necessary to keep the last attribute, so that we could segregate student learning-cum-performance patterns on the basis academic study years in engineering curriculum.

In very beginning, the first few academic seasons did not used to provide student registration database with the attribute, “Year_Of_Passing”, hence, these could not be considered during recommendation tasks. The test student instances for the analytics model were described with attribute schema of the set “T”, where each student is seen participating in ongoing NPTEL academic session and whose scholarship recommendation is asked for; $T = \{Reg_Cnt, Exam_Cnt, Sch_Nom_Cnt, Sch_App_Cnt, Final\ Avg\ Score, Year_Of_Passing\ (study_year_ID)\}$.

The whole recommendation system framework is outlined in following step sequence:

Step 1. Initially, the set of students participating in ongoing NPTEL academic session (i.e., test instances) were segregated into four groups, based on which year of study, viz. 1st year, 2nd year, 3rd year, or 4th year, as: T_1 , T_2 , T_3 , and T_4 . Each of these instances per study_year group is then treated upon for recommendation task individually.

Step 2. A novel concept of rank value computation is used to arrive at attribute value “Attr” for each test student instance, such that we can order these student instances in order of their residual scholarship preferences, looking into their past NPTEL academic profiles; the higher the “Attr” values, it implies, the higher is the preference allotted to that test instance. The rank value “Attr” was conceptualized from two contributive terms, namely α and β . In the expression (1), the first term (α) takes into consideration of ascending orders of preference owing to the student’s proportionate involvement in online NPTEL certification examination, every time, he/she registers for the same examination as well as the second term (β) contributes the share of negative bias upon the preference assigned due to the first term (α), hence expressed as $(1 - \beta)$

$$\text{Attr} = \alpha + (1 - \beta) \quad (1)$$

Suppose, a student who has been awarded that many number of times, as he was nominated for NPTEL scholarship in past his/her past performances, should be retrograded with lower preference (rank) value. The expressions for & are formulated in Eqs. (2) and (3).

$$\alpha = \frac{\text{Exam_cnt}}{\text{Reg_Cnt}} \quad (2)$$

$$\beta = \frac{\text{Sch_App_cnt}}{\text{Sch_Nom_Cnt}} \quad (3)$$

Step 3. From the given set “T”, the seed data points were fetched out as those student instances, who had already been awarded with NPTEL scholarship at least once in their past academic seasons, provided these academic seasons were considered for formulation of attributes in the schema of set “T”. The criteria for selection of such seed points in each of the four study_year groups (corresponding to each of the study years) were decided as: “Reg_Cnt \geq 1 AND Exam_Cnt \geq 1 AND Final Avg Score \geq 40 AND Sch_App_Cnt \geq 1”.

Step 4. In this way, there may be multiple seed points per study_year_group, where each such group is treated as a non-overlapping subset of input to our clustering framework. It may noted that the experiment was performed for even semester of academic season and hence, every test student instance could be scanned

for his/her past academic session's performance to obtain attributes; {Reg_Cnt, Exam_Cnt, Sch_Nom_Cnt, Sch_App_Cnt, Final Avg Score, Year_Of_Passing (study_year_ID)}.

Step 5. This step marks the beginning of our clustering activity. It begins with computation of cluster centroids from the set of seed point(s) by averaging upon their respective (three) attributes of only those seed instances whose rank values coincide; the rest of unique seed instances exist as they are in that study_year_group. Thus, the total number of centroids was equal to the number of seed instances bearing distinct "Attr" values. The attributes put for averaging computation are namely examination registration counts (Reg_Cnt), examination_appeared counts (denoted by Exam_Cnt) and final average scores (Final_Avg_Score).

Step 6. Having obtained the set of initial centroids (C_s) to each of study_year_groups (s), we assume the initial clusters formed by singleton centroid points. Now, we perform clustering operation to obtain finally stable (settled) centroids in these groupings. This task is implemented by comparing distances of each non-seed points with existing centroids and re-assigning such points to that cluster, whose corresponding centroid is at minimum distance from these points. Each iteration of this step ends up in computing new centroids of each cluster by averaging the above three attribute values of all student instances inclusive of that cluster [1, 2].

Step 7. The above step is repeated until the convergence criterion is achieved, i.e., (there exists no change in the centroid points of the functional clusters in each of the study_year_groupings). For an input parameter, "k" representing the upper bound of nominations provided by a chapter, so that $k = k_s$ calculated proportionate to total examination registrations done in each study_year, ($s = 1-4$); top k recommendations are picked and displayed as the final scholarship nominations from that chapter in a specific academic season; the recommendation results highlighted in yellow as a result of implementation from one such run instance are exhibited in screenshot as in Fig. 3.

3.2 Impact Analyses

Also, one cannot negate the awareness and motivational factors as obvious reasons for vast variation in statewise engagement of MOOCs. One of the major motivation factors is anticipated to be adoption of credit transfer initiatives pioneered by some of the universities and AICTE affiliated government and private institutions. The states of Andhra Pradesh, Gujarat, Karnataka, Kerala, Maharashtra, Punjab, Uttar Pradesh, Uttarakhand, and West Bengal were the pilot local chapters who led the initiative of supporting and abiding by the credit transfer scheme according to the guidelines released under a gazette notification by University Grants Commission (UGC) and All Indian Technical Education (AICTE) in August 2016. Coincidentally, there was seen, an increasing rise in learners' engagement counts in some of the geographic zones of the country which had been leading the league to implement AICTE's credit

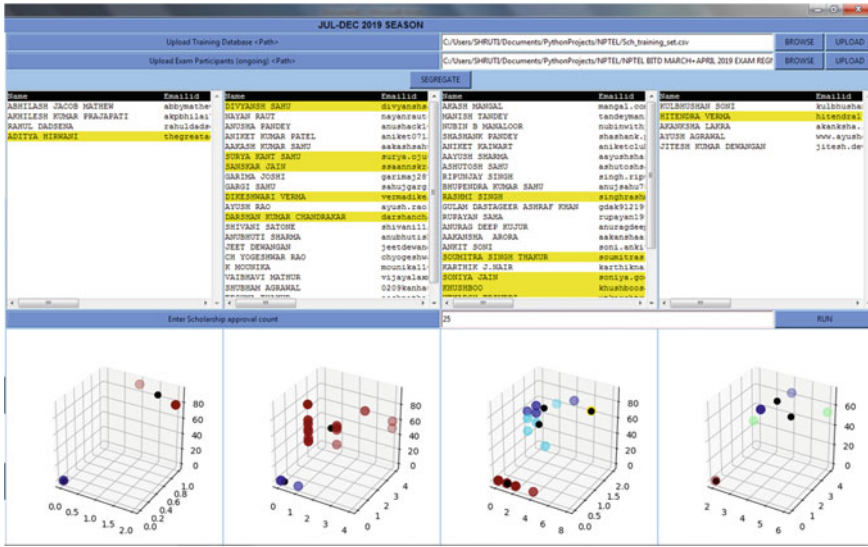


Fig. 3 Variant k-means clustering within four study year groups for NPTEL scholarship recommendation rankings

transfer initiatives. On analyzing upon their enrolled local chapter counts, active local chapter counts and top 100 active local chapter counts, a specific kind of hypotheses was formulated; (1) *AICTE credit-transfer initiatives do not have an impact on the performance of these states as “top 100 active local chapters”*; (2) as “*active local chapters*”. To resolve such queries, a chi-square test was performed on three numerical attributes: active (statewise) local chapter counts (academic sessionwise) and average top-100 local chapter counts (academic sessionwise) for these states, both these counts summing up to the third attribute: enrolled (statewise) local chapter counts (academic sessionwise). The χ^2 contingency table can be well illustrated in Table 1. The χ^2 statistic for tagging of average top-100 local chapters pertaining to

Table 1 Chi-square test for Indian states adopted with credit transfer initiatives from SWAYAM NPTEL

State	1	2	3	4	5	6	7	8	9	10	Sum(.)
Ox	1	2	3	4	5	6	7	8	9	10	104
Oy	14	2	4	5	10	1	30	8	0	30	559
Sum (LCs)	74	25	61	52	115	3	148	55	3	23	663
Ex	14	4	10	9	20	1	28	10	0	8	
Ey	74	23	55	48	105	3	150	53	3	45	
$\chi^2 = \sum(Ox - Ex)^2/Ex$	0	1	3.6	1.8	5	0	0.1	0.4	0	60.5	
$\chi^2 = \sum(Oy - Ey)^2/Ey$	0	0.2	0.7	0.3	1	0	0	0.1	0	10.8	

the ten mentioned states is 72.4, while that for tagging of active local chapters is 13.1; the p -value derived from these χ^2 values (with degrees of freedom = 9) rests upon the universal thresholding condition: $p < 0.05$ is the usual test for dependence.

For the hypothesis 1, as χ^2 value is greater than 27.877 (for significance level, $p < 0.001$), we are compelled to reject this NULL hypothesis and choose for alternate hypothesis of dependency, thereby concluding that the states that took a lead in implementing AICTE credit transfer initiatives did have a high impact on their excelling performance and tagging as “top-100 active local chapters”. For hypothesis 2, χ^2 value (13.1) provide p -value in the range [0.1, 0.2] approximating to 0.15 and hence supporting the hypothesis in partial truth.

Remarks. The notations for Observed (Top 100 Active LCs): O_x , Observed Active Only LCs: O_y , Expected (Top 100 Active LCs): E_x , Expected (Active Only LCs): E_y .

4 Frontiers Ahead

The resonating effect created by AICTE credit transfer scheme initiated by the ministry of human resource development (MHRD), Govt. of India compels nationwide intermediary-and-coordinating component of MOOCs stakeholder hierarchy, viz. the local chapters to carry a five-year foresight of evaluating the impact of the credit transfer scheme on MOOC learners. Local chapters need to engage themselves in self-evaluation tasks of exhaustive analyses within their boundaries for establishing themselves and retaining the “Active” tag among the list of competent peer participants to the scholarship national initiative. Last but not the least, the rising mentee figures of the local chapters demands a lot of improvement in the quality of mentoring system at the local chapters. No doubt, the delegatory mode of monitoring the progress of learners’ rests at the account of SPOCs; however, the delegated instructions are said to be well managed, if they reach every mentee in a serious tone and at stitch-in-time moment. Such an analytical framework is already into development stages and is sure to raise quality standards of a local chapter in appraising learners’ skills by being more punctual, more disciplined, more interactive with peers and course instructors.

Acknowledgements The work is an outcome of a couple of real-time data analytical experiments carried as a part of research and development activity, sponsored by NPTEL local chapter, Bhilai Institute of Technology, Durg, Chhattisgarh, India.

References

1. Han J, Kamber M, Pei J (2012) Data mining: concepts and techniques. Morgan Kaufmann Publishers, Elsevier, New York
2. Dunham MH (2006) Data mining: introductory and advanced topics. Pearson Education, Prentice Hall

Offline Voice Commerce Shopper Tracker Using NLP Analytics



Nikhita Mangaonkar, Vivek Venkatesh, and Maulika Arekar

Abstract Modern life offers a surplus of options for services and goods for consumers. As a result, peoples expenses and cost of living have been increasing day by day compared to that of a year ago. It has thus become essential to keep check on expenses in order to be sustainable in the long run. When we go shopping, make a list of items we want to purchase, but we always end up buying more than we needed. This disrupts the overall planned budget. It has become extremely important to keep a real-time check at the money we spend while shopping. Also in a fast-paced environment like a shopping market, it may be a bit frivolous to manually calculate everything. Hence, in order to ease the shopping experience, we developed an application called offline voice commerce shopper tracker. This application allows users to locate shop, set budget, create shopping list and provides them with chart analysis of their budget spending. The biggest draw of this application is that it can work offline, and creation and updating of shopping list can be done through their voice. As the Internet may not always be available, an offline solution will be highly appreciated.

Keywords Shopping · Tracker · Neuroscience · Neurology · Natural language processing · Flutter · Speech-to-text · Voice commerce · Analytic

N. Mangaonkar · V. Venkatesh (✉) · M. Arekar
Sardar Patel Institute of Technology, Mumbai, India
e-mail: vivek.v@spit.ac.in

N. Mangaonkar
e-mail: nikhita.mangaonkar@spit.ac.in

M. Arekar
e-mail: maulika.arekar@spit.ac.in

1 Introduction

1.1 Introduction

Shopping is done for essentials, and while other times, it is leisurely. But it is indeed helpful to set a budget if you ever plan a shopping trip. Expenses tracking are important for recording and analysing the incomes and expenses of a person over a period of time. We are living in a society where many people are always in a rush and are looking for efficient ways to manage their time and money. However, it has been noted that in most cases the management of the budget is mostly done mentally or manually on paper which makes expense tracking difficult. It may also lead to problems like calculation errors, mismanagement [1]. Shopping also has a more neuroscience reasoning for the justification of purchasing unwanted items. The human brain is geared towards recognition of patterns they perceive daily [2]. So when they see an anomaly in pattern, the brain responds with intrigue. When a customer sees a shelf in a shopping mart and notices an anomaly in the arrangement of the items on the shelf as compared to their previous visit, it leads to activation of various specific areas of the brain and also causes a release of dopamine. Also while shopping, in the brain, there is an efficient and rapid back and forth alteration between reward anticipation, pain, and mathematical thinking [3]. This fact is taken advantage of by stores by rearranging items in the store and also giving discounts to give a sense of satisfaction to the customer.

During uncertain times like a pandemic, it may be necessary to change our ways from traditional methods of list taking for example by hand to modern techniques. This gives rise to the need to have voice commerce. Voice recognition is a non-contact, non-intrusive and easy to use system. Arguably, this has become more important during the times of a pandemic, and voice authentication and voice to text have become the need of the hour. Voice commerce is the amalgamation of all the above features with the purpose of easing the process of shopping [4]. Voice commerce provides a real hands-free approach. It reduces the need of having physical contact with the devices we own. Due to development of technology voice querying has become swifter than it used to be. It also reduces the barriers by being more accessible and more convenient to use.

This shopping budget application serves the purpose of tracking and analysing the finances of individuals when it comes to shopping [5]. The application is developed for both the Android as well as the IOS users in mind, following the process of software development life cycle. To be more far reaching the user experience while using the app is really simplified. The main vision with this application is to cut the spending incurred down during shopping to only the bare necessities and encourage the users to develop a habit of being financially aware of their transactions.

1.2 Survey

For understanding the problem of overspending, we have conducted a survey. This survey was conducted among people from a varying age group and sex to get consumers feedback about their behaviour in relation to shopping and setting of budgets. We asked them the following questions and got the following answers (Figs. 1, 2 and 3).

We also asked our peers whether they make a shopping list before going out for shopping and with exception 10.8% the others agreed. But we figured out that 89.2% of them do not use any shopping budget apps but rather they would use Google

Do you set a budget before going to shopping?

66 responses

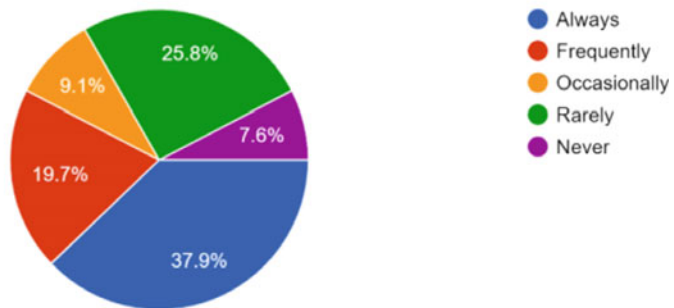


Fig. 1 Chart showing breakdown on how many people set a budget before shopping

By how much do you go over budget?(in percentage)

66 responses

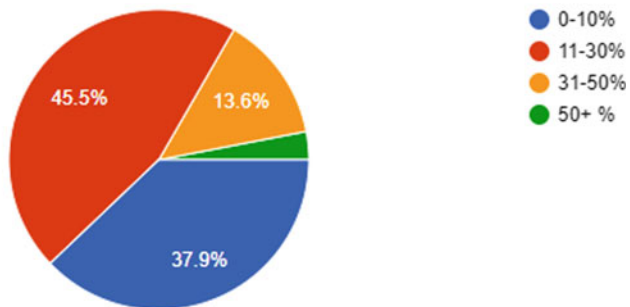


Fig. 2 Chart showing breakdown on how often people go shopping

Do you make a shopping list before going to shop?

66 responses

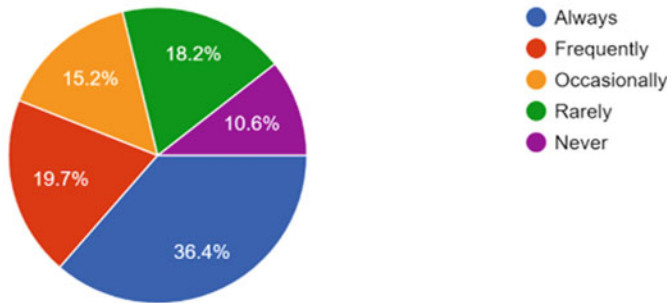


Fig. 3 Chart showing breakdown on how many people make a shopping list

Keep, traditional way of pen-and-paper, Notepad while some prefer to memorize their shopping list.

Based on the above survey conducted, we concluded that although people make a shopping list and also set their budget most of them to go over budget. This can be attributed to the methods of calculating their budget during shopping. As we have seen people are using static means like handwritten notes or note-taking apps which do not provide active feedback about the cost that they incur at the time of the purchasing the items but only when they checkout from the shop, they realize about their actual spending. A flaw with note-taking applications is that after each item you have to manually type out the purchased product, this process gets tedious and time consuming especially while in a shopping market.

We asked people if they would be interested in an application which will let them update their shopping list and budget in real time through their voice. The response was 60% will readily embrace the app, while 30.8% would like to try the app first, and the remaining 9.2% will not try the app.

When asked about why this application would prove to be useful, they responded it would save their time, money will keep track of all their spending, and it is also convenient and dependable. It would save them the time required in finding a store. It would help them to automate all things like manually updation of their shopping list and of performing quick math.

2 Methodology

2.1 Modules and Features

The offline voice commerce shopper tracker application is created using Flutter SDK. Flutter SDK is based on the Dart language and is used to create applications for IOS and Android devices at the same time using the same code base, and hence, the application maintenance is easier for both the platforms [6]. Material design by Google is used for designing the UI. Firebase and SQLite act as our NOSQL and SQL databases, respectively. Various concepts of AI like natural language processing are used. The model used for developing the application is an iterative waterfall model with phases like planning, analysis, design, implementation, testing taken into considerations.

Key Features of the Application are discussed below

Offline The main advantage of this app is that it can be used offline. The application takes the users current budget and then would allow the user to create a shopping list. This shopping list is stored locally so that there is no need to be connected to the Internet when updation happens onto the shopping list the budget is also updated without an Internet connection. Only time Internet is required is for locating shops and for analysis for large amount of data.

Voice driven The key feature of this application is that its voice driven. Users can add items onto the shopping list using their voice. They can say product name, quantity as well as price through their voice. To ease the process of searching, users can use their voice to search for shops and also to look through order history. Various NLP techniques are used to check the pronunciation, adding predicates, and spell check.

Robust/Real-time Feedback When a user decides to buy an item added into their shopping list, they can swipe right on their item to add it to the cart page. In doing so, the current spending of that visit is updated and that can be compared with the set budget for that trip. Whereas if they swipe left, then that item is removed from the list.

Visual Feedback of spending While purchasing items if the current spending would equal 25% of the set budget, then the colour of current budget icon will change to green. Similarly, when the current spending will equal 50% of the actual budget, the colour of current budget icon will change to orange. When the current spending will equal 75% of the actual budget, the colour of current budget icon will change to red. Also when you will go over budget, the colour of the budget will change to black, and a popup indicating the same will open.

Store locator The store locator helps in tracking the nearby stores upto a 10 km radius to your current location. By clicking on the icon of the store, you can see your purchase history for that particular location. This is achieved by using the GPS functionality of the smartphone. The store can also be located using your voice. When you say a name of a store, the application will narrow down the stores based on your query.

Order history/expense history The order history shows all of your previous visit to the particular store. This is shown along with the purchased items and amount spent at that location each visit. This information is also achievable using your voice. Just say the name of the store, and then, the application will sort out and show the past visits to the store in descending order.

Chart/graph analysis The data is analysed using charts and graphs. Three types of graphs are available, the first graph is a bar graph which shows the comparison between the three different shops from the past three months, second is a pie graph that shows the top six stores and the spendings, the last is a line graph that shows the budget Vs actual spending graph.

2.2 Methodology

The user first selects a store they may want to shop from. The list of available stores within a 10 km radius is displayed on the screen, and user can either scroll and select a store or can search for a store using their voice. They can see their previous purchases made from that particular store as well. Once a store is selected, a user is needed to provide their budget limit. After that, the user can create their shopping list. To create a shopping list, the user needs to either speak out the product name, quantity, price or can type it. If the user is purchasing the item, then they will swipe right on the item indicating that its purchased, and if they swipe left, it will denote that they have deleted the item. After every purchase, the spending gets updated, and according to the current spending Vs the set budget, the user is notified when the spending goes above a certain threshold of the budget. Once the shopping trip is over, the items are added onto the database with the entry for the visit to the shop. This data can then be accessed through the order history. Once sufficient data is collected, the ability to generate charts is also available. Based on the need, three different charts can be generated.

2.3 Algorithms

For the store locator and order history page, there is a feature to search the list of available stores and products through your voice instead of typing it out. It also provides a feature where we can search for shops through the products. Each shop is

considered as a predicate set that has certain predicates based on the type of product, this way when you search a product, it will see in which predicate set that is available in and accordingly show the shops that may serve those products. For example, milk will be a predicate in the classifier set for shop, dairy, and supermarket. Pen will be a predicate in a stationary supermarket. The classifier set is possible to be achieved due to the use of Google Places API that along with the details of the place such as name, location also provides us with information such as type which shows whether a place is a supermarket, shop, garage, etc.

Algorithm for voice-based searching in store locator by product name

1. Voice spoken is transcribed into text using the hidden Markov model [7] which is implemented using the Viterbi algorithm.
2. The product name is converted into lowercase
3. White spaces and punctuations are removed from the product.
4. Check the product name for misspelling by comparing it with the list of proper words.
5. If the product name does not exist in the proper words list, then find the most likely word. It could be calculated by finding the weight of the product and comparing it to the weight of the words in the list and the one with the most similar weight is the most likely word. Weight of the product can be found out by finding out how many times a letter occurs and in which position for all letters in the word.
6. Replace the word with the most likely word
7. We will now have the processed word.
8. Compare the processed word with every predicate in every predicate classifier set.
9. If the word is found in a predicates classifier set, then all stores where the store type equals the name of predicate classifier set should be displayed
10. Now, we search for a particular store to reduce the number of entries (Fig. 4)

For the shopping list when spoken, the conversion of information from voice to text happens with the help of NLP. The user can say there we use the speech recognizer package and implement it in flutter. First the word is converted from speech-to-text, and then, it goes through another process that corrects any misspells. The speech recognizer channel will wait for input. If some data is being passed, then the program will read the input and store it into a list format.

Algorithm

1. Select the store and then set the budget through the slider.
2. Set the language with language name and language code.
3. Set the Boolean listening variable as false.
4. Initialize the method for speech recognizer, call variable platform call handler.
5. Call the function activate recognition

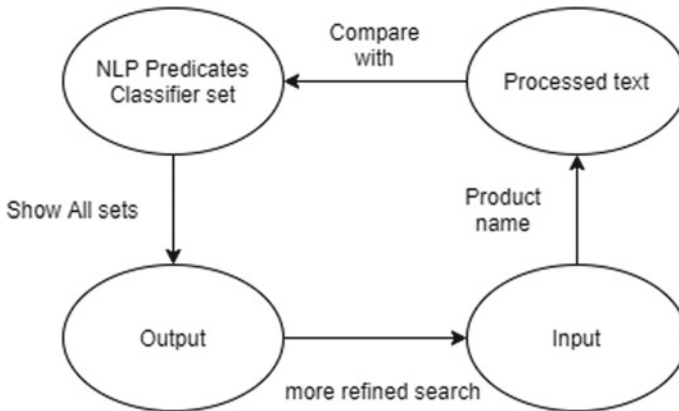


Fig. 4 Diagram showing how the search functionality works

6. Initialize method channel variable with output of speech recognizer method.
7. Await input from use
8. When input is received, change speech channel variable in speech class and set method call handler variable.
9. If input is “start”, then set method call handler as start.
10. Say the products name, quantity, and price
11. If input is “stop”, then set method call handler as stop.
12. If method call handler is stopped, then set variable listening to false.
13. If method call handler is set to stop, then append the list
14. In the transcription method if the list given is empty, then vocal data is converted into text data by following the algorithm and storing the value in the transcription element.
15. If the Todos list is empty, then append Todos list with the transcription element received.
16. If the Todos list is not empty, then pop the top element from the Todos list and then append the cart element, and update the current spending. Give an alert if overbudget.

3 Screenshots

See Figs. 5, 6, 7 and 8.

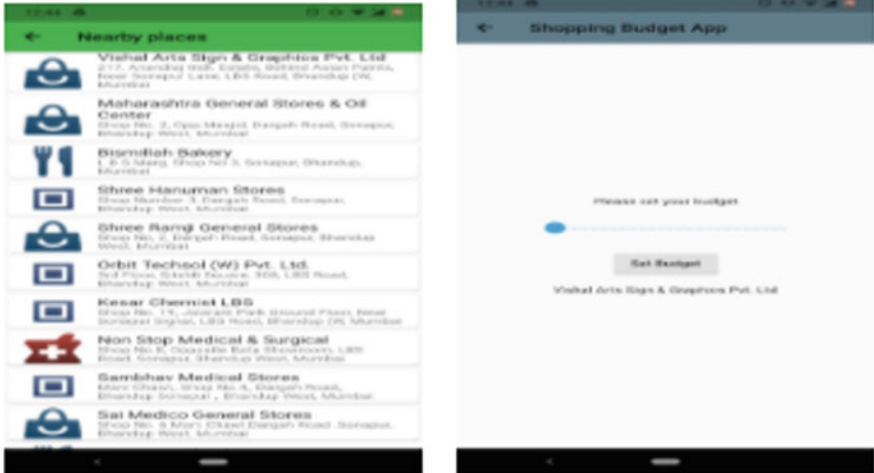


Fig. 5 Screenshot from the application showing shop locator and budget setting page

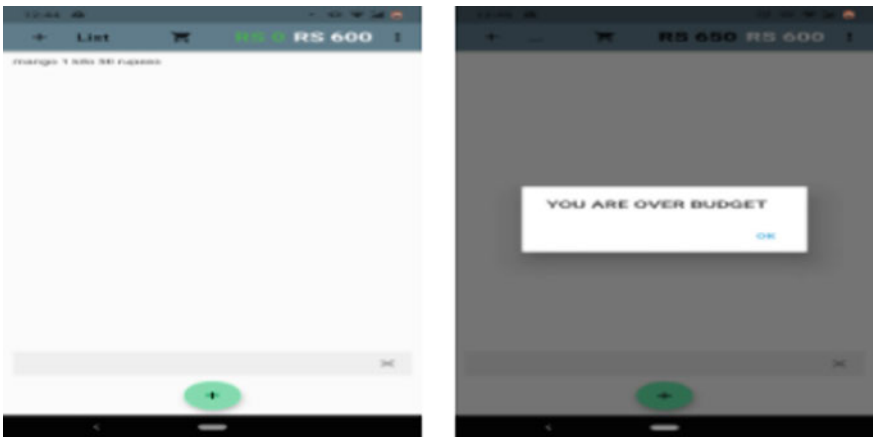


Fig. 6 Screenshot of shopping list page and over budget popup

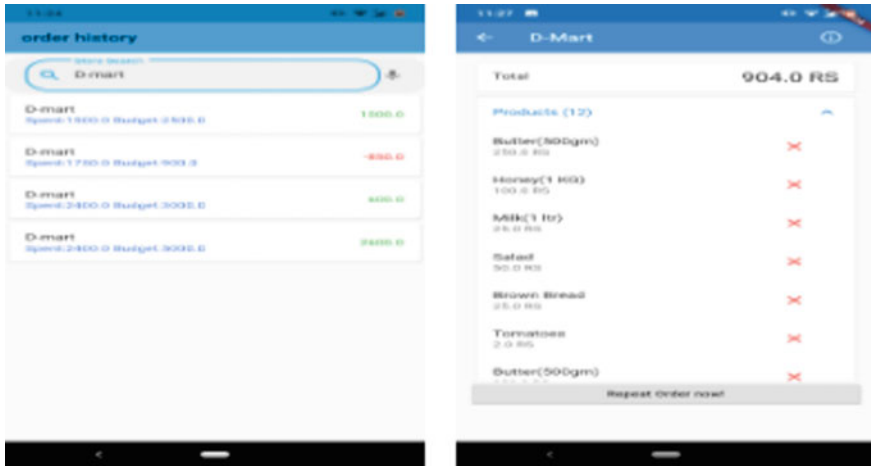


Fig. 7 Screenshot of shop searching in order history and also product search



Fig. 8 Screenshot of graphical analysis of budget and comparison between shops

4 Future Enhancements

The future plan is to provide more graphical analysis like comparing the cost of a particular item in various shops and determining which shop is best for which item. Create machine learning algorithms that will generate shopping list based on past purchases and spending habits. Also, it is possible to go a level deeper in categorization by sub categorizing into type of product to perform even more optimized search than what is currently possible.

5 Conclusion

We have developed a shopping budget application for Android and IOS which can be operated through voice. In this paper, we have discussed a solution for keeping track of budget rather than the contemporary way of making mental and sometimes erroneous calculations. Our applications deals with developing an application to overcome the shortcomings of the existing budget-tracking methods and at the same time curb the user from having the tendency to purchase over the necessary needed. This application will hopefully help in minimizing overspending and manage a person's expenses while on a shopping trip but also acts as guide to realize the cost of goods in different shops and help decide which store should be used. The user experience is also been kept in mind. Ultimately, the goal is that this application will be a one stop solution to keeping a track of shopping expenses.

References

1. Pennings JME, Itterssum, Wansink B (2005) To spend or not to spend? The effect of budget constraints on estimation processes and spending behavior. In: *Advances in consumer research*, volu 32, pp 328–329
2. Goldstein B (2016) The shoppers brain: what neuro science can teach us about customer behaviour. <https://www.credibly.com/incredibly/evaluating-capital-needs/shoppers-brain-neuroscience-key-understanding-customer-behaviour/>. Last accessed 15 May 2020
3. H. Moawad, Shopping and the brain, “<https://www.neurologytimes.com/view/shopping-and-brain>” , Last accessed 2020/06/24
4. Mari A (2019) Voice commerce: understanding shopping related voice assistants and their on brands. In: IMMAA annual conference, Northwestern University in Qatar, Doha (Qatar), 4–6 Oct 2019
5. Bekaroo G, Sunhalo S (2007) Intelligent online budget tracker. In: *Computer science and IT education conference*, Oct 2007, pp 111–124
6. Wu W (2018) React native vs flutter, cross-platform mobile application frameworks. Metropolia University of Applied Sciences, Information Technology, Mar 2018
7. Khilari P, Bhope VP (2015) A review on Speech-to-text Conversion methods. *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)* 4(7):3067–3072

Intelligent Content-Based Hybrid Cryptography Architecture for Securing File Storage Over Cloud



Neeraj Ghule, Aashish Bhongade, Ashish Ranjan, Rahul Bangad, and Balaji Patil

Abstract With the rise of computer and mobile usage, and availability of affordable Internet and free social media, huge amount of data is generated every minute. The data consists of personal and sensitive information about users, which could be their photos, videos, documents even their passwords, and what not! But there is a question regarding the security and privacy of their data. Nowadays, cryptography and steganography techniques are more popular for data security. Using only a single algorithm does not prove to effectively provide high-level security to the data. In the proposed methodology, hybrid encryption technique for the DOCX files has been used in which the input source file is divided into different parts and encoded using two symmetric cryptography algorithms RC6 and Blowfish and then these encrypted parts are stored over the cloud.

Keywords Sensitive data · Hybrid encryption · Blowfish · RC6 · Symmetric cryptography

N. Ghule (✉) · A. Bhongade · A. Ranjan · R. Bangad
Maharashtra Institute of Technology, Pune, Maharashtra 411038, India
e-mail: neerajghule80@gmail.com

A. Bhongade
e-mail: ashishcsb98@gmail.com

A. Ranjan
e-mail: Ranjan.ashishmit@gmail.com

R. Bangad
e-mail: rahul16398@gmail.com

B. Patil
Dr. Vishwanath Karad, MIT World Peace University, Pune, Maharashtra 411038, India
e-mail: balaji.patil@mitwpu.edu.in

1 Introduction

Cryptography is a technique, in which original data is translated into unreadable form called ciphertext. There are two types of cryptography techniques namely symmetric key cryptography and public key (asymmetric) cryptography. Both the techniques use keys to translate original data into ciphertext. Symmetric key cryptography algorithms use single key for encryption and decryption purposes. Some of the symmetric algorithms are AES, DES, 3DES, RC6, Blowfish, and IDEA. Public key cryptography uses two or more keys (public and private) for encryption and decryption purposes. Asymmetric algorithms are RSA and ECC. These algorithms have high-level security but there is increased delay in data encoding and decoding.

In many existing system, only a single cryptography algorithm is used. But, use of a single algorithm does not guarantee high security and efficiency. Use of a single symmetric key cryptography algorithm presents security problems because this type of algorithm applies a single key for data encryption and decryption. Also, there are various proposed systems in many research papers, wherein the similar security issues are dealt with. But the inside contents of the file are not analyzed, i.e. if the text contains sensitive information like password, bank a/c. number, Aadhaar number, etc. In our proposed system, the encryption of data (DOCX file) using hybrid cryptography is done based on the contents of the file.

This paper is organized into five Sections namely first section give the introduction, Sect. 2 about related work, Sect. 3 introduces proposed system architecture, Sect. 4 discuss experimentation & results and paper ends with conclusion in Sect. 5.

2 Related Work

In the proposed methodology, hybrid encryption technique for the text file has been used in which the input text file is first divided into normal and sensitive text and then encrypted using RC6 and Blowfish, respectively.

As in our existing project initially we have taken Docx file which contains different data types such as text images. So initially we are breaking this Docx file into parts. i.e., we are segregating it into text and images. And also we are dividing text into two parts sensitive and non-sensitive text. Sensitive text contains ATM pin, CVV, etc.

To propose this technique, we have taken different ideas from different research papers. The survey of related work is done on how to break the file in to blocks, which security algorithm should be used considering its merits and demerits.

Authors Maitri and Verma [1] proposed the idea of breaking the file into parts in their hybrid cryptography system. In which they have followed an approach where the original file to be encrypted (text file) is first split into eight different blocks and these blocks are encrypted by AES, RC6, BRA, and Blowfish algorithms. Each part of file is encoded simultaneously using multithreading technique. Encoded file is

then stored on the cloud server and keys which are to be used for encryption are stored into cover image.

Magdelin Jennifer Princy [2] in his survey of comparison between symmetric algorithms DES, AES, Blowfish, RC4, and RC6. It has been concluded that RC6 extends good operation in terms of security and compatibility compared to other algorithm. While the Blowfish is more secure and fast processing algorithm. It reduces the execution time and it gives better security and it consumes less memory usage compared to any other algorithms.

Rivest [3] developed RC6 algorithm and came up with the results: RC6 is a secure, compact, and simple block cipher. It delivers good performance with considerable flexibility. Furthermore, its simplicity would allow cryptanalysts to quickly improve over estimates of its security. Meaning RC6 has a flexible, structured can be modified easily like its predecessors RC4 and RC5. Ghorpade and Talwa [4] in their review have concluded Blowfish algorithm can be utilized for the transmission of the encrypted texts and images over the web. Blowfish algorithm cannot be broken effectively by the cryptanalysts until they locate the right mixes. This is more complicated to shape the accurate mixes of the lock.

3 Proposed Work

3.1 Methodology

In the proposed methodology, DOCX file is targeted. A docx file is scanned using Apache POI library, containing XWPFDocument class. User is prompted to opt for encryption or decryption. If encryption is selected, the path of source.docx file and the path of destination folder are asked. The destination folder will contain the extracted images, their encrypted form, a text file named SAMPLE.TXT that will contain the text from the docx file and ENCRYPTED.TXT file that will contain the encrypted text. While scanning, images are extracted first. The location of an image is stored in the form of paragraph number. Then, text is extracted and stored in a SAMPLE.TXT file. This text is scanned for keywords like: password, id, number (phone number, account number, Aadhaar number, etc.), OTP. These keywords are stored in string named SENSITIVE_TEXT. Rest of text is kept in NORMAL_TEXT. The location of images, in the form of paragraph number is also kept in NORMAL_TEXT. While storing these words in the two strings, their respective position is stored too. For example, if the text is: ‘Your otp is 411012’, this will be stored as:

NORMAL_TEXT: 1#Your 2#otp 3#is

SENSITIVE_TEXT: 4#411,012.

Newline character is stored too, in the normal_text string. Once all the words are segregated, NORMAL_TEXT is encrypted using RC6 algorithm, and SENSITIVE_TEXT is encrypted using Blowfish algorithm.

If ‘Decryption’ is chosen, the user is prompted to enter the path of the folder containing encrypted images and ENCRYPTED.TXT file. Then, user is asked to enter the location where.docx file is to be created. Images are decrypted and kept in the same folder. Then, algorithm scans the ENCRYPTED.TXT file and segregates it into NORMAL_CIPHER and SENSITIVE_CIPHER. These are decrypted using RC6 and Blowfish algorithm, respectively. The algorithm extracts the image locations from the resultant normal text. A loop is then used to place the words and images in their respective positions in the.docx file.

3.2 System Architecture

Figure 1 shows the proposed system. Cloud user uploads the data file on the cloud server. The file is split into different blocks of images and text. Text data is further categorized into normal and sensitive text. These blocks are then encrypted using different encryption algorithms. RC6 algorithm is used for normal text and Blowfish for images and sensitive text. Encoded files are stored on the cloud server. Cloud storage is the multiuser environment where more than one user can access files from a cloud server. Cloud user requests for file, on request of the file user gets access to key using email which consists of key information. Using this key, user can download the file; during this process, different files on cloud are decoded and merged again to the original Docx format file.

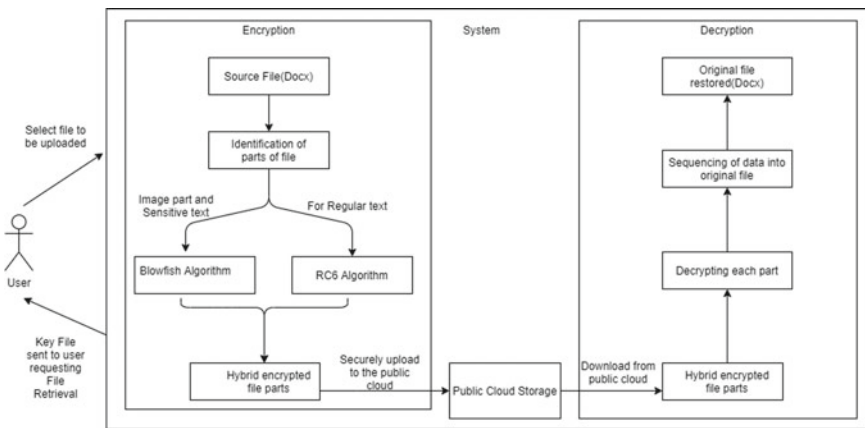


Fig. 1 System architecture

3.3 Existing and Proposed Algorithms

Here, we have explained existing Blowfish and RC6 and proposed hybrid algorithm.

Blowfish Algorithm

Blowfish is a symmetric key algorithm designed by Bruce Schneier. Blowfish provides strong encryption rate in software and no effective cryptanalysis of it has been found to date.

Blowfish algorithm is not subjected to any patents, and therefore, it is freely available for anyone to use.

Algorithm: Blowfish

Design specifications:

Key size: 32–448 bits Block size: 64 bits Rounds: 16 Structure: Feistel.

Working of Blowfish:

1. It is a 16-round Feistel cipher and uses large key-dependent S-boxes. In structure it resembles CAST-128, which uses fixed S-boxes.
2. There are five subkey-arrays: one 18-entry P-array and four 256-entry S-boxes (S0, S1, S2 and S3).
3. Action 1 XOR the left half (L) of the data with the r th P-array entry.
4. Action 2 Use the XORed data as input for Blowfish's F-function.
5. Action 3 XOR the F-function's output with the right half (R) of the data.
6. Action 4 Swap L and R.
7. The F-function splits the 32-bit input into four eight-bit quarters, and uses the quarters as input to the S-boxes. The S-boxes accept 8-bit input and produce 32-bit output. The outputs are added modulo 232 and XORed to produce the final 32-bit output.
8. After the 16th round, undo the last swap, and XOR L with K18 and R with K17 (output whitening).
9. Decryption is exactly the same as encryption, except that P1, P2...P18 are used in the reverse order.

Image Source: Internet.

RC6 Algorithm

RC6 is a symmetric key algorithm, successor to the RC5 algorithm, designed by Ron Rivest, Matt Robshaw, Ray Sidney, and Yiqun Lisa Yin. It is a fast and secure encryption algorithm.

RC6 is much similar to RC5 in terms of structure, use of data-dependent rotations, XOR operations, and modular addition, RC6 can be viewed as combination of two parallel RC5 encryption processes, although RC6 uses an extra multiplication operation which is not present in RC5.

Algorithm: RC6

Design specifications:

Key size : 128–256 bits Block size: 128 bits Rounds: 20 Structure: Feistel

Working of RC6:

1. Divide X into two blocks XL and XR of equal sizes. Thus both XL and XR will consist of 32 bit each
2. For $i=1$ to 16 $XL = XL \oplus P_i$ $XR = f(XL) \oplus XR$ Swap XL,XR (undo last swap)
3. $XR = XR \oplus P_{17}$
4. $XL = XL \oplus P_{18}$

The decryption will be identical to encryption algorithm step by step in reverse order.

Image Credit: Ref. [3]

Hybrid Algorithm

Hybrid algorithm is a proposed algorithm in which the input file is split up into different parts. Firstly, as text and image. Further, the text is classified into sensitive and non-sensitive data, which is further encrypted by suitable encryption algorithm. Hybrid algorithm is the combination of RC6 and Blowfish algorithm.

Algorithm: Hybrid Algorithm Input: Docx file

Output: Encrypted file

Step 1: Ask the user for the length of blowfish key (128 bits, 256 bits or 448 bits).

Step 2: Ask the user if he/she wants to perform encryption or decryption.

Step 3: If encryption is selected:

Step 3.1: Input the location of docx file along with its name.

Step 3.2: Insert the output folder location.

Step 3.3: The image files from docx file are extracted along with their locations in terms of paragraph number and encrypted using blowfish algorithm.

Step 3.4: Text from the docx file is extracted and kept in sample.txt file.

Step 3.5: This sample.txt file is scanned for sensitive data like IDs, passwords, OTPs, numbers (like Account number, Telephone number, Aadhaar number, PAN number).

Step 3.6: Two strings are created, normal text and sensitive text. The sample.txt file is given as input.

Step 3.7: Every word of input string is analyzed and is placed either in normal text or sensitive text along with its location (in terms of number of words before the current word).

Step 3.8: The image locations (paragraph numbers) is appended to the normal text String.

Step 3.9: The normal text is encrypted using RC6 and sensitive text is encrypted using blowfish encryption algorithm.

Step 3.10: An EncryptedFile.txt text file is created in the specified path of the output folder and the normal cipher as well as the sensitive cipher is written in it.

Step 4: If decryption is selected:

Step 4.1: Input the folder location containing encrypted image files and Encrypted File .txt.

Step 4.2: Insert the output folder location.

Step 4.3: The Encrypted .txt file is split into normal cipher and sensitive cipher and is decrypted using RC6 and blowfish respectively.

Step 4.4: Image location (in terms of paragraph) is extracted from the normal text.

Step 4.5: A docx file is created in the specified output folder. A FOR LOOP is used to select the words of normal and sensitive texts and to keep a track of paragraph numbers.

Step 4.6: A blank string 'str' is created.

Step 4.7: Continue this until the length of normal text + sensitive text is exhausted

4.7.1: Check if the current paragraph matches with the paragraph number of image. If yes, create a new paragraph in the docx file and add the image to the docx file.

Step 4.7.2: Check if the selected string contains a newline character. If not, add the word to the string 'str'.

Step 4.7.3: If the selected word is a newline character, write the contents of 'str' to the docx file and reset str to blank.

Figure 2 shows the working of the algorithm step wise as mentioned in the above algorithm.

4 Experimentation and Results Analysis

In this section, we have explained the experimentation setup and result analysis on various parameters.

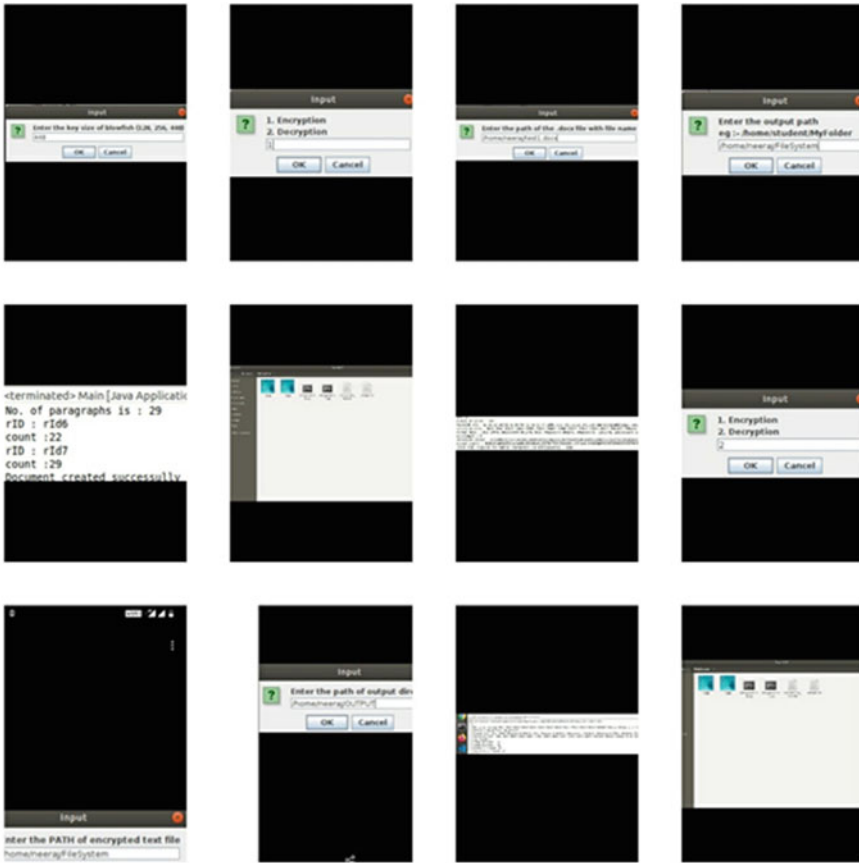


Fig. 2 System working

4.1 Experimental Setup

Experimentation is carried out on Ubuntu 16.04 OS, with front end IDE: Eclipse Oxygen and online support softwares like Google Docs used for creating performance graphs and plot, Draw IO(website) to develop the System Architecture image and Amazon Web Service(AWS) to store the data along with external libraries used: Apache POI, XWPF and Openxml4j.

Test cases considered for result analysis:

- (i) Document file docx containing image(s)
- (ii) Document file docx contains text as well as images at different locations in file
- (iii) A docx file containing document starting with image(s)
- (iv) A docx file containing document ending with image(s)

- (v) A docx file containing document image in between text
- (vi) A docx file containing document with multiple images and text

4.2 Result Analysis (Encryption & Decryption)

The graph in Fig. 3 shows the comparative analysis of hybrid algorithm versus RC6 versus Blowfish for encryption time for files (Docx) of different sizes. And it can be deduced from the graph that the system takes consistent time for encryption till the total file size is 980 byte, but after doubling the size to 1.5 Kb the time taken takes a sharp rise and this trend is also observed for larger file size values.

The graph shown in Fig. 4 is the comparative analysis of hybrid algorithm versus RC6 versus Blowfish for decryption time for different file sizes (Docx). It is observed that the decryption time increases gradually with file size, as the size of the ciphertext in the larger files needs more time to be decoded to become plaintext (original).

From above two graphs, it can be concluded that that Blowfish algorithm provides the strongest security but at the expense of more time, whereas RC6 takes the least time but is not strong enough compared to the Blowfish. Our hybrid algorithm combines the strengths of both the algorithms very smartly and gives optimum performance.

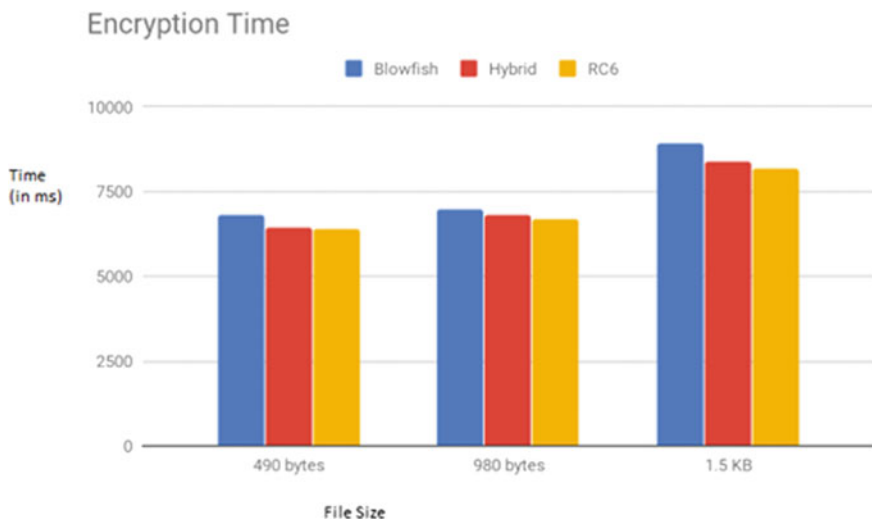


Fig. 3 Encryption time

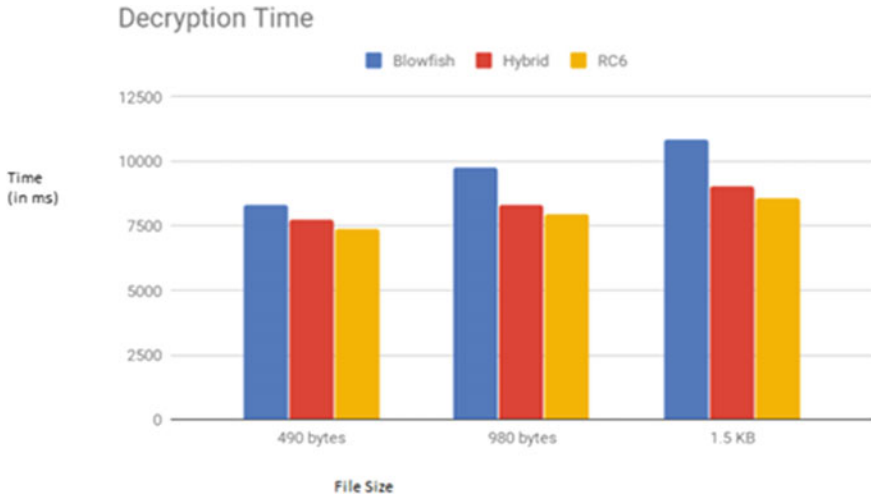


Fig. 4 Decryption time

4.3 Scatter Plot

Scatter plots are used to display values for typically two variables for a set of data to identify the relationship between two quantitative variables.

In this scatter plot shown in Fig. 5, performance evaluation of the system (encryption/decryption) is done, where in the original file contains only text. It can be inferred

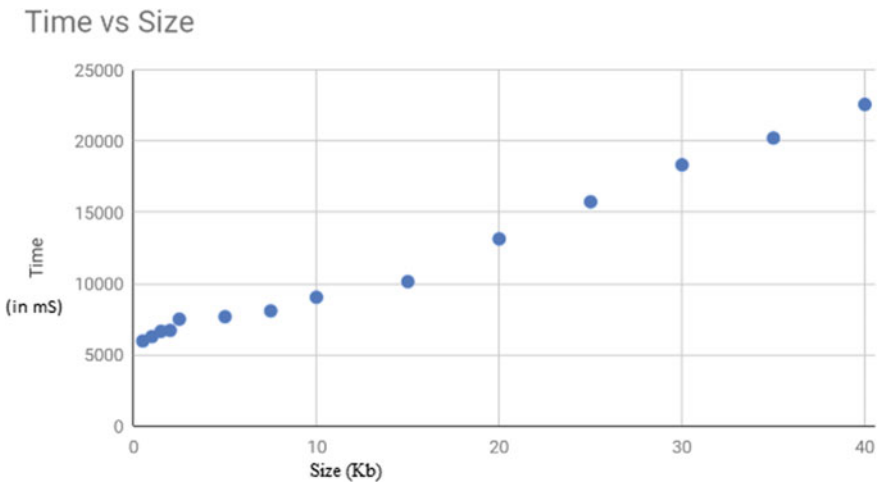


Fig. 5 Time versus size comparison plot

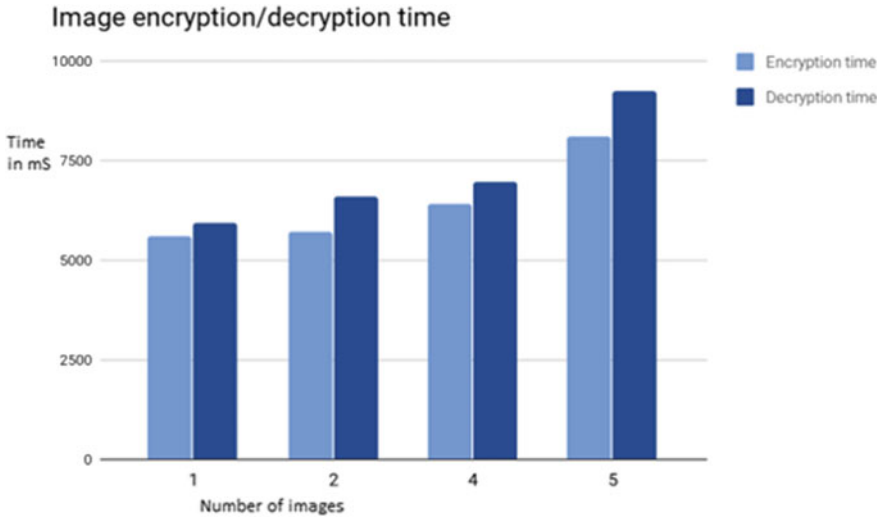


Fig. 6 Image encryption and decryption time

that for the file size till 15 Kb, the time gap for every 5 Kb of text there is little time growth but after 15 Kb it is observed that for every 5 Kb increase in file size the time taken growth increases linearly.

4.4 Image Encryption and Decryption Time

Figure 6 shows the analysis of time required for encryption and decryption of file (Docx) which contains only images which has been achieved using Blowfish algorithm. The size of each image for this graph is 150 Kb and it was observed that time taken by the system gradually stays consistent till four images but after that more time is taken, system performs well when number of images are 5 or less, this trend has also been observed for different image files of different sizes.

4.5 Blowfish Versus Hybrid Versus RC6 Output Size Comparison

Table 1 shows the analysis of Blowfish versus hybrid versus RC6 cryptography algorithms output file size comparison to analyze the impact of overheads. It gives

Table 1 Blowfish versus hybrid versus RC6 output size comparison

Blowfish		Hybrid		RC6	
Plain	Cipher	Plain	Cipher	Plain	Cipher
490 b	2.2 Kb	490 b	2.2 Kb	490 b	3.7 Kb
980 b	4.4 Kb	980 b	4.4 Kb	980 b	7.9 Kb
1.5 Kb	6.7 Kb	1.5 Kb	6.7 Kb	1.5 Kb	12 Kb

the comparison between hybrid, RC6 and Blowfish algorithm for the output cipher text when they are given the same input plaintext to each. And it is observed that hybrid algorithm has identical results as Blowfish and gives output cipher text in minimal size compared to the RC6 algorithm.

From Table 1, above graphs and plots we can say that our hybrid algorithm combines the strengths of both encryption algorithms which is strong encryption strength and lesser sized output of Blowfish and lesser execution time of RC6.

5 Conclusion and Future Work

In our proposed system, we have solved the security aspect of encrypting the Docx file by using hybrid cryptography coupled with the text prioritizing system. Using the entrusted Blowfish and RC6 algorithms for encrypting the data along with our text prioritization system. We have obtained the results as: For encryption, 4.9% less time is required compared to Blowfish algorithm which has benchmark security but slower execution and 56.9% less space than RC6 which is faster but takes more space a comparatively than Blowfish and our hybrid algorithm.

In future, more advanced symmetric or asymmetric cryptographic algorithm can be implemented to ensure more data security from any malicious activity. Image Processing can also be applied to identify and extract crucial data like ATM pin, OTP, Account number, Aadhaar number, PAN number from images and encode them into separate files using even stronger algorithm.

References

1. Maitri PV, Verma A (2016) Secure file storage in cloud storage using hybrid cryptography algorithm. In: International conference on wireless communications, signal processing and networking (WiSPNET), 20–22 June 2016
2. Magdelin Jennifer Princy P (2015) A comparison of symmetric key algorithms DES, AES, BLOWFISH, RC4, RC6: a survey. *Int J Comput Sci Eng Technol (IJCSET)* 20–22

3. Rivest RL, Robshaw MJB, Sidney R, Yin YL (1998) The RC6 block Cipher. In: 1998 M.I.T. Laboratory for Computer Science, 545 Technology Square, Cambridge, 1–3 Feb 1998
4. Ghorpade AM, Talwar H (2016) The Blowfish Algorithm simplified. Int J Adv Res Electr Electron Instrument Eng (IJAREEIE)

LegitANN: Neural Network Model with Unbiased Robustness



Sandhya Pundhir, Udayan Ghose, and Varsha Kumari

Abstract One of the widely used models in Machine Learning is an Artificial Neural Network. Ingenuous and legitimate results are the utmost need of every model. These models work like a black box, and no factor as trust is considered. Machine Learning algorithms are stuck with the risk of relying on biases encoded in the data or algorithms used for predictions. Prejudice or Bias is the root cause of the poor performance of the model. A novel approach is proposed, named legitANN, which has used three levels of bias control and an AvgNew loss function to make the model legitimate and reliable. Median Absolute deviation is a statistical measure that can guarantee the learning of neural networks to achieve unbiased robustness. Median Absolute deviation used here to prove the effectiveness of the method legitANN. The work proved that feed forward neural network is better at approximating. But for uniformly dense sets or preprocessed data, a radial basis network may be a right choice. This paper presents legitANN a new model that work on bias control using radial basis network and achieve better performance in comparison to other Artificial Neural Network in terms of improved evaluation metrics values for link prediction.

Keywords Artificial neural network · Bias variance · Customized artificial neural network learning · Robustness · Fairness learning

S. Pundhir · U. Ghose
University School of ICT, Guru Govind Singh Indraprastha University, New Delhi, India
e-mail: Sandhya.pundhir@gmail.com

U. Ghose
e-mail: Udayan@ipu.ac.in

V. Kumari (✉)
Department of CEA, GLA University, Uttar Pradesh, Mathura, India
e-mail: Varsha.kumari@gla.ac.in

1 Introduction

Machine Learning (ML) is a pervading technique that is defined as the computational machine having the capability to learn and search patterns among the input data to predict the needed output. Artificial Neural Networks (ANN) includes various technologies like Machine Learning (ML) and Deep Learning (DL), as a part of Artificial Intelligence (AI). Different architectures of ANN work differently using the various principles in determining their rules, and each has its merits and demerits as per Dey et al. [1]. The application of ANN is not limited to one fields as per Sandhya, et al. [2]. ANN can be applied to any problem area for which there is no known solution or problem area where the solution involves a lot of parameters. The neural networks model can be used for different types of learning. One of its uses is as a classifier, the input nodes will take the input features, and the output nodes will depicts output classes. And another application is a function approximation based on input and gives value into output node as per Sandhya et al. [3].

Learning algorithms have the risk to reckon on various biases as per Romanov et al. [4] as per Belinkov et al. [5]. Statistical Bias and variance relation can highlight the problem in ML bias. Large scale voting and error-correcting codes require colossal memory and CPU (Central Processing Unit) time. Many works on training machines with learning systems that give fair decisions have defined several important measurements for fairness, such as Demographic Parity, Equality of Odds, and Equality of Opportunity as per Zhang et al. [6]. Algorithmic Bias should be taken as a standard metric for evaluating data-based models as per Smith et al. [7]. A metric is a precise quantitative way to quantify the unfairness concept as per Zhang et al. [8] as per Díez et al. [9].

ANN can be implemented in different domains with different architecture and customization. It consists of an arbitrary number of essential computing elements called nodes placed on parallel and independent layers. Each node is associated with some weight. The bias node is like a particular node added to each layer in the neural network, and it mostly stores the value as one. The bias node is analogous to the offset in linear regression and is represented as $y = mx + c$ where “ m ” is called slope and “ c ” is the constant value or coefficient of independent “ x ”. The role of Bias in the prediction can be crucial. Bias allows the line to move up and down to be a good fit as per the prediction with the given data. Without ‘ c ’ the line always passes through the origin (0,0), and you may get a more inadequate fit. The Bias in neural network nodes is not equivalent to the threshold of a perceptron, which only outputs ‘1’ if sufficient input is supplied. So the role of Bias is not to act as a threshold, but to help ensure the output best fits the incoming signal. Adding a bias neuron without a changing value allows you to control the behavior of the layer. Bias takes any number and has the activity to shift the graph shape or the output, thus representing more complex situations. The primary function of Bias is to give a constant value to a node and the other inputs received by the neural network node.

The importance of bias value is that it enables activation function to either left or right, for training the neural network. But from ML’s point of view about bias

was like giving importance to specific features or groups to generalize better. Bias in ML makes us generalize satisfactorily and make the model less sensitive to a single data point. The concept of Bias, which is central to both facts: how ML models are designed and how we put trust in output produced by these models. Bias and variance reduction must be a combined effort to reduce mean error as per Thomas et al. [10]. This can also be a guiding factor to control Bias here. Accepting the possibility and reason of Bias is the need to mitigate Bias. Users of ML algorithms should audit for Bias.

Regularization is one of the processes of introducing additional information to prevent over fitting and also plays an important role in bias prevention. By reducing the poor performance or failures, ML becomes more robust. Bias-free, robust, multi-level and multi perspectives characteristics are considered here to make any ML model working in a true sense complete and optimized.

2 Related Work

To make ML models more robust and bias-free is the main aim of every optimization. Cross-validation is not a silver bullet. It is the best tool available so far because it is the only nonparametric method to check for generalization as per Raschka [11]. To address algorithmic Bias problem, model-validation processes should be updated to ensure appropriate algorithms are selected in any given context. Another approach is to develop ‘challenger’ models, using alternative algorithms to benchmark performance as per Fomin [12]. Data acquisition and use must also be managed by keeping a balance of system performance with transparency on how AI systems make their predictions about diagnosing Bias in machine-learned classifiers by examining performance on the testing set clusters, as said in Thomson et al. [13]. There may be a delicate balance needed between fairness and preferences—for instance, supposing that users are more likely to connect with others of similar age or experience level, one needs to ponder whether to recommend people of different ages to people of different ages as per Bird et al. [14]. Generalize the methods to be less dependent on individual words so that we can more effectively deal with biases tied to terms used in many different contexts as per Dixon et al. [15].

Algorithms seeking to reduce the risk and difficulty of adverse outcomes, the promotion and use of the mitigation method can create a pathway toward fairness.

To mitigate the risk and complications of adverse outcomes, these are five ways as per Niedoba et al. [16]:

- *Use statistical calibration:* Like resample or reweigh data to reduce Bias.
- *Use a regularizer:*-fairness regularizer (mathematical constraint) with ML algorithms.
- *Use surrogate models:* Wrap a fair algorithm around baseline ML algorithms.
- *Use fair ML models:* Adopt completely new ML algorithms that ensure equitable outcomes.

- *Calibrate the threshold*: Calibrate the prediction probability threshold to maintain fair outcomes for all groups with protected and sensitive features.

There are several useful measurements for fairness as per Zhang et al. [6] as per Acharyya et al. [17] as per Díez et al. [9] are such as Individual fairness, Group fairness, Demographic Parity, Equality of Odds, and Equality of Opportunity.

Bias can be broadly classified as follows as per Sebastian Raschka [11] as per Fomin [12] as per Thomas et al. [10, 18] as per Smith et al. [7] as per Acharyya et al. [17]:

1. *Statistical Bias*: A Statistical Bias is the conscious or unconscious weightage of one group, parameter, or output over other groups or outputs in the data. There are two critical types of Bias: Selection bias and Response bias.
2. *Prediction bias*: Today the concept of machine bias is often associated with data used to train ML systems: “programming that assumes the prejudice of its creators or data,” as said in [17]. It is the error that occurs, and that is difference between the expected prediction accuracy of a model and its exact prediction accuracy [16]. Human’s direct involvement makes a model more Bias.
3. *Algorithmic based*: Algorithm get biased on some attributes even if we omit those. They do this by learning the underlying representation of those features from other provided features.
4. *Data based*: Suffer in following form of Bias like (i) Sample Bias, (ii) Prejudice Bias, (iii) Group attribution Bias, (iv) Confirmation Bias, and many more.

It is adequately highlighted that uneven distribution affects model training and model predictions. It also pinpointed that the performance of the ML algorithm depends on the nature of the dataset. Trust must be established by some process or parameter on the working of the model to be unbiased and robust. A necessity of such model is there which can establish the trust and reliability.

3 Proposed Work: LegitANN

Just as we expect a level of trustworthiness from human decision-makers, the same way level of reliability is expected from our models. This model is more robust and overcome a variety of bias issues. We have taken a supervised ML classifier for testing the accuracy of different model; primarily we have used algorithm legitANN mentioned below.

Our method mitigates Bias by using evolutionary training as the strategy. Using appropriate mathematical or statistical procedures ensures the minimized risk of errors is and prevents discriminatory effects. We utilize AvgNew loss method for measuring fairness and three popular methods of bias correction as per Acharyya et al. [17] as per Díez et al. [9]:

- *Bias check before modeling*: is also known as preprocessing, involves processing the data to detect any bias and mitigate unfairness before training any classifiers.

Here we have converted the input matrix into an upper triangular matrix and preprocessing to make it more Bias free.

- *Bias check during modeling*: also known as in processing using the loss function of the classifier. MAD(Mean Absolute Deviation) is used.
- *Bias check after modeling*: also known as Post-processing, which makes the classifier fair under the measurement of a specific metric. Here Used AvgNew loss method to check Bias.

Algorithm legitANN Steps

1. Take input data and use a bias check before modeling. If needed minimize Bias, move to the next step.
2. Uses a random sampling method to choose a sample and apply a bias during minimize Bias, moves to the next step.
3. Split chosen sample into the train and the test sets.
4. Choose RBF ANN architecture, as discussed above.
5. Apply Bias check after modeling if needed minimized Bias, move to the next step.
6. Check the accuracy of output.

This algorithm works with the level of bias control and statistically proves its correctness over other ANN models used here. The algorithm is implemented in the next section of the paper, and the results obtained are tabulated and discussed. When these Bias mitigating steps are used with other configurations also, it gets improved. But the best improvement was achieved using the RBF with two layers. So with hit and trial, we made legitANN using RBF and that Bias controls steps.

4 Data Sources and Evaluation Methods

Datasets used here cover observable types of networks. The data sets used here are open-source categories and readily available for research work. Datasets used are the collection of nodes and edges to the connecting node, if any, exist there.

4.1 Data Sources

Detailed information about the datasets used here are as described in as per Sandhya et al. [19].

1. C-elegans-frontal:-It has many Nodes as much as 297 and many vertices as many as 2148. It is a neural network of C. Celegans.
2. karate:- It has 34 nodes and 78 edges. It is a social relationship network of 34 members of the karate club.

3. **USAir**: It has 332 number of nodes and 2126 number of edges. It is about domestic flights of the US, and Nodes are the US airports, and edges are about air travel connections.

4.2 Evaluation Method

The various methods used for evaluation are mentioned in the above data sources and evolution section. As mentioned above, we will use those evaluation methods, and the value returned is in the range 0–1. Zero indicates poor results and no match and value one indication of the perfect match. These methods are used to measure the quality of results obtained by various similarity measures. These are as follows using these standard terms, let x and y are the pairs of observations at one time. Y is the actual output, T is the expected output, and n is the dataset's size.

1. *AvgNew*: A modified cross entropy method proposed in [19].

$$AvgNew = \Sigma(-T * \log(Y) + (1 - T) * \log(1 - Y)/(ll * \log(ll))) \quad (1)$$

2. *Mean Absolute error (MAE)*: is the difference among two continuous variables where two variables let x and y are the pairs of observations.

$$MAE = \sum |T - Y|/r \quad (2)$$

3. *Mean*: In statistics mean is the average of a set of data. Mean is calculated by take the sum of all the number and then divide by the number of items in the set.

$$Mean = 1/n \sum x \quad (3)$$

4. *Cross-Entropy (CE)*: It is used in optimization. It can easily estimate small probabilities accurately. It is used to measure classification performance.

$$CE = -T * \log(Y) - (1 - T) * \log(1 - Y) \quad (4)$$

5. *Elapsed Time*: Time taken to get the final output. Less the time taken, fast the result obtained. But sometimes a bit more time taken is also considered as not bad.

$$ET = FinishTime - StartTime \quad (5)$$

6. *MAD*: Mean Absolute Deviation is the average distance between each data point and the mean. It gives us an idea about the variability in a dataset. Where X_n is the mean value. It can also refer to the population parameter that is estimated

by the MAD calculated from a sample

$$\text{MAD} = \left(\sum |x_i - x_m| \right) / r \quad (6)$$

7. *MSE*: is the mean of the squared difference between your estimate and the

$$\text{MSE} = 1/n \sum |T - Y|^2 \quad (7)$$

Here a wide range of evaluation methods are used which focus on different aspect of the prediction.

4.3 Why Use These Evaluation Methods and Dataset

In this section evaluation method's strength are highlighted in the light why and how these are used.

- *MAE*: is the average error measure and is unambiguous. MAE error measures are better for dimensioned assessment and comparison. MAE is more robust to error.
- *Mean*: is the average error indicator of the error of all cases.
- *Cross entropy*: shows accounting of better adaptation of the self-correction model. Cross entropy are used when you do not have good data distribution. The log function is a non-linear magnifier.
- *Time*: A indicator of the good or poor performance of algorithm.
- *MAD*: This metric is used as a robust statistic here. Robust statistics are statistics with good performance for data drawn from a wide range of probability distributions, especially for distributions that are not normal. It is helpful as it let us understand the data property such as the spread of data. Lower is the MAD smaller are the chances of deviation.
- *AvgNew*: It is the modified cross entropy using the size of the dataset also and making it more data-dependent.
- *MSE*: magnifies the error. The error is higher and gets even on a single higher error in the dataset.

USAir and C-elegans-frontal datasets have a larger number of nodes than karate dataset. By choosing these data sets a small size and large dataset size, applicability of the proposed work is proved. A good range of applications of the proposed method can also be proved.

5 Experimental Work and Results

The above-proposed work is implemented we using MATLAB 2016 on Mac Book Air with 1.3 GHz Intel Core i5, memory 4 Gb, and 125 GB HD. All the work implemented using MATLAB for all types of neural network experimented here. Every experiment is run five times and then average of that is noted. Here pre processing steps as discussed in legitANN is also used for all five architectures.

Configurations or type of ANN used here are as follows:

- I. *CFFANN*: Customized FFANN with (bias) initialization function. The most common weight and bias initialization function is *rands*, which generates random values between -1 and 1 .
- II. *TDFANN*: It is tailored FFANN as discussed in [3].
- III. *Cascadeforwardnet*: A type of FFANN include a connection from the input and every previous layer to following layers.
- IV. *Pattern-net*: It is another type of FFANN. The output data should be consist of vectors of all values zero except for the class they are correct.
- V. Radial Bias network (RBF): Radial basis networks can be used to approximate functions. newrbe very quickly designs a radial basis network with zero error on the design vectors.

Tables 1, 2 and 3 are having values of important metrics and different ANN configuration. Table 1 shows various ANN for link prediction using C-elegan-fronatal dataset. CFFANN and TDFANN have similar results. Cascade-forward and patterned

Table 1 Various ANN for link prediction using C-elegan-fronatal datasets

Metrics/ANN types	I	II	III	IV	V
MAE	0.095	0.08	0.11	0.112	0.004
CE	0.095	0.09	0.11	0.113	0.0037
MSE	0.048	0.138	0.09	0.092	0.0035
AvgNew	0.054	0.014	0.028	0.027	0.009
AvgNn	0.134	0.137	0.27	0.262	0.096
Time	0.51	0.45	0.85	0.82	0.6

Table 2 Various ANN for link prediction using Karate datasets

Metrics/ANN types	I	II	III	IV	V
MAE	0.123	0.173	0.098	0.08	0.0194
CE	0.123	0.174	0.1003	0.09	0.018
MSE	0.062	0.165	0.089	0.084	0.019
AvgNew	0.0227	0.017	0.03	0.021	0.07
AvgNn	0.171	0.127	0.226	0.22	0.56
Time	1.45	1.32	0.99	0.71	0.7

Table 3 Various ANN for link prediction using USAir datasets

Metrics/ANN Types	I	II	III	IV	V
MAE	0.144	0.234	0.138	0.135	0.0035
CE	0.144	0.225	0.138	0.136	0.0029
MSE	0.091	0.241	0.145	0.131	0.00035
AvgNew	Inf	0.0113	0.068	0.025	0.0007
AvgNn	Inf	0.131	0.263	0.298	0.051
Time	0.62	1.25	1.32	0.808	1.04

are having similar results. RBF the V configuration has more better result as compare to other four here in terms of better MAE, CE, MSE, AvgNew, AvgNn but little more time in execution.

Table 2 has various ANN for link prediction using Karate dataset, RBF the V configuration has more better result as compare to other four here in terms of better MAE, CE, MSE, AvgNew, AvgNn and similar execution time. Table 3 shows USAir dataset based link prediction by different ANN used here. RBF the V configuration has more better result as compare to other four here in terms of better MAE, CE, MSE, AvgNew, Avgnn but little more time in execution.

From these results of Tables 1, 2 and 3, it is clear RBF works better for all three dataset used here and in terms of all metrics used here except time. Figures 1 and

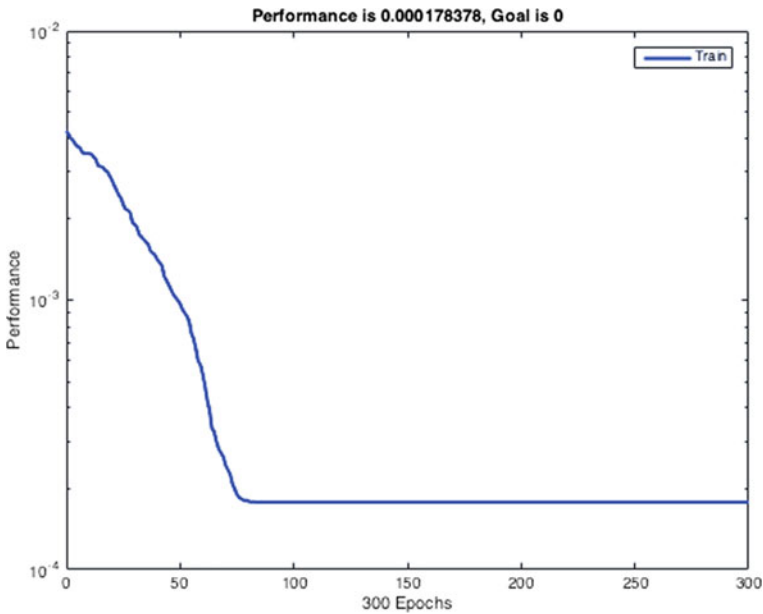


Fig. 1 USAir database based training efficiency of legitANN

2 shows the training performance of legitANN using USAir and C-elegant-frontal datasets. Table 4 has before training results of metrics and after training highlighting the bias controlling effects. It can be easily elaborated through comparing results. Less the metrics value lesser is the error. MAE, MAD and AvgNew got improved or less value after the training. Less value obtain for Mae and AvgNew in all three

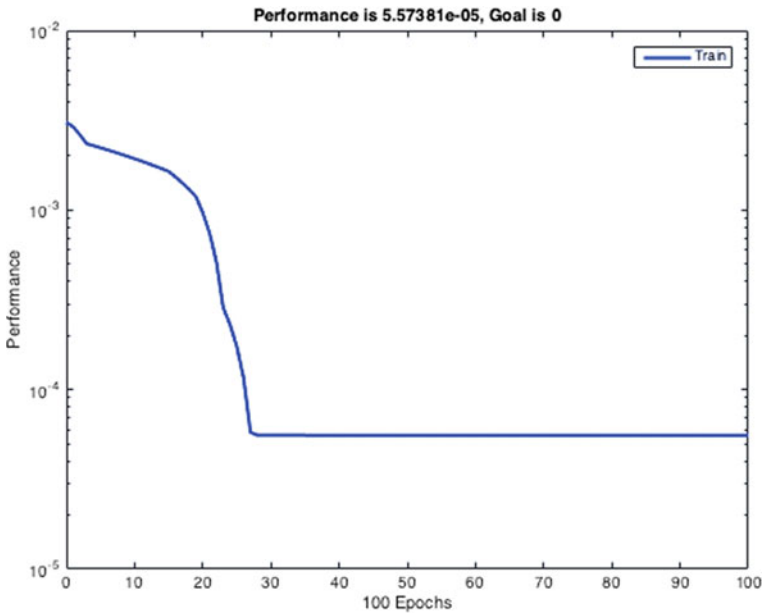


Fig. 2 C-elegans-frontal database based training efficiency of legitANN

Table 4 LegitANN model for link prediction using above datasets

	Metrics/LegitANN	Karate	USAir	C-elgans-frontal
Before training	AvgNew	0.089	0.063	0.0142
	MAE	0.841	0.372	0.225
	MAD	1.68	0.744	0.45
After training	MAE	0.011	0.005	0.0037
	CE	0.013	0.006	0.0034
	MSE	0.011	0.0041	0.0032
	AvgNew	0.042	0.008	0.0083
	AvgNn	0.312	0.092	0.082
	Time	1.12	85.6	5.95
	MAD	0.011	0.0031	0.003

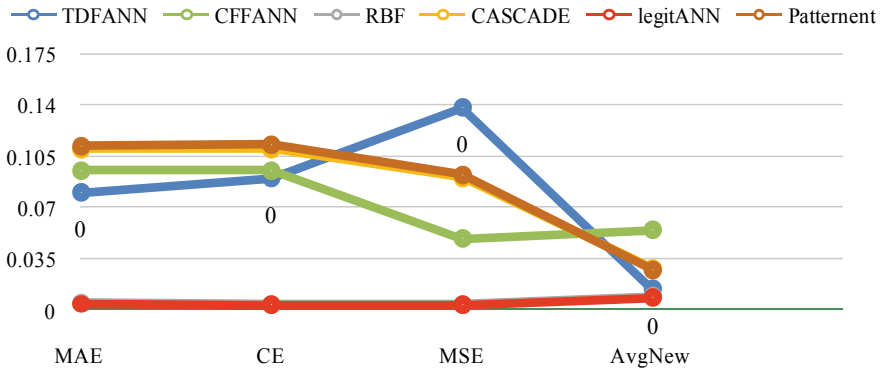


Fig. 3 C-elegans-frontal database legitANN and other ANN metrics value comparison

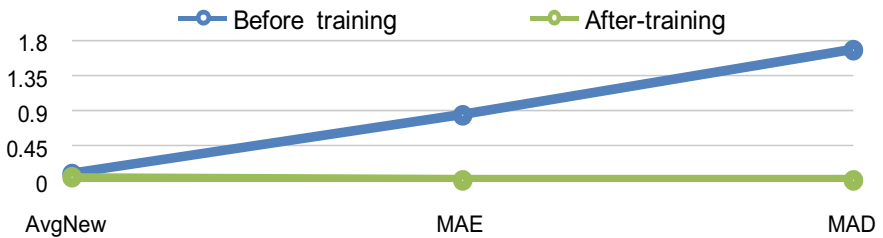


Fig. 4 Change is metrics value before and after trining by legitANN on karate dataset

datasets is the indication of good performance of bias controlling. MAD is the indicator of mean disperse of data in the data set. Less value of MAD after training is the indication that results are in similar range of values.

Figure 3 shows various metrics value for all ANN configurations used here in this paper. For MAE, CE MSE and AvgNew all these metrics legitANN is better than other ANN configurations. Even Fig. 4 shows a good comparison in before training and after training results on detests using metrics AvgNew, MAE and MAD. All metrics shows a good improvement.

As goal set for performance value is zero and nearly zero performance value is obtained. Improvements obtained are for karate dataset AvgNew got 52% better, MAE got 98.7% better and MAD value changed to 99.3%. For USAir dataset AvgNew got 87.3% better, MAE got 98.7% better and MAD value changed to almost 100%. Improvements obtained are for C-elegant-frontal dataset AvgNew got 41.3% better, MAE got 98.4% better and MAD value changed to 99.3%.

From the experiment it can be seen FFANN are better at approximating smooth functions since they tend to interpolate rather than average as can be seen from CFFANN results. They also seem better at approximating sparse data. They also seem to be more accurate for non-uniform point distribution; again mainly because cross-validation works better for uniformly dense sets as proved with TDFANN. But for

uniformly dense sets or preprocessed data, RBFs may be better since, theoretically, cross-validation should provide a more accurate response surface.

Here legitANN using RBF and three level bias control proves to be good choice as it perform better in terms of better MAE, AvgNew and also MAD metrics values.

6 Conclusion

Bias or unfairness may arise at any step in ML process. One of the leading causes of Bias is data, and another is the algorithm. Therefore creating a data collection process that is free from Bias is essential for the model's ultimate success. In contrast to previous work, our method eliminates the need to specify which biases are to be mitigated and allows simultaneous mitigation of multiple biases, including those related to group intersections. We are controlling generalized bias problem, not the specific ones.

A simple algorithm is implemented for diagnosing and mitigating Bias here. One advantage of this method is that there is no need to identify the source of Bias. Data is the most important, and so does the integrity of data. The fundamental challenge in neural modeling is more with data integrity than the learning. Using proper modeling principles, Bias can be greatly reduced or eliminated and those working on AI can help expose accepted biases.

References

1. Dey A (2016) ML algorithms: a review. *Int J Comput Sci Inform Technol* 7(3):1174–1179
2. Sandhya GU, Bisht U (2020) Assessment of effectiveness of data dependent activation method: MyAct. *J Intell Fuzzy Syst* 1–13
3. Sandhya GU, Bisht U (2019) Tailored feedforward artificial neural network based link prediction. *Int J Inform Technol* 1–9
4. Romanov A, De-Arteaga M, Wallach H, Chayes J, Borgs C, Chouldechova A, Geyik S, Kenthapadi K, Rumshisky A, Kalai AT (2019) What's in a name? Reducing bias in bios without access to protected attributes. *arXiv preprint arXiv:1904.05233* (2019).
5. Belinkov Y, Poliak A, Shieber SM, Van Durme B (2018) Mitigating Bias in natural language inference using adversarial learning
6. Zhang BH, Lemoine B, Mitchell M (2018) Mitigating unwanted biases with adversarial learning. In: *Proceedings of the 2018 AAAI/ACM conference on AI, Ethics, and Society*, pp 335–340
7. Smith P, Ricanek K (2020) Mitigating algorithmic bias: evolving an augmentation policy that is non-biasing. In: *Proceedings of the IEEE winter conference on applications of computer vision workshops*, pp 90–97
8. Zhang Y, Wu H, Liu H, Tong L, Wang MD (2019) Improve model generalization and robustness to dataset bias with bias-regularized learning and domain-guided augmentation. *arXiv preprint arXiv:1910.06745*
9. Díez VA, Egido MF, García Sánchez JJ, de Dios Zapata Cornejo E, Baquero Triguero JC, Fuente PG, Buenache AB, Samusev AM. Analysis and mitigation of bias in ML

10. Dietterich TG, Kong EB, Bias ML (1995) Statistical bias, and statistical variance of decision tree algorithms
11. Raschka S (2018) Model evaluation, model selection, and algorithm selection in ML. [arXiv:1811.12808v2](https://arxiv.org/abs/1811.12808v2) [cs.LG] 3 Dec 2018
12. Fomin VV (2020) The shift from traditional computing systems to artificial intelligence and the implications for bias, Jan 2020
13. Thomson R, Alhajjar E, Irwin J, Russell T (2018) Predicting bias in machine learned classifiers using clustering. In: Annual social computing, behavior prediction, and modeling-behavioral representation in modeling simulation conference
14. Bird S, Kenthapadi K, Kiciman E, Mitchell M (2019) Fairness-aware ML: practical challenges and lessons learned. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp 834–835
15. Dixon L, Li J, Sorensen J, Thain N, Vasserman L (2018) Measuring and mitigating unintended bias in text classification. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp 67–73
16. Niedoba M, Cui H, Luo K, Hegde D, Chou F-C, Djuric N (2019) Improving movement prediction of traffic actors using off-road loss and bias mitigation. In: Workshop on ‘ML for autonomous driving’ at conference on neural information processing systems
17. Acharyya R, Das S, Chatteraj A, Sengupta O, Iftekar Tanveer Md (2020) Detection and mitigation of bias in ted talk ratings. arXiv preprint [arXiv:2003.00683](https://arxiv.org/abs/2003.00683)
18. [https://en.wikipedia.org/wiki/Bias_\(statistics\)](https://en.wikipedia.org/wiki/Bias_(statistics)). Last Used on 06 June 2020
19. Sandhya GU, Bisht U (2020) Performance evaluation of various ANN architectures using proposed cost function. In: IEEE conference on ICRITO, 4–5 June 2020

Analysis of Implication of Artificial Intelligence on Youth



Deepika Pandoi 

Abstract Artificial intelligence (AI) promises to convey essential alterations to society, moving all from commercial to administration and professional life to private life. As AI structures are established and organized universally, the effects on society are uncertain. AI will generate significant alterations for society. The intention of this research is to investigate the prospective dangers AI creates for children and teenagers. In exploring these potential dangers, the types of AI are researched to develop a strong understanding of the technology, how AI is used today and the threats that are created through AI devices. AI devices and technologies have become common for the children of today's society, but the children and even the parents often do not focus on the privacy concerns that may linger after the use of Internet-connected devices. In order to protect children and even adults from the dangers of AI, people need to be educated on how to properly use the technology and even to be aware of when the use of AI is out of their control. The requirements of policy advancements and changes are also imperative to properly support a secure and private environment for children. Without government controls and policies, companies will misuse and capitalize on collected data. As the AI technology expands, user information will continue to be collected from adults, children, and companies on the Internet. It is imperative that the data harvested is properly utilized and stored to maintain proper privacy. As society continues to connect to the Internet with an expanding number of devices, the importance and impact of online safety is becoming increasingly important.

Keywords Artificial intelligence · Privacy laws · AI usage statistics · Technology · AI and youth

D. Pandoi (✉)

Institute of Business Management, GLA University, 17 KM Stone, NH#2, Mathura-Delhi Road, Mathura 281406, India

e-mail: deepika.pandoi@gla.ac.in

1 Introduction

Children today are growing up with technology all around them. Technology devices are used to communicate, learn, play, and have become tools that are always by their side. Specifically, artificial intelligence (AI) has become a major piece of children's lives and they are unfamiliar in a world without AI. AI is now built into many of the common devices used by children like phones, tablets, and even appliances around the house. Children see adults utilizing these devices everywhere and they expect to use them the same way. Children expect technology devices to function as planned and be reliable for offering fast and correct information, even at a young age [26].

In addition to the social implications surrounding a child's use of Internet-based devices and artificial intelligence, there are concerns associated with common vulnerabilities and exposures (CVE) of the stored physical data and the digital footprint created by children at a very young age. This footprint may also have an impact on their lives as adults. The consequences of stored Internet data on a child are still unknown and smart parenting skills are always required when a child is using any Internet-connected device. In addition to the oversight from parents, laws, and policies are beginning to catch up and are working to protect children on the Internet. Maintaining this safety for children can seem to be a difficult task, but policies and laws from higher levels are necessary and are required to remain relevant as technology changes [26].

1.1 Purpose of the Study

The intention of this research is to investigate the prospective dangers AI creates for children and teenagers. The author will present the research and explain what AI is and create a clear understanding of the technology. Next, the author will explore the research in the project to answer questions on both uses of AI. Finally, the author will also detail specifics on how children can be protected from the potential dangers of AI.

1.2 Research Questions

The subsequent research questions were evaluated in the investigation of AI and its function in children's security. These questions were chosen to extend an in detail perspective at AI while accepting the effects and limitations of society's adolescence.

RQ1. What is artificial intelligence?

RQ2. How can artificial intelligence be used against children?

RQ3. How can children be protected against AI?

2 Literature Review

The concept of AI has been a focus of conversation for many science fiction authors over the past century and the continued advancements in computing have made it a reality. AI technology is driving continued growth at an individual, business, and economic levels [13]. Children live in a digitally focused world and even the education industry encourages the use of technology by children to prepare them for the digital requirements that lie ahead [19]. The purpose of this literature review is to explore the specifics of AI and how children can be protected from the potential dangers that may be created from AI. Along with the use of any technology today, comes the participation of AI. AI has become the foundation of most technologies, from robotics to the Internet of Things and it is said that AI will change all sectors of human activity. At a very young age, children are interacting with AI through the normal means of tablets and laptops, but even many of the toys and games used by children may be connected via AI [11].

2.1 Defining Artificial Intelligence

AI can be difficult to define. Many people have their ideas of what AI is and where it may fit into their lives. In simplistic form and the overarching goal, AI is the ability of a man-made system to replicate human thoughts [8]. There are two layers of AI and the differences should be well understood while discussing the topic. The first type is referenced as strong AI, which is intelligence that can be capable of doing anything that humans can do, or even better. Strong AI is the end goal and what many strive for AI to be. Strong AI does not exist today, but there are hopes, to see it one day [8]. A machine that is as smart as a human will be built and possess intelligence that is both broad and adaptable [29].

The second type of AI is called weak AI. This AI will perform specific tasks and presents human-like experiences. Weak AI acts as a simulation of human processes and is in use today with technologies such as Siri and Alexa. The research performed in this research paper will mostly cover weak AI [18].

A deeper understanding of AI will be explored and the importance of gathered user data by companies will be detailed. User gathered data, including information collected from children, is valuable to help improve experiences offered to customers and allow companies to understand the demand in interests. The higher quality of data creates more opportunities and economic benefits [14].

The human experience cannot be replicated, is unique, and despite all that is known about the human brain, it remains an enigma. Artificial intelligence cannot compete with the human brain because even humans do not know enough about the human brain and humans are the creators of AI [6].

Uses of Artificial Intelligence. The uses of AI are boundless. AI is used to mimic human senses and is excellent at language translation, extracting the meaning of

written and spoken words, and searching Internet sites for specific shapes, colors, and details of photos. In the business world, AI has had major successes in the customer service process and creating operation automation with transferring email and call data, resolving billing issues, and personalizing advertising for customers [27]. In addition to playing a major role in customer service and business, AI has also been used in writing news stories and developing statistical articles like sport statistics and financial reports. This type of technology is common for many major news organizations, including the Wall Street Journal [27].

Today, AI is performing many of the simple tasks that people have learned to utilize and appreciate. AI is answering questions through smartphones or smart devices like Amazon's Alexa, completes text messages as they are being written, assists in providing the quickest route when traveling, and presents ads for products of particular interest to the user [12]. The functionality of AI is only going to become more sophisticated as self-driving cars are rapidly developing, home appliances are being controlled with location sensors, and factory robots are packing and delivering products to doorsteps. AI is not the next big thing. It is here and it is only going to greatly expand the global impact affecting people's lives [15] (Fig. 1).

AI Usage Statistics. AI statistics has initiated several functions in the industry and among customers. These techniques are usually being used for conversing with portable instruments [17] (Table 1).

2.2 *Changes in Privacy Laws*

Artificial intelligence and the protection of personal data are uniquely connected. On May 25, 2018, the European Parliament adopted the General Data Protection Regulation (GDPR) and AI became subject to the new GDPR AI had to adapt and overcome hurdles quickly to meet compliancy. These hurdles were primarily focused on big data and where the data or controllers processing the data resided [3].

This regulation is becoming a core driver in the push for user privacy; not just in the EU, but worldwide. The GDPR is becoming the primary stake of data protection for all digital users, even to a generation that only knows about Internet-connected devices and social lives driven by apps and touch screens [5].

The Personal Data Protection Bill, 2019, was introduced in Lok Sabha by the Minister of Electronics and Information Technology, Mr. Ravi Shankar Prasad, on December 11, 2019. The Bill pursues to deliver for the fortification of personal data of people and creates a Data Protection Authority for the same. The Bill administers the handling of personal data by: "(i) government, (ii) companies incorporated in India, and (iii) foreign companies dealing with personal data of individuals in India. Personal data is data that pertains to characteristics, traits, or attributes of identity, which can be used to identify an individual. The Bill categorizes certain personal data as sensitive personal data. This includes financial data, biometric data, caste, religious or political beliefs, or any other category of data specified by the government,

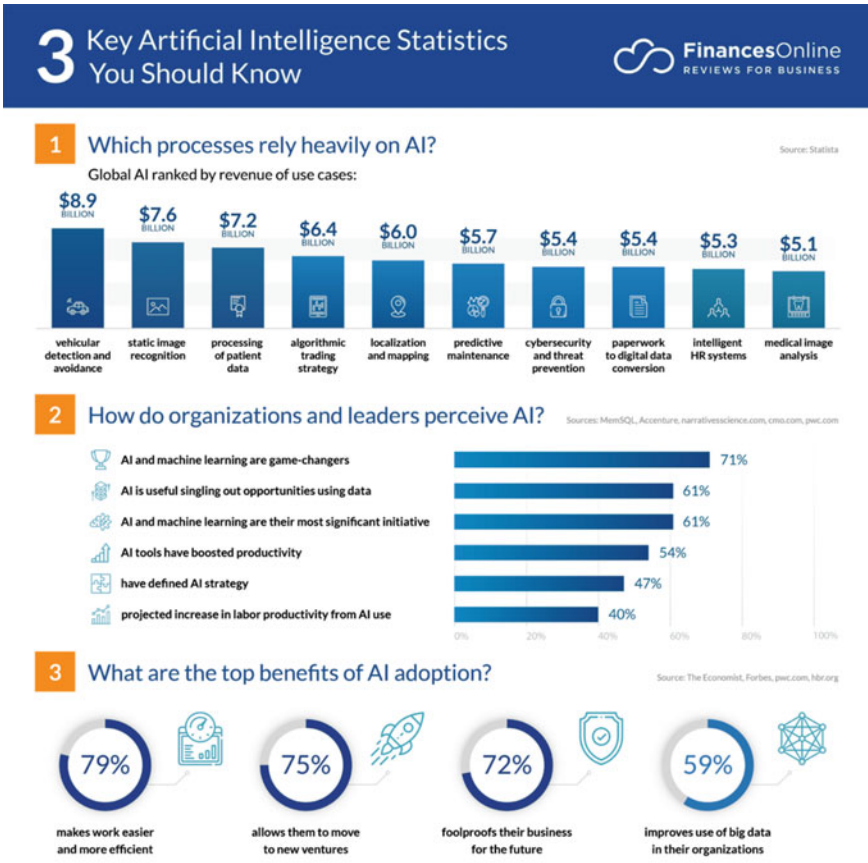


Fig. 1 Three key artificial intelligence statistics [15] Source: financesonline.com

in consultation with the Authority and the concerned sectoral regulator.” A data fiduciary is a person who selects the medium and drive of handling personal data. Such handling will be matter to definite objective, gathering, and storage restrictions. They need also to introduce tools for age confirmation and parental accord when handling complex personal data of children [16].

2.3 AI and Youth

The media is filled with stories told by adults about the future of AI. These stories are focused on self-driving cars, virtual assistants, and devices that support everyday business tasks [4]. The security and advancements put into AI are for both today and tomorrow’s generation. Responsible AI needs to be developed for the children

Table 1 “Cumulative global AI revenue forecast 2016–2025” [17]

Applications	AI revenue forecast in millions of dollars
“Static image recognition, classification, and tagging”	8097
“Algorithmic trading strategy performance improvement”	7540
“Efficient, scalable processing of patient data”	7366
“Predictive maintenance”	4680
“Object identification, detection, classification, tracking”	4201
“Text query of images”	3714
“Automated geophysical feature detection”	3655
“Content distribution on social media”	3564
“Object detection and classification, avoidance, navigation”	3169
“Prevention against cyber security threats”	2471

Source www.statista.com

of today and tomorrow [4]. Tablets, phones, and laptop devices are connected to the Internet and children may need proper supervision when using these devices, but the extended sprawl of Internet-connected devices has created security vulnerabilities in atypical places.

Children interact and use artificial intelligence every single day as they engage with electronic devices because almost all electronic devices used are considered smart or at least connect to an artificially intelligent source [1]. The concept of responsible technology is beginning to emerge and play a role in today’s world to help develop safe and ethical computing. Responsible technology is the practice to build practical, usable guidelines around ethical ways of creating and deploying anything digital [25].

With the continued progression and growth of AI, there are issues related to jobs, health, and security that need to be handled to both maintain the growth of responsible AI and to support children. Government intervention is needed to ensure that children have skills ready for the future. Intervention could mean changes in the school curriculum that alter both technical content and teaching methods that focus on interpersonal skills and emotional intelligence [4].

Many children today grow up surrounded by AI-powered voice assistants that sound or act human. The concern is that it is unknown how this interaction will

influence children's well-being? Research done by MIT shows early evidence that interactions between children and AI devices may alter children's perception of their own intelligence [9]. Additional research is needed to understand how digital devices may serve as responsible companions for children when embedded in toys, games, or devices.

When looking into children's privacy in technology advancements, both data privacy and data ethics are questioned. If properly developed and implemented, AI-powered toys and games could become a force for good that aids in empowering children by personalizing adaptive learning techniques [9]. In order for positive outcomes to arise with the fast development of technology, regulations on products and devices will need to be enforced.

What is being called the Internet of Toys continues to raise privacy and security concerns. Even with the constant link to the Internet, Mattel says it will not conduct any research with the details left behind. Similar toys have been taken off the market in various countries due to the hacking vulnerabilities left with either connection types or data storage options [9].

Information on people is collected and shared every day over the Internet and with connected devices. The screens being watched, the websites being viewed, the apps utilized all day, and even the toys and gadgets that have become necessary, are all collecting information and creating a pool of historical data on people. Data footprints continue to grow, and the bigger concern is that the data footprints of children now start at a very young age [21].

2.4 Education

Today's middle school children belong to the first generation to grow up as AI natives. These children interact with Amazon smart speakers and are familiar with YouTube algorithms. It is how things work for children, but it becomes important to teach children how to be responsible consumers and users of AI technology [22].

AI is an emerging technology that is speeding up change in the user environment. There is a responsibility to foster children's curiosity and creativity to prepare them for an unpredictable future. Children should not be pressured to utilize AI in a fixed way, and they need to be asked to challenge the digital systems used to foster leadership strengths [22]. It is significant to recognize that institutes are making youngsters to prosper in a domain where smart robots have changed the humans. It is on these skills that the government and corporations need to focus on. Humans in general need to be educated to go beyond what computers can do and take advantage of shifts in the job market [7].

2.5 *Dangers of AI and Youth*

The potential for transformative interactions with children or technology continues to expand dramatically. AI devices are used at home and in classrooms and children are using them without realization of any potential impact. Scoring is becoming a more important metric in all walks of life, particularly on social media. The idea of how many hits or how many likes someone incentivizes people to reach higher levels of ranking. Information on people is collected and shared every day over the Internet and with connected devices. The screens being watched, the websites being received is what pushes people to post comments, images, or videos. The system viewed, the apps utilized all day, and even the toys and gadgets that have become necessary, are all collecting information and creating a pool of historical data on people. Data footprints continue to grow, and the bigger concern is that the data footprints of children now start at a very young age.

3 Summary

The literature review provided factual insight into AI along with details of both concerns and benefits of AI, specifically with children. AI solutions continue to be customized to meet the requirements transversely in many sectors such as automotive, healthcare, education, finance, and entertainment. For example, in the automotive area, AI is used to power and control autonomous cars. These types of vehicles are expected to become standard shortly. In the healthcare industry, the growth of AI and machine learning has accelerated the pace of innovation and changed the entire operating models of hospitals. AI is also making major advancements in the education industry, where it is customizing learning programs for students [23].

The research also showed the connection and reliance between AI and big data. Big data and AI are currently two of the most powerful and useful technologies today. Artificial intelligence has been in existence for over a decade and even though the existence of big data is much less, the power of the two technologies has already shown great strength and opportunity. Computers can be used to store and analyze millions of records and data that will continue to provide deeper machine learning [2].

4 Discussion of the Findings

This research project analyzed substantial works concerning AI and the concerns of AI around children. AI makes it possible for machines to learn through experience and to attempt to function as a human, but there is a cost to doing this. The growth of AI is bringing numerous positive changes to society, but it seems the technology is

moving faster than the regulations that are required to be built around it. The larger technology companies like Google, Amazon, and Facebook are driving AI, but their focus is not on the welfare of children or even adults.

4.1 Artificial Intelligence

AI is a technology that is ever-changing and continues to re-shape society. Society is learning to rely on AI and AI is causing both technological and societal changes in the way people live. The rate of transformation between deep learning and big data is boosting the utilization of AI and creating new opportunities at a very rapid rate.

There is a long-term question with AI and whether or not the quest for strong AI succeeds. If there is a point where AI systems become stronger than humans at all cognitive tasks, then these AI technologies would be able to strengthen them to become a smarter AI system. This process of constant self-improvement could create a path that leaves the human intellect behind. In a positive opposition, the development of strong AI could revolutionize intelligence in such a way that it could aid in eradicating war, disease, or even poverty. The creation of strong AI could be the biggest event in human history but proves the need to align the goals of the AI with human, properly before the opportunity is too late [28].

4.2 Dangers of Artificial Intelligence with Youth

AI warrants concerns for children in various ways. Researchers are only beginning to learn how children will be affected and what the long-term results will be. AI can alter the way children communicate and even handle real relationships. With the omnipresence of AI devices always, a child may never feel alone or their desire to learn can be diminished, as their own intelligence is questioned. The growth of AI toys and devices is expanding rapidly and now is the time to focus on children's future [20]. There is a good chance that children will grow up thinking and feeling that toys or objects are just as good as actual pets or actual friends. A time where a child confuses a human with an AI managed toy or device is very real. As children get closer and more comfortable with AI devices, the level of information given to the AI data pool will increase and make children even more vulnerable to exploitation and manipulation by marketers, large companies, and even hackers.

Machine learning technologies are expected to impact everyone's daily lives increasingly and this generation of children will be most affected by AI technologies. Children today have been born and raised in an era of big data and machine systems. These systems will make decisions related to their education, employment opportunities, and much more [24].

4.3 Protecting Youth from Artificial Intelligence

The consequences of the gathered information on children are unknown, but in light of this uncertainty, humans should not be compliant with how organizations continue to collect and share children's data. Schools need to educate students about the significance of protecting private data and parents should be aware about what they have to share and what not to share. Also, the potential consequences need to be understood well by both parents and children.

The organisations that manufacture apps and software's for children require ending the option to track location and activity. Children do not understand privacy, nor will they think about their future when deciding to play with a toy or Internet device. They live in the now and like many parents, agreements are quickly skipped over, and the terms are just agreed to. The use of location tracking has also become very popular and dangerous at the same time. Now, even more details of children are stored and if the information ends up in the wrong hands, it could be dangerous. Also, the government needs to remain close to technology advancement to understand changes and effects on children fully. Legislation in accordance with data protection needs to be aligned with AI advancements in order to genuinely protect children.

4.4 AI Risks

There are plenty of risks associated with AI and much to be aware of as we progress deeper into AI. The highest short-term risks are around security and created bias. The opportunity to perform cyber-attacks on Internet-connected devices or harvested data is real and it needs to be protected. The systems and data stores can be seen as vulnerable targets depending on access processes and storage methods. The other immediate concern is the susceptibility of machine learning algorithms to bias. This bias may be based on race or gender, but it is an area that requires research and monitoring. If AI is going to be a powerful life tool in the future, this area needs to focus now as there is no room for bias in AI algorithms [10].

Other immediate concerns, such as privacy, will continue to be linked to AI as the gathered data is desired by many and can contain important personal data. This concern will inevitably always be there and requires constant focus. To aid in privacy concerns, laws and regulations are required to remain current with technology. Perhaps cite examples of data linked laws, such as PDP bill, 2019?

5 Conclusion

This research paper conducted and created a clear understanding of AI and details the dangers of AI around children, and the impact it may have on their future. AI is a

life-changing technology that is expanding rapidly. The development and progression of AI requires much attention and guidance around the development of algorithms used by AI. The potential dangers discussed in this paper can very easily overshadow the multitude of positive benefits AI creates. The true danger within AI comes from people not understanding it in its entirety. People must ask questions and understand where the data from a recording, Internet search, or a home smart assistant is going to be stored and understand that one day, it may be used again.

Many people fear AI and the change AI is bringing to society. Some say the fears are exaggerated, but the risks of AI are very real. The risks of AI require attention from users of AI, developers of AI algorithms, lawmakers, and even just anyone in public spaces. The reliance on AI alone could carry risks that are already here. Smart devices are a growing part of society and already aiding in decisions being made. These types of systems will continue to expand into areas like health care, criminal justice, and education. AI is already touching on educating children, but soon AI may be aiding in policing the streets or prescribing medications.

Both the growing number of Internet smart devices and the increasing quantities of stored data are creating severe privacy concerns for all AI users. The availability of devices collecting data is increasing and creating potential unsecured environments for children.

Today, as AI remains to be referenced as weak AI, the development and regulations installed will become the foundation of what one day may be called true strong AI. The data collected today, and the regulations enforced upon this collected data, will aid in creating a trusted and unbiased AI structure that can one day prove to be beneficial to humankind. Those working with AI today need to make it a priority to define the problems that specific AI will solve or the benefits it holds for society. At this time, the primary objective for most is not to have AI operate like a human brain, but to use its unique capabilities to enhance society.

References

1. 2 Ways artificial intelligence is affecting our families without us knowing. <https://factsandrends.net/2018/09/19/2-ways-artificial-intelligence-is-affecting-ourfamilies-without-us-knowing/>
2. Arora V (2018) How AI and big data are connected. AI Zone
3. Artificial Intelligence and the GDPR: how do they interact? <https://www.avocats-mathias.com/technologies-avancees/artificial-intelligence-gdpr>
4. Axente M (2018) Raising children in the era of artificial intelligence. Part One. PWC
5. Buttarelli G (2017) Teenagers on privacy. European Data Protection Supervisor
6. Cawley C (2018) Why artificial intelligence will never compete with the human brain. Tech. co.
7. Devlin H (2016) Schools not preparing children to succeed in an AI future. MPs warn. The Guardian
8. Dinh T (2018) Silicon minds: the science, impact, and promise of artificial intelligence. Wise Fox Publishing
9. Druga S, Williams R (2017) Kids, AI devices, and intelligent toys. MIT Media Laboratory
10. Ford M (2019) The biggest risks associated with the future of AI. Thrive Global

11. Goldstein B (2019) DUALITY: prepare yourselves and your children for the age of artificial intelligence
12. Greenwald T (2018) Artificial intelligence (a special report)—what exactly is artificial intelligence, anyway? Everybody’s talking about AI these days. Here’s what all the fuss is about. Wall Street J R.2
13. Hamke D (2019) In-depth: artificial intelligence 2019: Statista Digital Market Outlook. Statistica Content and Design, Hamburg
14. How to Monetize Your Data. <https://www.lotame.com/how-to-monetize-your-data/>
15. <https://financesonline.com/artificial-intelligence-statistics/>
16. <https://www.prsindia.org/billtrack/personal-data-protection-bill-2019>
17. <https://www.statista.com/chart/6810/the-future-of-ai/>
18. Kerns J (2017) What’s the difference between weak and strong AI. Machine Design
19. Leon Straker J (2018) Conflicting guidelines on young children’s screen time and use of digital technology create policy and practice dilemmas. J Pediatrics 4
20. Leung W (2018) How will AI technologies affect child development? The Globe and Mail
21. Longfield A (2018) Who knows what about me? Children’s Commissioner
22. Ma M (2019) How to teach kids about AI; a researcher at MIT Media Lab has designed a middle-school curriculum to help demystify algorithms and their effects. Wall Street J
23. Mehta D, Ann-Kristin A (2018) In-depth: artificial intelligence 2019
24. Memorandum on Artificial Intelligence and Child Rights. <https://www.unicef.org/innovation/media/10501/file/Memorandum%20on%20Artificial%20Intelligence%20and%20Child%20Rights.pdf>
25. Miller C (2017) What is responsible technology, anyway? Doteveryone
26. Saeed F (2017) 9 Powerful examples of artificial intelligence in use today. IQVIS
27. Sraders A (2019) What is artificial intelligence? Examples and news in 2019. The Street
28. Tegmark M (2019) Benefits & risk of artificial intelligence. Future of Life Institute
29. Vincent J (2018) This is when AI’s top researchers think artificial general intelligence will be achieved. The Verge

The Categorization of Artificial Intelligence (AI) Based on the Autonomous Vehicles and Its Other Applications



S. V. Aswin Kumer, P. Kanakaraja, L. S. P. Sairam Nadipalli,
N. V. K. Ramesh, and Sarat K. Kotamraju

Abstract The artificial intelligence (AI) is creating the great impacts and revolution in the technological world, education world, industry world, and business world. The concepts in artificial intelligence are improving themselves day by day, and it is becoming a part of human's everyday activities. It plays a major role in the human life, and the situation becomes the people cannot live without this AI. In this book chapter, the categorization of artificial intelligence based on the autonomous vehicles and its other applications is to be discussed in brief. The artificial intelligence technique is based on the artificial neural network (ANN), machine learning (ML), deep neural network (DNN), recurrent neural network (RNN), and convolution neural network (CNN). To implement these types of neural network, the system should use the programming languages like MATLAB programming, Python programming, R programming, etc.

Keywords Artificial intelligence (AI) · Internet of things (IoT) · Artificial neural network (ANN) · Machine learning (ML) · Deep neural network (DNN) · Recurrent neural network (RNN) · Convolution neural network (CNN) · Autonomous vehicles

S. V. Aswin Kumer (✉) · P. Kanakaraja · L. S. P. Sairam Nadipalli · N. V. K. Ramesh · S. K. Kotamraju
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh - 522502, India
e-mail: svaswin@gmail.com

P. Kanakaraja
e-mail: pamarthikanakaraja407@gmail.com

L. S. P. Sairam Nadipalli
e-mail: sai.nadipalli@gmail.com

N. V. K. Ramesh
e-mail: nv.krishnaramesh@gmail.com

S. K. Kotamraju
e-mail: kksarat@gmail.com

1 Introduction

At first, the learner should know the difference between the artificial intelligence (AI) and artificial neural network (ANN). The artificial neural network [1] can be formed by the input excitation layer, hidden layer, and the output response layer as shown in Fig. 1. For example, the artificial intelligence is implemented using deep neural network (DNN), the DNN also has excitation layer and response layer. But there are multiple hidden layers present in the DNN as shown in Fig. 2.

This is the major difference between an artificial neural network and deep neural network. Both of the neural networks using the artificial neurons in the nodes available in the hidden layer. When comparing to the artificial neural network [2], the deep neural network will produce the results with more accuracy and more efficiency. So, the deep neural network is preferred to implement the artificial intelligence. The training is necessary for any neural network, and based on the training data, the neural network will generate the results. The accuracy and the efficiency of neural network are based on the training data.

2 Artificial Intelligence

The artificial intelligence can be majorly classified into two parts. They are artificial embedded intelligence (AEI) and artificial integrated intelligence (AII) as shown in Fig. 3. The artificial embedded intelligence will perform only specific tasks based on the training data. For example, face recognition, iris recognition, fingerprint scanning and processing, image quality enhancement, object detection, predicting the weather forecast, automatic maintenance of room temperature, remote sensing, satellite imaging, speech recognition, text to speech engine, speech to text engine, smart translator, searching the concepts in the websites, smart assistant, self-driving cars, smart farmer in agriculture, smart worker in industries, smart speaker, smart physician for diagnosing the diseases, etc., The artificial integrated intelligence will perform

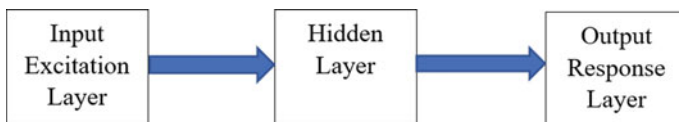


Fig. 1 Stages of artificial neural network (ANN)

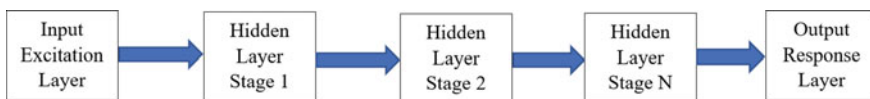
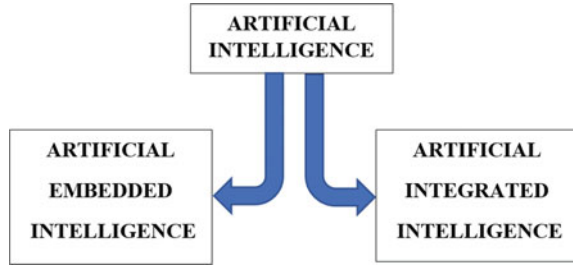


Fig. 2 Stages of deep neural network (DNN)

Fig. 3 Classification of artificial intelligence



multiple tasks, which is the major competitor for the human being. Because the artificial integrated intelligence will perform the tasks as the human being can do, also, it will do faster and more accurate than the human being. At this point, the artificial intelligence overtakes the natural human intelligence, and most of the times, the artificial intelligence is a step ahead than the natural human intelligence. The humanoid robots are the suitable example for the artificial integrated intelligence.

3 Machine Learning

Before implementing the artificial intelligence, the learner must know about the concept of machine learning [1]. It is the type of artificial intelligence which learns the mappings from input excitation to output response. This type of learning mechanism is called supervised learning. For example, if a person is receiving an email, then supervised learning type of AI system checks the category of that email, whether the email belongs to primary inbox or promotion inbox or spam message. Based on the previous actions and steps done by the user, it categorizes the received email as shown in Figs. 4 and 5.

Another example is, if a person received an audio file, then the supervised learning type of AI system converts that audio into script form by using speech to text engine. The audio may consist of tracks and voices, but the system will find the type of music and the variety of language. If the user wants to translate the text and speech of that particular audio into other languages, it is also possible with the supervised learning type of AI system. The major application of the supervised learning type of AI system in the current situation is, posting the advertisement in the sidewall of the websites. Based on the previous search history of the user, the supervised learning type of AI posting the advertisements in the websites which is visited by the user. For example, if the user searches about the mobile phone and its specifications, then the user will

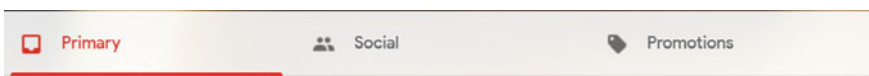
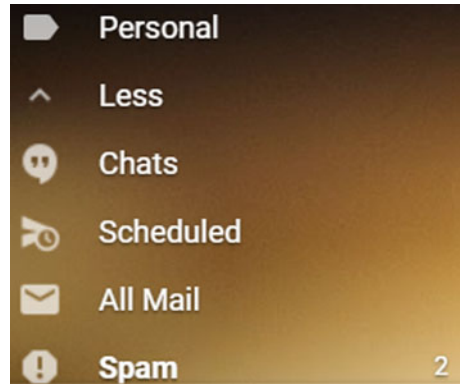


Fig. 4 Categorization of emails into three different parts

Fig. 5 Categorization of emails into personal and spam



get the advertisements which are related to the mobile phones in the sidewalls of the other websites. And also, the user receives the offers and discounts to purchase a mobile phone in his inbox. Another application of supervised learning type of AI system is self-driving cars. At first, the system finds the position and location of the vehicle. After that, it checks whether all the vehicle parts like power steering, gearbox, clutch, accelerator, speedometer, fuel indicator, antilock braking system, air bags, water viper, headlamps, fog lamps, turning indicators, stop indicators, wheel alignments, power windows, and central door locking system are working properly or not. And also, it checks whether the vehicle parts can be operated, controlled, and monitored by the AI system or not. It also verifies the working of GPS module, sensors, and actuators connected in the cars to operate, control, and monitor the vehicle parts. The ultrasonic sensor is used to calculate the distance of the nearby vehicles and objects. Likewise, many sensors are used for different purposes in the car which is shown in Fig. 6.

The AI can also be applied in the assembling unit in the industries. Based on the picture available in the industries for assembling the products, it will assemble the products or the AI will provide the assembling instructions through voice commands for the low-tech employees for perfect assembling of products. In all the above cases, there is an input to output mappings happened in the system which is so-called supervised learning.

The efficiency of supervised learning is varying for different applications, and it is very valuable for some applications. To increase the efficiency of the supervised learning, the huge amount of input data should be given to the hidden layers. After that, the neural network should be trained well based on the available huge data. The performance and accuracy of the system depends on the training of neural networks with large amount of data as shown in Fig. 7. Nowadays, most of the industries are digitizing the data available with them which was collected over the decades. The newly generated data is already in the digitized form, so, it is easy to collect and it is stored in the database.

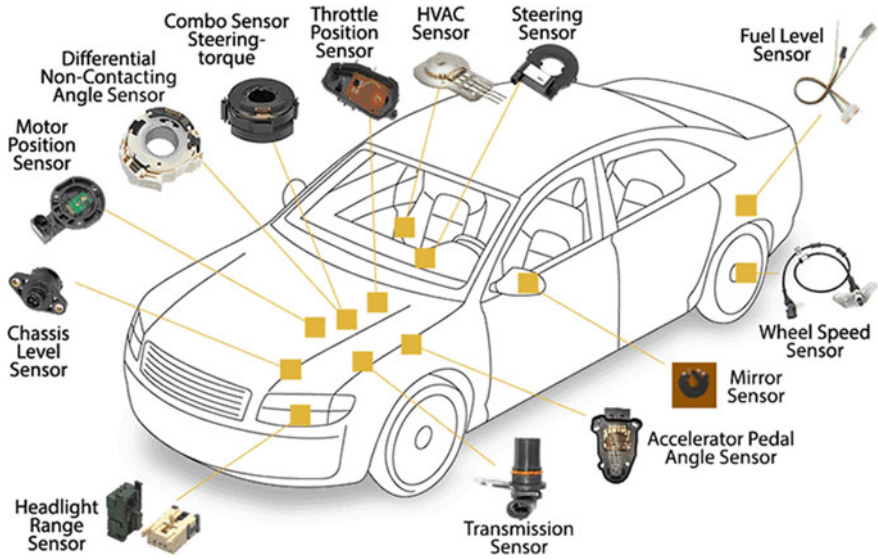


Fig. 6 Car schematic diagram with different sensors

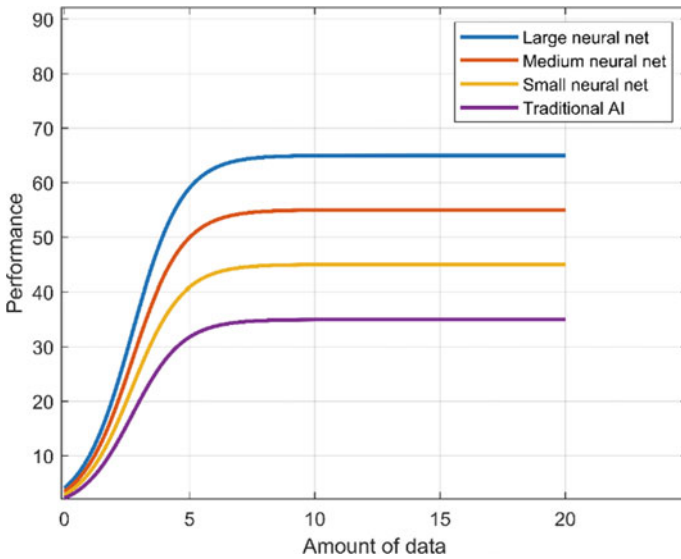


Fig. 7 Performance of different sized neural network based on amount of data

4 Data Input

To build an AI system, the data is necessary to give it as input. Depending upon the input data, the neural networks process the task and produce the output data. The training data is also necessary to produce the output data. So, the learners are having the confusion about data, input data, training data, and output data. This segment deals with the concept to clarify the above questions. For example, if anybody wants to purchase a laptop table, then the dataset has been created in Table 1.

The data available in the dataset can be separated by input IN and output OUT. The dataset can be created by using spreadsheet, and the AI system learns about IN to OUT mappings. The IN and OUT should be decided based on the customer. The customer can choose only the material as an input and length, breadth, height, and price as an output as shown in Fig. 8 or the customer can choose the material, length, breadth, and height as an input and only price as an output as shown in Fig. 9.

Based on the customer choice, the AI system will perform the IN to OUT mappings. But, sometimes, it will happen in the reverse that, the input of the customer is price of the table and the output required for the customer is material, length, breadth, and height. For that case, the AI system will perform reverse mappings as shown in Fig. 10. The dataset can be developed further like thickness of the table,

Table 1 Example of a dataset

Material	Length (in.)	Breadth (in.)	Height (in.)	Price (USD)
Wood	20	40	25	700
Plywood	20	40	25	800
Plastic	20	40	25	900
Fiber	20	40	25	1000
Iron	20	40	25	1100
Steel	20	40	25	1200

Note The above table is the assumption and mentioned as an example for better understanding

S.NO.	MATERIAL	LENGTH	BREATH	HEIGHT	PRICE
1	WOOD	20 INCHES	40 INCHES	25 INCHES	700 USD
2	PLYWOOD	20 INCHES	40 INCHES	25 INCHES	800 USD
3	PLASTIC	20 INCHES	40 INCHES	25 INCHES	900 USD
4	FIBER	20 INCHES	40 INCHES	25 INCHES	1000 USD
5	IRON	20 INCHES	40 INCHES	25 INCHES	1100 USD
6	STEEL	20 INCHES	40 INCHES	25 INCHES	1200 USD

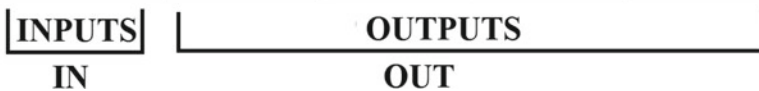


Fig. 8 IN and OUT mappings of AI

S.NO.	MATERIAL	LENGTH	BREATH	HEIGHT	PRICE
1	WOOD	20 INCHES	40 INCHES	25 INCHES	700 USD
2	PLYWOOD	20 INCHES	40 INCHES	25 INCHES	800 USD
3	PLASTIC	20 INCHES	40 INCHES	25 INCHES	900 USD
4	FIBER	20 INCHES	40 INCHES	25 INCHES	1000 USD
5	IRON	20 INCHES	40 INCHES	25 INCHES	1100 USD
6	STEEL	20 INCHES	40 INCHES	25 INCHES	1200 USD

IN

OUT

Fig. 9 IN and OUT mappings of AI

S.NO.	MATERIAL	LENGTH	BREATH	HEIGHT	PRICE
1	WOOD	20 INCHES	40 INCHES	25 INCHES	700 USD
2	PLYWOOD	20 INCHES	40 INCHES	25 INCHES	800 USD
3	PLASTIC	20 INCHES	40 INCHES	25 INCHES	900 USD
4	FIBER	20 INCHES	40 INCHES	25 INCHES	1000 USD
5	IRON	20 INCHES	40 INCHES	25 INCHES	1100 USD
6	STEEL	20 INCHES	40 INCHES	25 INCHES	1200 USD

OUT

IN

Fig. 10 IN and OUT mappings of AI





weight tolerated by the table, reliability of the table, power wire through hole option, and ventilation for the laptop exhaust. This will improve the efficiency of the AI system to meet the customer requirements. Also, the database has the data in the ranges with respect to price, properties, and features as listed in following Table 2. These are the input dataset given by the customer to purchase the table in the online shopping portal. Based on the inputs from the customer, the AI system provides the information related to that with the help of supervised learning.

The AI system performs the input to output mapping to satisfy the customer requirements. The customer should choose the options available in the shopping portal to specify the inputs as requirements. The customer has multiple choices to select to expose their requirements. Different customers have various needs and

Table 2 Example of an input dataset

Material	Length	Breadth	Height	Price
Wood	Below 15 in.	Below 15 in.	Below 15 in.	Below 700 USD
Plywood	15–20 in.	15–20 in.	15–20 in. in.	700–800 USD
Plastic	20–25 in.	20–25 in.	20–25 in.	800–900 USD
Fiber	25–30 in.	25–30 in.	25–30 in.	900–1000 USD
Iron	30–35 in.	30–35 in.	30–35 in.	1000–1100 USD
Steel	35–40 in.	35–40 in.	35–40 in.	1100–1200 USD

Table 3 Input data for AI system to find the name of the animal

Image	Description	Image	Description
	Dog		Dog
	Cat		Cat

choices selected by them are unique. If many customers have selected the same requirements, then that will be highlighted for the other customers who are not choosing any options. For example, the customer choice of input is highlighted in Table 2.

The customer needs plastic table with the length of 25–30 in., breadth of 35–40 in., height of 30–35 in., and the price should be in the range of 700–800 USD. Based on this input, the AI system maps the output data with the help of training data using supervised learning. To get clarity about training data, another example is creating the AI system to find the animals in the available images. Table 3 shows the input data for AI system and the result of AI system about finding pets.

The AI system clearly examines the dogs and cats in the input image. It is possible only by training data. Before giving input to the AI system, the AI system should undergo the training with lots of images. Based on the training data, it finds the animal available in the input images. The training data should be huge to label the input images. It needs thousands of training images to label the image with more accuracy. The AI system observes different images in the search engine as the training data. If the AI system learns from the data, then that data is called training data. And also, the AI system observes the user behavior in the websites and shopping portals and these types of observation are also called as machine learning. If the AI system produces the response based on the comparison between the input data and training data, then it is called output data. Finally, the above discussion is enough to know about data, input data, training data, and output data.

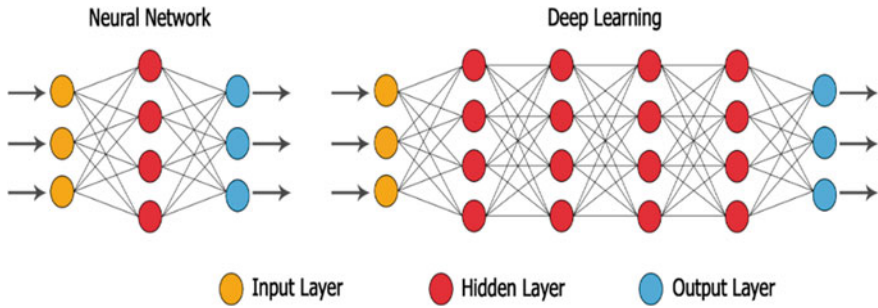


Fig. 11 Structure of artificial neural network and deep neural network

5 Deep Neural Network

The deep neural network has multiple hidden layers which are used to improve the accuracy and efficiency of the task [3]. But there are many computations have to perform because of the multiple layers. There are huge number of data to be processed in this neural network which occupies more memory for processing and storing the results [4]. When compared to the conventional neural network which is also called as artificial neural network, the deep neural network has more hidden layers [5]. The artificial neural network has only one hidden layer. Figure 11 shows the basic structure of the artificial neural network and deep neural network. The similarity between these two neural networks is they have only one input layer and only one output layer. This deep learning neural network is mainly used in computer vision, image understanding, machine vision, image processing, audio processing, and speech processing.

6 Recurrent Neural Network

The recurrent neural network is also called feedforward neural network. If the artificial neural network has the feedback neural network, then it is called recurrent neural network which is shown in Fig. 12. This is the neural network having dynamic behavior which includes gated memory to store the previous state [6]. The controlled states are the part of the gated recurrent units (GRU) and the long short-term memory (LSTM). The directed graph can be formed by using the connections of the nodes which are available in the neural network [7]. The major applications of this recurrent neural network are speech recognition and handwritten-based text recognition which needs dynamic behavior of the system.

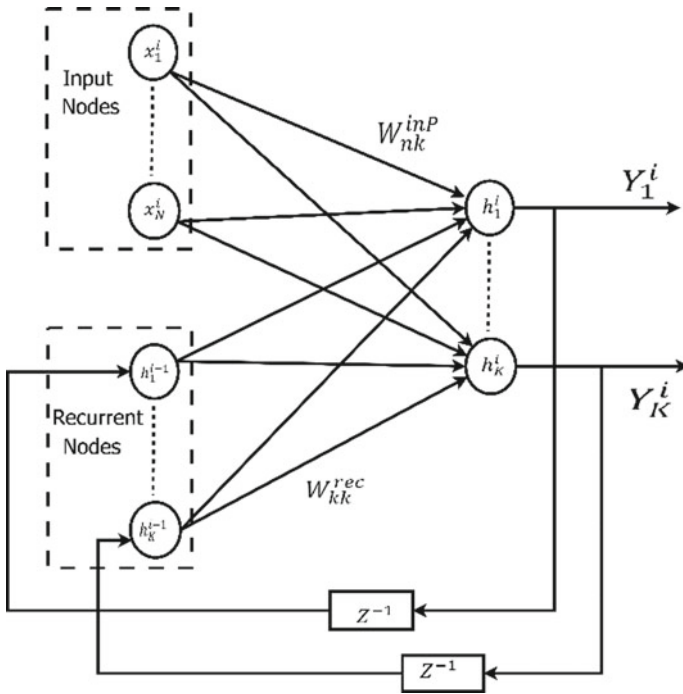


Fig. 12 Structure of recurrent neural network

7 Convolution Neural Network

The shift-invariant neural network is the other name for this convolutional neural network because of its translation invariance characteristics [8]. This convolution neural network is also having multiple hidden layers which are fully connected layers [9]. The neurons in each stage of the hidden layer are connected with the adjacent layer for continuous and sequence processing. So, it is also called as fully connected layer neural network. It has the shared weights architecture in the layers, so, it is called space invariant artificial neural network. The subsamples are the pool layer and the convolutions are performed by the convolution layer [10]. The structure of convolution neural network is shown in Fig. 13. The major applications of this neural network are the image recognition and image classification in medical image analysis.

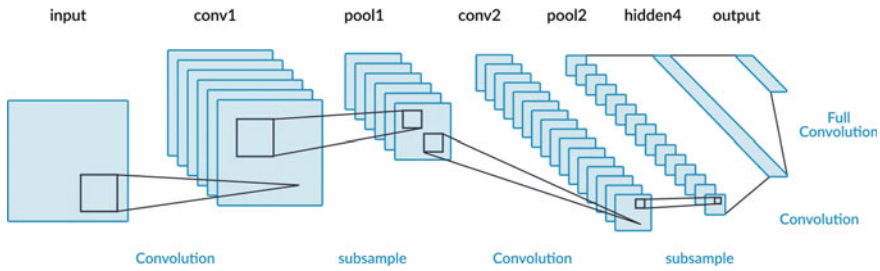


Fig. 13 Structure of convolution neural network

8 Conclusion

Thus, the classification of the artificial intelligence based on the applications is deeply discussed in this chapter which helps to select the particular type of neural network suitable for the specific applications. The major applications are image processing, embedded systems, and Internet of Things (IoT). The contribution of the artificial intelligence in Internet of Things (IoT) and image processing is massive and many researchers are solving their research problems using the different categories of the artificial intelligence and machine learning.

References

1. Maddiseti L, Senapati RK, Ravindra JVR (2019) Training neural network as approximate 4:2 compressor applying machine learning algorithms for accuracy comparison. *Int J Adv Trends Comput Sci Eng* 8(2):211–215
2. Sekhara Babu V, Venkat Ram N (2019) Rotation invariant object classification using convolutional neural network. *J Cri Rev* 6(6):479–482
3. Srinivasa RS, Suman M (2018) Microaneurysm extraction with contrast enhancement using deep neural network. *J Adv Res Dyn Control Systems* 10(11):313–320
4. Anantha RG, Kishore PVV, Sastry ASCS, Anil KD, Kiran KE (2018) Selfie continuous sign language recognition with neural network classifier. *Lecture Notes Electr Eng* 434:31–40
5. Reddy SS, Suman M, Prakash KN (2018) Micro aneurysms detection using artificial neural networks. *Int J Eng Technol (UAE)*, 7(4):3026–3029
6. Lakshmi Prasanna P, Rajeswara Rao D (2019) Probabilistic recurrent neural network for topic modeling. *Int J Innov Technology and Exploring Engineering*, 8(4), PP.165–168.
7. Prasanna PL, Rao DR (2019) Development of topic modeling framework using probabilistic recurrent neural network. *Int J Adv Trends Comput Sci Eng* 8(4):1761–1767
8. Prasad MVD, Lakshamma BJ, Chandana AH, Komali K, Manoja MVN, Kumar PR, Prasad CR, Inthiyaz S, Kiran PS (2018) An efficient classification of flower images with convolutional neural networks. *Int J Eng Technol (UAE)* 7(1.1):384–391
9. Kishore PVV, Anantha RG, Kiran KE, Teja Kiran Kumar M, Anil KD (2018) Selfie sign language recognition with convolutional neural networks. *Int J Intell Syst and Appl* 10(10):63–71
10. Raja C, Balaji L (2019) An automatic detection of blood vessel in retinal images using convolution neural network for diabetic retinopathy detection. *Pattern Recogn Image Anal* 29(3):533–545.

Context Awareness Computing System—A Survey



Pranali G. Chavhan, Gajanan H. Chavhan, Pathan Mohd Shafi,
and Parikshit N. Mahalle

Abstract Context term had been sensing in various areas like pervasive computing, intelligent environment, ubiquitous computing, etc. In such areas to sense data and then adapting the moments or behavior according to environment or context information like physical context, user task, and computational context which captured the data, an implementation of such a context-aware computing system is more difficult in many ways. The prime focus of such implementation is always to provide realistic and very specific data required for understanding the actual scene being monitored. Internet of Things (IOT) has changed the overall process of gathering the data as well as processing of that data. In smaller systems like an embedded systems data collected by the sensors and processed on a computing platform to understand the process parameters or to sense actual changes occurred during the given process execution, an attempt has been made by the researcher in this paper to review various fundamentals of the context-aware computing by reviewing the available literature, and a system-based approach is presented in comparative manner so as to justify the uniqueness of this in pervasive computing. So in this paper, the concept of context, context awareness, context awareness algorithms, and some techniques are discussed.

Keywords Pervasive computing · Context · Context awareness · Context modeling and techniques and Context ontology model

P. G. Chavhan (✉)

Department of Computer Engineering, VIIT, Pune, Maharashtra, India
e-mail: pgchavhan@gmail.com

G. H. Chavhan

Department of Electronics & Telecommunication Engineering, VIIT, Pune, Maharashtra, India

P. M. Shafi · P. N. Mahalle

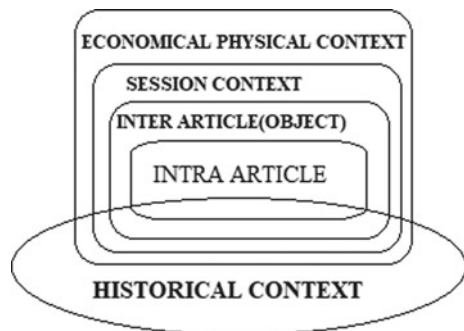
Department of Computer Engineering, SKNCOE, Wadgaon, Pune, Maharashtra, India

1 Introduction

The context-aware computing is very specific process of pervasive computing environment which deals with the conscious understanding of the data being sensed. This refers to an adaptable but general class of pervasive computing environment. Context is set of conditions over a specific situation confined either in logical or in physical domain. The most significant parts of defined context are like the place you are (where you are), who you are with, and what assets are close by (what resources are nearby). Apart from the physiological significance of the context, it is presumed that the system should respond according to the user's behavior. Sometime the user may respond to a situation in a mechanized way, but still there could be many more differences at perception level of sensing in a system. The only physical parameter sensing is not sufficient to learn and feel the environment and actual the underline message is need to be decode in many case. Likewise knowing a person's digital personality profoundly can lead to understand the actual user needs. To have this in place, the systems have to sense the location, object, time, some activity, and other sources. Many scholars have documented the fundamental characteristics of context according to their extracted meaning of context. It is very evident from these literatures that the overall concept of the context awareness is somehow more inclusive term and concept. Jinhong Wu, Wehun University [1] used the term context in 2010 and explores the concept of context as personal knowledge, social knowledge, and context ontology. This is an attempt to provide a cocoon structure of associated parameters which includes sensed information, physical behavioral, and conceptual parameters. The context ontology includes economical–physical context, social systematic, conceptual, emotional context, session context, and inter-object and intra-object layers (Fig. 1).

The intra-object context is deepest in the settled structure of a system like a general counting data about items themselves, for example, word, sentence, passage, etc. The object context depicts connections among objects. In this layer, each object was viewed as other items' specific situation. For instance, graphs can be viewed as the context of content and section as the unique circumstance of single sentence.

Fig. 1 Basic element of context



The context can be learnt more precisely by considering a scenario and analyzing it under the scanner of context oriented pervasive computing. For example, we may consider an educational case where a student's digital identities and behavioral parameters can be collaborated to extract the higher level of student-specific information. The session context contains specific and individual data about the student including preferences, style, understanding, information, etc., and exists during connection among student and framework. This information can be then utilized by learning framework to translate the student's subjective practices. The accelerative contexts are the outer context in the settled structure, including social foundational context and monetary techno-physical context. They depict impact in all clients, units, and connections affected by outer condition. A users contextual data points are static elements till the system does not find the correlation between these separate data points. History of a user is such a data point container. The history context is an extra and time measurement and over all levels and can be shaped by the client desires for past understanding. It has acquaintances with contexts all above.

2 Literature Survey

Defining the context awareness is a common issue tried to resolve by many researchers through their understanding of the process and the anticipated outcome of these processes. Context awareness is a higher dimensional data set if we compare it with other static data-sensing methods. The underlined expectation for context-aware systems and pervasive computing is to have more precise understanding of the data that a system is processing. A brief summary of all such definitions given by researchers is given below.

Context applicable to the fields and raw information collected at various stages of the processes is always client centric. In some cases, these raw information is also treated as per the conditional requirements or action requirements [2]. This context awareness was introduced by Xiufen Fu, Zhande v, and Weiqiang Huang. Framework familiarization works begin with physical sensors recognition. The collected data by such physical sensors are to be uniquely identifiable. The sensors are working at physical as well as logical level and provide an interface between the context and the source.

Physical sensor is used to detect and to obtain the original data. System parameters sensing and awareness of status items like battery poser, network conditions, etc., along with the history of the users sensed by the logical sensors [2, 3]. Ferit Topcu has introduced SPATIAL MODELS FOR CONTEXT, which are fact-based models. The context of the model is organized by the physical locations of the sensors [4]. Selviandro et al. proposed case -based reasoning technique [5]. In this techniques, the author proposed four stage for CBR. These four stages are retrieve, reuse, or revise and retain [6]. The data is fully collected in the very first stage of the data sensing. Supervised and unsupervised rule based systems are used to develop the fuzzy logic-based algorithms.

3 Context-Aware Techniques and Algorithms

The context modeling schemes are compared with different algorithms, and a brief summary is presented for discussion and review. The algorithms selected for comparison is the main criterion for classification and has better level of separation index as expressed by many researchers. A detailed study of the present literature shows that this area is not yet explored (Table 1).

The context-aware applications and administrations can be found in different diverse application spaces as beneath which are increasing serious considerations around the world. These are sample applications and are used in many such cases in real-life scenarios.

1. Travel agencies are mostly using location information's and routing paths as per the likes and selections of the customers.
2. Smart living spaces where understand the environment status
3. Smart home scenarios: Many smart homes are built around and the systems are gathering huge amount of data under this. The services so offered are tailor-made solutions and expected to have learning approach to be maintained so as to provide maximum personalized services to the user. Critical parameters to be considered in this case are response time and accuracy [5].
4. Service selections are always difficult problems to be solved in personalized service providers. A user profile plays a major role in such areas, and many more decisions are made on the basis of these already provided data inputs by the user.
5. Multi-agent systems are one of such condition where individual performance matrixes are always at higher priorities.
6. Hotel industry is having many differentiated services to be offered and high level and complex bunch of orders to be processed.
7. Human presence detection by using PIR sensors is a common practice nowadays to be used in public places. This also collects all other detection records, but no such data can deliver output since those outputs are masked [9].
8. User activity-based services are offered to clients like providing support to execute routine activities like preparing a coffee at a particular time [7].
9. Based on the dryness of the soil, sprinklers can be operated.
10. In cleaning-based situation, cleaning machines can clean the floor surfaces wherever human presence is not detected, and these machines can stop if movement is detected in that area to safeguard them [10].
11. Temperature-based situation, where the room air conditioning system will be operate on the basis of user profiles and sensor temperature data [7].

Table 1 Context-aware uses or scenario

S. No.	Algorithms	Advantages	Disadvantages	Explanations
1	Fuzzy logic	Provide more accuracy	Less response time	FLBA uses two important parameters like response time and accuracy
2	Supervised context-aware algorithms	Accurate and reliable results	Only uses off-line analysis	All data is labeled which is learning to predict the output from the input data
3	Unsupervised context-aware algorithms	It generates higher level user contexts from low-level information in a simple manner and also uses real time analysis of data	Adequate reliable and accurate results	All data is unlabeled which learn to inherent structure from the input data
4	Ontology-based context-aware algorithms	It assembles a typical ontology-based setting model with Web Ontology Language (OWL), and this gives a viable and helpful method of philosophy. This includes working area explicit ideas into this basic model. The basic philosophy-based model is sensible and viable [7]	Huge DB required for computational and time consumption and difficult to interface with outer face model with OWL	Ontology-based propose a typical setting model with OWL and present the arrangements with the regular model to tackle the issues of position and setting history. A BCIA model worked with the basic model is given, in which some thinking rules portrayed by SWRL is characterized to catch the changing action of a specific traveler
5	Rule-based context-aware algorithms	Rule-based model ought to be expressive, vigorous, and expandable, and context-aware frameworks is the model for portraying and misusing setting [8]	Problematic to define rule as per various list requirement	A standard-based technique based on the concerning the event condition movement model. End customers show when and what should be educated to them by using an ontology RBO rule definition

(continued)

Table 1 (continued)

S. No.	Algorithms	Advantages	Disadvantages	Explanations
6	Fuzzy logic context-aware algorithms	This method is very easy and easily can extend and it needs less resources and use natural language	Some mistakes are done by developer while entering manual values	In fuzzy logic, the author utilizes fuzzy deduction framework as a system. The contribution from sensors and afterward changed over in to fluffy sets with the assistance of participation capacities
7	Predicate detection-based context awareness (PD-CA) framework	Fault-tolerant mechanisms	Need more realistic and comprehensive experimental evaluations are required	Using PD_CA, it present the programming toolkit to simplify the development of context-aware applications based on MIPA [8]

4 Conclusion

In this paper, here, we describe the fundamentals of context and context awareness. We analyzed different technique and algorithms along with popular context-aware applications. Context-aware algorithms or techniques are basically focused on interaction among people and machines (IoT and IoE). Context awareness makes a deliberate examination on client inclination and proposed a technique which can use to compute network for different properties.

References

1. Malik S, Jain S (2017) Ontology based context aware model. ISBN 978-1-5090-5595-1/17/\$
2. Subbu KP, Vasilakos AV (2017) Big data for context aware computing—perspectives and challenges. Big Data Res. <https://doi.org/10.1016/j.bdr.2017.10.002>
3. Liu X, Mo X, Wang C, Wang H (2009) A rule-based ontology for context-aware computing
4. Topcu F (2011) Context modeling and reasoning techniques. Department of Telecommunication Systems, Technical University of Berlin, Berlin
5. Selviandro N, Sabariah MK, Saputra S (2016) Context awareness system on ubiquitous learning with case based reasoning and nearest neighbor algorithm. ISBN: 978-1-4673-9879-4. <https://doi.org/10.1109/ICoICT.2016.7571882>

6. Xin M, Cao H, Niu Z (2014) An approach to a location context awareness service prediction algorithm. In: 2014 IEEE workshop on electronics, computer and applications. ISBN: 978-1-4799-4565-8
7. Patel A, Champaneria TA (2016) Fuzzy logic based algorithm for context awareness in IoT for smart home environment. ISBN: 978-1-5090-2597-8
8. Yiling Yang, Yu Huang, Xiaoxing Ma, and Jian Lu, 'Enabling Context-awareness by Predicate Detection in Asynchronous Environments', IEEE TRANSACTIONS ON COMPUTERS, VOL. X, DOI <https://doi.org/10.1109/TC.2015.2424879>, IEEE Transactions on Computers
9. Huang S, Yin B, Liu M (2017) Research on individualized learner model based on context-awareness. In: 2017 international symposium on educational technology
10. Qiao L, Zhang R (2015) Personalized recommendation algorithm based on situation awareness. ISBN: 978-1-4799-1891-1
11. Luo L, Lv T, Chen X (2013) Network selection based on context-awareness services
12. Jiang Q, Tang R, Liu P, Qiu Y, Xu H (2014) Research on dynamic data fusion algorithm based on context awareness. ISBN: 978-1-4799-2030-3

An Adaptive K-Means Segmentation for Detection of Follicles in Polycystic Ovarian Syndrome in Ultrasound Image



N. S. Nilofer and R. Ramkumar

Abstract One among the frequent disorder existing in females is caused due to the hormonal change in reproductive age group which is called as polycystic ovarian syndrome (PCOS). PCOS is mostly interrelated with type 2 diabetes mellitus, obesity in addition to high cholesterol levels, hence it is necessitated to detect in early stage besides treatment. Various forms of ovulatory failure require to be recognized and diagnosed ensuing to infertility, which is regarded a significant part. The characteristic of PCOS is that several follicles are formed in the ovary, and this may be regarded as an endocrine disorder. The various effects due to this disorder are cardiovascular disease, obesity, diabetes, and infertility. Ultrasound imaging has an eminent role in PCOS diagnosis since significant information about the number of follicles in addition to size is acquired. The follicles obtained through manual detection may be prone to error and laborious. An adaptive K-means clustering method is greatly utilized for small follicles recognition and rapid segmentation. The testing is accomplished on PCOS ultrasound images, and it is validated that this method outperforms well in contrary with prevailing methods.

Keywords Ultrasound image · Polycystic ovary syndrome · Adaptive K-means clustering

1 Introduction

Most of the women in reproductive age predominantly get affected due to frequent endocrine disorder namely polycystic ovary syndrome (PCOS). The range of PCOS occurrence particularly in India is around 3.7–22.5% which is on the basis of population studied besides criteria exploited for diagnosis [1]. Diagnosis processes are

N. S. Nilofer (✉)

Computer Science, Nandha Arts and Science College, Erode, India

e-mail: nilofers7@gmail.com

R. Ramkumar

School of Computer Science, VET Institution of Arts and Science Erode, Erode, India

acquired through the number of follicles present in the ovary [2]. Thus, ultrasound imaging is greatly utilized for scanning of ovary.

The automatic follicles segmentation such as active contour method [3], edge-based method [4], object growing method [5], and morphology [6] is derived through various researches. These researches are not the end, still there are lot of scope to be explored in this area.

A chronological way of presentation for discussing the previous work made on ovarian ultrasound image processing is as follows: Samsi et al. [7] proposed follicles identification system in automatic manner in IHC stained tissue segments. Dong et al. [8] exploited genetic programming and rotationally invariant local binary patterns for corpora lutea (CL) segmentation and detection.

Potočnik et al. [9] presented a survey on prevailing computer methods for detection, recognition, and analyses of follicles in two-dimensional (2D) and 3D ovarian ultrasound recordings. In [10], fuzzy logic is utilized for follicle detection.

Gopalakrishnan and Iyapparaja [3] utilized modified Otsu threshold value for automatic follicles discovery. In [11], ovarian tissue follicles counting is achieved in automatic way through a novel methodology. CR-Unet is presented in [12], for simultaneous segmentation of ovary and follicles in transvaginal ultrasound (TVUS).

As a result, this research work is focussed on the basis of clustering technique namely adaptive K-means clustering. The paper structure is specified here: Sect. 2 gives the outline of associated work on segmentation for detection of follicles in PCOS. Section 3 describes about adaptive K-means clustering algorithm-based identification of follicles in ovary ultrasound images. Section 4 discusses about experimental results, and lastly Sect. 4 confers conclusion with future work.

2 Proposed Methodology

The steps comprised in suggested approach are explained in this section. Initially, preprocessing phase is done mainly for speckle noise reduction. Adaptive K-means is greatly utilized for follicle area discrimination and identified area marking is also accomplished. Here, the objects confers to follicles. The overall workflow is depicted in Fig. 1.

2.1 Data Collection

Nineteen ultrasound images of ovaries from two radiology center, i.e., Gama Imaging and Diagnostic Center, Singatala, Malda, West Bengal, India, and Swagat Diagnostic Centre, Dhubri, Assam, India, respectively, are collected for this research work. These images are considered as reference for the experimentation. This suggested technique is utilized for prediction of number of follicles as performed manually for the input depicted in Fig. 2.

Fig. 1 Overall flow diagram of follicle detection using adaptive K-means

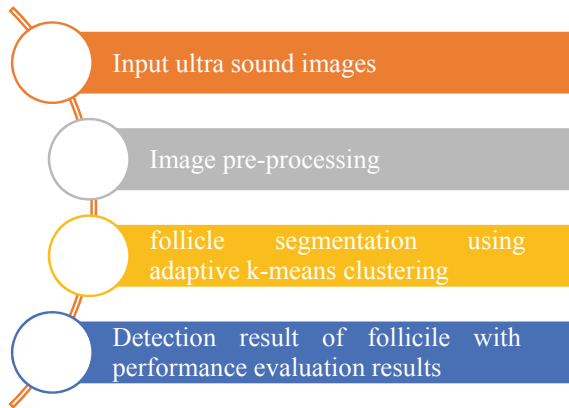
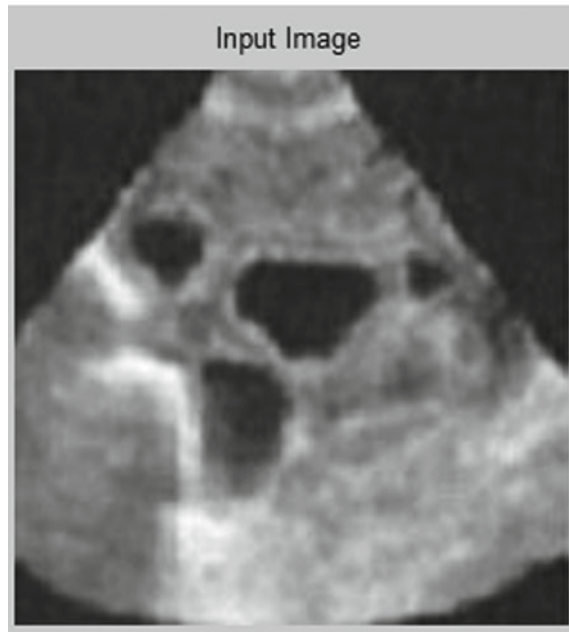


Fig. 2 An input image



2.2 Image Preprocessing

Preprocessing is a process which deals with image improvement having lowest insight level. The pixel area clarity is greatly achieved through the below four steps:

- Pixel brightness alterations
- Geometric alterations

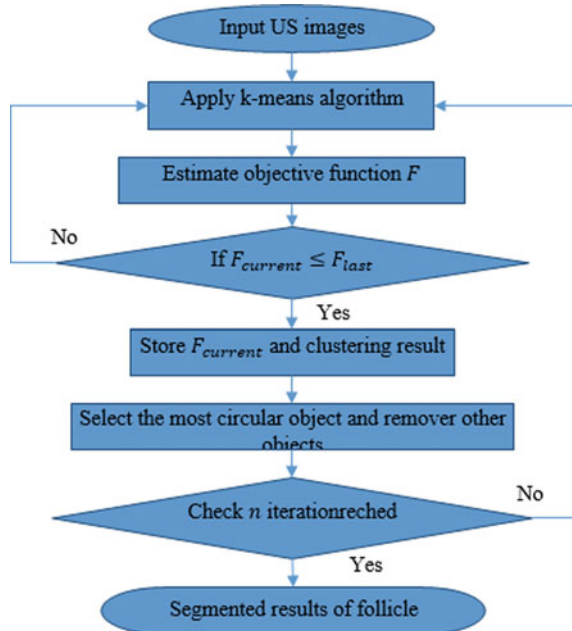
- Utilizing local region of the treated pixel for preprocessing methodologies
- Image renovation necessitates information nearly the whole image.

2.3 Adaptive k-mean Segmentation Approach

To diagnose PCOS in women, this research work employs the adaptive K-means segmentation methodology for segmenting the follicle US images, where two features (apart from conventional K-means) have additionally involved during the segmentation, namely brightness and roundness. The proposed method has divided into two significant steps, i.e., the initialization phase and the adaptive segmentation phase as represented in Fig. 3.

Initialization phase: An objective function has estimated through the image clustering process by applying the K-means for a least iteration count n (8 iterations). This entire phase has been considered as the first iteration in the proposed adaptive segmentation strategy. For the generation of initial new clusters, the new centers have been distributed. During this iteration, the estimation of objective function has been carried out for the first time. Here, the average circularity of the overall obtaining objects (i.e., follicle) has been referred as the objective function. The ratio of the area of the shape (ovary) A to the area of a sphere with equal perimeter Pr has been contemplated as the roundness ratio. It can be formulated as follows,

Fig. 3 Flowchart of adaptive K-means-based segmentation process



$$f_o = \frac{4 * \pi * A}{Pr^2} \quad (1)$$

In f_o function, its value is one for circle; its value is less than one for all other shapes. In every individual cluster, the area has been derived through the overall count of pixels n . The subsequent equation expresses the estimation of objective function F ,

$$F = \sum_{i=1}^n \frac{f_{oi} * A_i}{A_i} \quad (2)$$

Here, the roundness ratio of i th object (follicle) has been denoted by f_{oi} ; the area of i th object (follicle) has been signified by A_i . In order to maximize the values of large object and to minimize the values of small objects, the roundness ratio of the object has been multiplied by the object area.

Adaptive segmentation phase: Besides, recalculation of F has been carried out by the feature-based computation using Eq. 2, subsequently the current F_{current} has been compared with the F_{last} of the previous iteration. Then, the current state has been kept, if the new F_{current} has the lower value than the previous, by eliminating the previous results; if the current goodness is poor (i.e., new goodness value is higher than the previous one), the generation of new centers has been initiated to create new clusters through the brightest two clusters (i.e., target objects), and archive remaining clusters deprived of modification with the help of distance (d) function that has been considered as the metric to estimate the distance amid a data point x and cluster center C . Though this process, the distance within the n data points and their corresponding cluster centers has indicated, which has been expressed as given below,

$$d = \|x_j^j - C_j\|^2 \quad (3)$$

This entire process has been iterated until it converges with either one of the two criteria, (i) the result of goodness should remain unchanged and (ii) should reach the number of algorithm iterations n . Ultimately, the application of object selection process accomplishes the identification of highly circular objects (circularity ratio close to 1). Algorithm 1 defines the steps involved in the suggested adaptive segmentation algorithm.

Algorithm 1. The proposed adaptive K-means segmentation algorithm.

Input: US images

Output: segmentation outcome of follicle

Activation step

Call K-means clustering [13] for n iterations

for all connected object (US images) proceed

Identify roundness ratio by Eq. (1)

End for

Determine objective function F by Eq. (2)

Replicate

Call K-means method starting from the last clustering result for n iterations

Determine objective function F by Eq. (2)

If $F_{\text{new}} \leq F_{\text{last}}$ then archive F_{current} as new F

Archive optimum clustering results of segmentation as depicted in Fig. 5. Else

Call the new center creation by Eq. (3), End if

Iterate the entire process until the convergence of n

For each resulted US images proceed

estimate roundness ratio by Eq. (1)

If roundness ratio is not close to 1 then

Eliminate other object Else

Take the most circular object outputs of segmented follicle results, as represented by Figs. 5 and

3 Experimental Results and Discussion

This section assesses the efficiency of the suggested DBN, besides compares the outcomes of the performance under prevailing strategy, namely morphological operation [3] and proposed adaptive K-means, during which the accuracy, sensitivity, specificity, and segmentation accuracy have taken as the performance parameters. The results of segmentation of follicle are given in Fig. 4a–c.

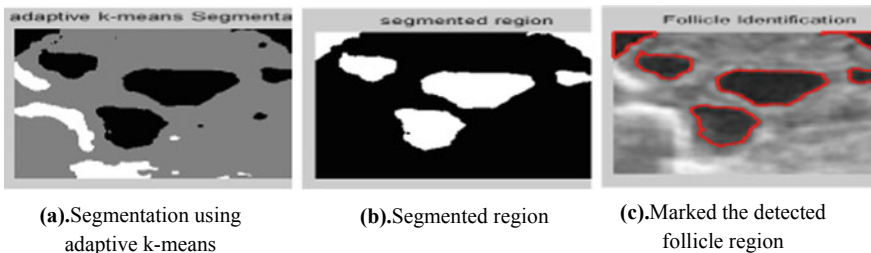


Fig. 4 Segmentation results

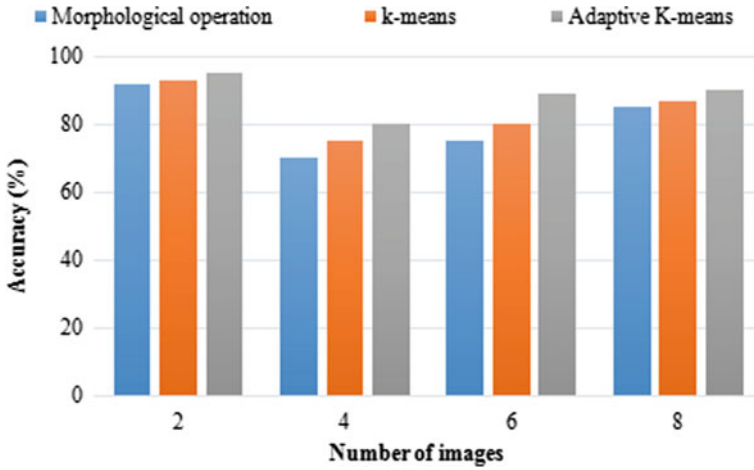


Fig. 5 Result of accuracy

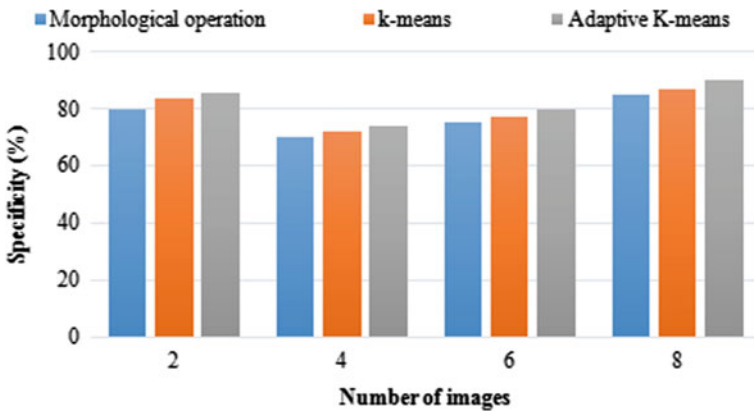


Fig. 6 Result of specificity

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

In which, the samples of TP have appropriately been categorized as no follicle; the samples of FP have inappropriately been categorized as follicle; the samples of

TN have appropriately been categorized as follicle; and, the samples of FN have inappropriately been categorized as follicle.

Segmentation Accuracy: Define I_n as the quantity of nonzero pixel in the nucleus of output; consider IM_n as the quantity of pixels belongs to manually segmented follicle. The estimation of the accuracy for follicle segmentation can be expressed as follows,

$$\text{segmentationaccuracy} = \left[1 - \frac{IM_n - I_n}{IM_n} \right] * 100\%$$

3.1 Accuracy Comparison

In Fig. 5, the proposed adaptive K-means establishes its proficiency to secure 95% of accuracy to choose the cluster center, which is superior to the existing methodologies, such as the morphological operation, and K-means as they solely obtain 92% and 93% of accuracy, respectively. Through this result, it can be observed that the competency of the conventional K-means clustering algorithm has considerably enhanced by the proposed strategy.

3.2 Specificity Comparison

Figure 6 shows the proposed adaptive K-means establishes its capability to secure 90% of specificity to choose the cluster center, which is combatively higher than the morphological operation, and K-means methodologies, as they solely obtain 85% and 87% of specificity, respectively. The empirical findings depict the adeptness of the proposed approach to assist with medical image segmentation, especially the follicle segmentation and recognition, as the significant follicle types are identically spherical. It proves that the medical image analysis and medical diagnosis can utilize the proposed approach efficiently.

3.3 Sensitivity Comparison

Figure 7 shows the proposed adaptive K-means establishes its competency to procure 95% of specificity and identify the cluster center. On the other hand, 80% and 85% of specificity have achieved by morphological operation, and K-means, respectively. The segmentation has processed through the adaption concept in each iteration and sustained the optimal outputs.

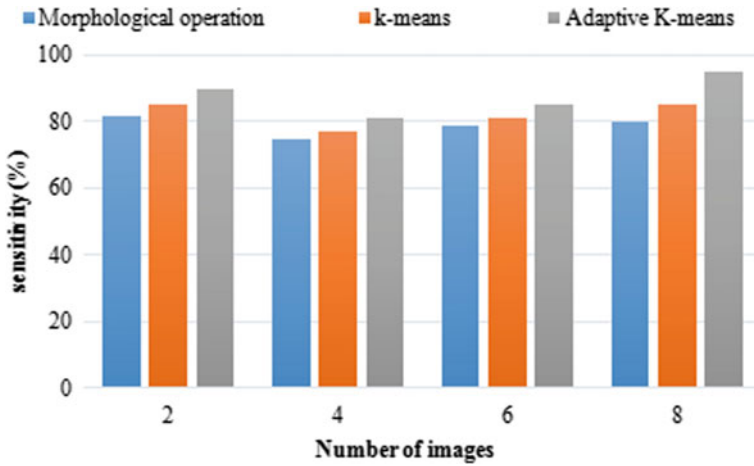


Fig. 7 Result of sensitivity

3.4 Segmentation Accuracy Comparison

In Fig. 8, the rate of segmentation accuracy proportionally improves in accordance with the increasing number of images. For the detection of follicle, the suggested adaptive K-means proves its efficiency to obtain 95% of segmentation accuracy and identify the cluster center, which is superior to other existing approaches. The reason is that, the follicle has been segmented and identified in the image clustering process

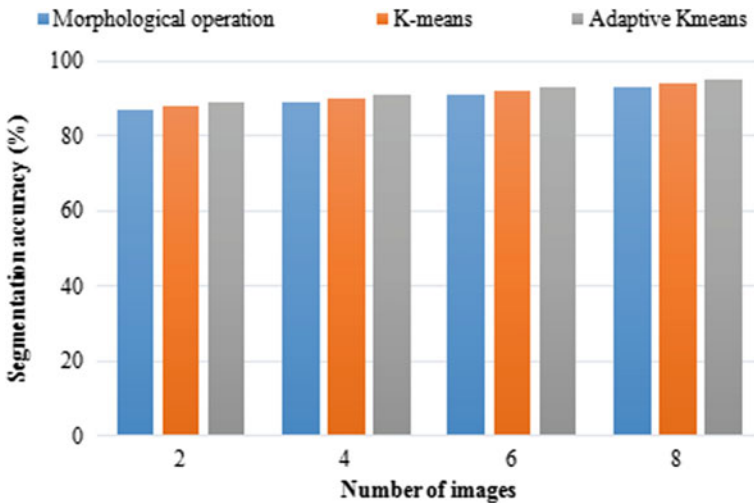


Fig. 8 Result of segmentation accuracy

itself, as it does not require further operations to identify the target objects (follicle) after clustering, whereas the prevailing approaches, like morphological operation and K-means, hold 93%, and 94% of segmentation accuracy.

4 Conclusion and Future Work

For automatically identifying the follicle, this study introduces a statistical approach. For segmentation purpose, an adaptive K-means clustering has further involved by the proposed algorithm. The suggested approach proves its adequacy being a basis for automatic follicle identification. This study can get further extended through focussing on the advancements of the image preprocessing task, and the appropriate attainment of image information by accelerating the algorithm through significantly decreasing the pixel quantity, in future. Besides, the adaption of various clustering algorithms (e.g., bio-inspired algorithms) must be considered, and additional objective functions will be included to surpass the new challenges in medical imaging.

References

1. Ganie MA, Vasudevan V, Wani IA, Baba MS, Arif T, Rashid A (2019) Epidemiology, pathogenesis, genetics & management of polycystic ovary syndrome in India. *Indian J Med Res* 150(4):333
2. Eskandari H, Azar RZ, Pendziwol L (2017) U.S. Patent No. 9,679,375. U.S. Patent and Trademark Office, Washington, DC
3. Gopalakrishnan C, Iyapparaja M (2019) Active contour with modified Otsu method for automatic detection of polycystic ovary syndrome from ultrasound image of ovary. *Multimed Tools Appl* 1–24
4. Hiremath PS, Tegnoor JR (2010) Automatic detection of follicles in ultrasound images of ovaries using edge based method. *IJCA Special Issue on RTIPPR* 2:120–125
5. Deng Y, Wang Y, Shen Y (2011) An automated diagnostic system of polycystic ovary syndrome based on object growing. *Artif Intell Med* 51(3):199–209
6. Padmapriya B, Kesavamurthy T (2016) Detection of follicles in poly cystic ovarian syndrome in ultrasound images using morphological operations. *J Med Imag Health Inform* 6(1):240–243
7. Samsi S, Lozanski G, Shanarah A, Krishanmurthy AK, Gurcan MN (2010) Detection of follicles from IHC-stained slides of follicular lymphoma using iterative watershed. *IEEE Trans Biomed Eng* 57(10):2609–2612
8. Dong M, Eramian MG, Ludwig SA, Pierson RA (2013) Automatic detection and segmentation of bovine corpora lutea in ultrasonographic ovarian images using genetic programming and rotation invariant local binary patterns. *Med Biol Eng Comput* 51(4):405–416
9. Potočnik B, Cigale B, Zazula D (2012) Computerized detection and recognition of follicles in ovarian ultrasound images: a review. *Med Biol Eng Comput* 50(12):1201–1212
10. Hiremath PS, Tegnoor JR (2014) Fuzzy inference system for follicle detection in ultrasound images of ovaries. *Soft Comput* 18(7):1353–1362
11. İnik Ö, Ceyhan A, Balçoğlu E, Ülker E (2019) A new method for automatic counting of ovarian follicles on whole slide histological images based on convolutional neural network. *Comput Biol Med* 112:103350

12. Li H, Fang J, Liu S, Liang X, Yang X, Mai Z, Ni D (2019) CR-Unet: a composite network for ovary and follicle segmentation in ultrasound images. *IEEE J Biomed Health Inform* 24(4):974–983
13. Kiruthika V, Ramya MM (2014) Automatic segmentation of ovarian follicle using K-means clustering. In: 2014 fifth international conference on signal and image processing. IEEE, pp 137–141

Image and Signal Processing

Two Layers Machine Learning Architecture for Animal Classification Using HOG and LBP



Sandeep Rathor, Shalini Kumari, Rishu Singh, and Pari Gupta

Abstract There is a social relation between animals and humans. Both are directly or indirectly dependent on each other. Therefore, it is our duty to maintain the healthy existence of both. To maintain this situation, we have to develop an efficient model that is able to protect the life of animals. To do the same, first it is necessary to identify different animals so that we can concentrate more on the endangered species (animal). But due to various varieties of animals, we need a system to recognize such type of animals automatically. In this paper, we propose a system based on two-layered machine learning architecture by using histogram of oriented gradients (HOG) method and local binary pattern (LBP) to perform the detection of animal. SVM and gradient boosting classifiers are used on different layers to classify the object efficiently. The result analysis shows that our proposed model works efficiently and effectively with an acceptable accuracy, i.e., 95.15%.

1 Introduction

The existence of animals and humans depends on each other. But due to extensive deforestation, rapid urbanization, and continuously changing environment, shelters of many animals are getting diminished [1]. As a result of which, numerous species are getting endangered and some of them are even becoming extinct. Poaching is also one of the significant reasons for the endangerment of animals, and in addition to this, the fast-growing automobile industry possesses an all new threat for animals. According to the recent reports, the number of animals getting killed in road accidents is rising alarmingly [2]. Such animal-vehicle collisions not only affect the life of the animal, but also affect the life of the driver and passengers of the concerned vehicle [3]. These accidents are on the rise mainly due to the increased number of vehicles on the roads and this number continues to increase exponentially [4]. The roots of all these problems are, in one form or other, connected to overpopulation because as the demand increases, mankind exploits the natural resources to fulfill its needs and

S. Rathor (✉) · S. Kumari · R. Singh · P. Gupta
GLA University, Mathura, India
e-mail: sandeep.rathor@gla.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 192,
https://doi.org/10.1007/978-981-33-6546-9_42

445

eventually, instead of making the situation better, we end up creating new problems for ourselves. It is required to bring this devastating situation under control before it is too late. Therefore, there is a need to carefully examine the present situation and search for the possible solutions. It can be observed that the prominent factors of this problem are poaching and road accidents of animals, which means that there is an urgent need to ensure the safety of animals from the menace of poaching and road accidents. To overcome the problem of animal–vehicle collisions (AVCs) we will need to develop an efficient road safety and alert system for detection of animals at the roadside so that further action can be taken. To handle the situation, we need a system which can perform the task of detection and classification of animals efficiently in minimum possible amount of time.

A lot of work has been done in the context of face detection and human detection and it has been successful too but animals have totally different facial features as well as characteristics compared to humans, and therefore, we need to work separately on animal detection. One of the major problems for animal detection is the presence of large varieties of animals, due to which we cannot at random make the system that detects all types of animals at once. Therefore, in this paper, we propose a model with two-layered architecture which is efficient enough to classify the animals with acceptable accuracy.

2 Related Work

Researchers have already proposed their research on human and vehicle detection; however, animal detection has not yet been explored in its full depth. Animal kingdom is a subject of great research because there are a large number of species with even a greater number of differences. However, some researchers have focused in this context as:

An animal–vehicle collision avoidance system is proposed by [2]. In this paper, author focused on the animal “moose” only. They used LBP and SVM classifiers. The results only cater to the detection of moose and all other animals are not considered in this paper. The research with an Arduino board and infrared sensor is proposed by [4]. In this paper, camera is set up on the front of the vehicle to capture the images of animals on the road. The images captured by the camera are matched with the animal database to check if an animal is approaching the vehicle. If an animal is detected, the LED is turned on and a hooter is sounded indicating the presence of an animal. This system is developed for the self-driving cars as well as the normal driver-driven vehicles but it is not efficient for all the animals and group of animals. Object detection framework using Haar technique is proposed by [5]. In the paper, author discussed the use of Haar wavelet to create the features. Initially, the process is focused to get adjoining regions at a location and then calculating the sum of the pixel intensities in that particular region. This method is good for extracting feature but due to the limited instances in the dataset this method is not efficient for animal detection. Baluja et al. [6] used neural network-based system for upright

face detection. The author divided image into various small windows and test if each window contains a face or not. The proposed model works only on human face detection and cannot work with the animal face detection. Felzenszwalb et al. [7] proposed a learning model to detect the objects. Jian Sun et al. [8] proposed animal head detection method. In this paper, the author tried to detect the animal heads of large head animals and used Haar of oriented gradients approach. After getting the image, they implemented two algorithms: deformable detection and brute force for extracting the shape and texture features. Idea on getting distinctive features from images is proposed by [9]. In this paper, the author performed the matching between different views of an object by using nearest-neighbor algorithm and Hough transformation. The results are satisfactory but cannot be used for animal detection. A new approach, histogram of oriented gradient (HOG) method, was proposed by Dalal et al. [10, 11] for human detection. In this paper, calculation of histogram is done by dividing the particular image into a number of cells. The research was mainly focused on human and pedestrian identification and not suited for animal detection. A trainable system using sliding window approach for detection of faces as gray image is proposed by [12]. Advantage of this system is that there is no need of manual interaction for choosing and extracting components. In this paper, the author trained a support vector machine (SVM) classifier to perform the detection. The method is not animal concerned.

Various researchers have proposed a number of methods in this context; however, it has its own drawbacks. In this paper, we propose a model with two-layer architecture which is efficient enough to classify the animals with acceptable accuracy.

3 Proposed Work

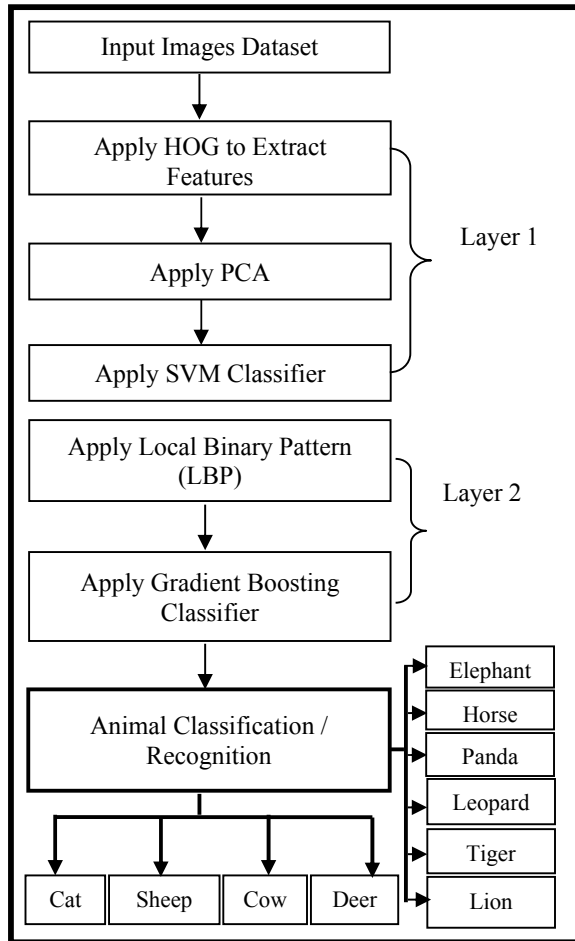
Our main aim is to successfully detect the animal from front (face) as well as side view. We as a human can identify an animal from different angles but for a machine we need to define the parameters through which it can detect the animal. Therefore, we are going to propose a model which can detect animals from various views with much more precision and accuracy.

We proposed two-layered architecture in which the first layer consists of three operations, i.e., HOG, PCA, and SVM. Feature extraction is done through HOG and principal component analysis (PCA) is used to reduce the features while SVM is used for classification purpose. The second layer comprises LBP and gradient boosting classifier. LBP is a simple texture descriptor technique based on binary numbers which consumes minimum time for processing and another classifier is used for the final predictions, i.e., gradient boosting. This layered architecture provides greater efficiency and accuracy. The whole process is represented in the Fig. 1.

A. Dataset Used

We have created our own dataset by combining the images from various available datasets such as Pascal VOC12c, KTH [13] and divided the image dataset in the ratio

Fig. 1 Proposed model for animal classification



of 70:30, i.e., 70% for training and 30% for testing purpose. Table 1 shows the total number of images of different animals along with training and testing dataset.

Table 1 shows that we have taken total 3483 images of elephant out of which 2438 images (70%) are used for training and 1045 images (30%) are used for testing purpose. In the same context, we have taken images for other nine animals as shown in Table 1.

B. Layered Architecture in the proposed Model

The proposed model consists of two layers. In the first layer, HOG technique is applied to describe the overall features of the particular object. It is basically for the processing of image for detecting the object. In the proposed model, HOG descriptor is used to find the texture, appearance, and shape. After that, image is divided into a number of cells which are further divided into pixels, and for each cell, a histogram is

Table 1 Total number of images (instances) in a dataset

Name of animal	Total number of images	Images for training	Images for testing
Elephant	3483	2438	1045
Horse	3190	2233	957
Panda	3203	2242	961
Leopard	2986	2090	896
Tiger	3246	2272	974
Lion	3150	2205	945
Cat	3743	2620	1123
Sheep	2943	2060	883
Deer	3283	2298	985
Cow	3263	2284	979

complied. Now, concatenate these histograms with the descriptor. For achieving high accuracy, these histograms are normalized by calculating the measure of intensity in the larger portion of the image, known as block, and after that we use this block to normalize all the cells residing in this block. To reduce the features, apply PCA and train the SVM for classification. The results are passed to the next layer.

In the second layer, local binary pattern (LBP) method is applied [14]. To calculate LBP, the proposed model divides the window into cells. After this, for each cell, the center pixel is compared with the adjacent 8 pixels. If the adjacent pixel contains a value larger than the center pixel, then the pixel value is changed to “1”, otherwise it is changed to “0”. Now, the weight of the cell is calculated. To calculate the weight of the cell, we traverse it in an anticlockwise direction starting from the mid-left pixel. We get an 8-bit binary number. For example, consider the following cell (Table 2):

Suppose the weight of the cell in binary = 00100101 (anticlockwise direction starting from mid-left cell) then, the weight of the cell in decimal = $1 \times 2^0 + 0 \times 2^1 + 1 \times 2^2 + 0 \times 2^3 + 0 \times 2^4 + 1 \times 2^5 + 0 \times 2^6 + 0 \times 2^7$. Therefore, we can calculate the result as $1 + 4 + 32 = 37$

The calculated 8-bit binary number can be any number in the range 0 to 255. So, if we create the histogram of all the cells, we will get a 256-dimensional feature. Then gradient boosting classifier is used to obtain the final results. It has shown great results and we are able to classify different animals from the front along with the side view also.

4 Result and Discussion

The proposed model shows good results in case of head detection as well as side body detection of animals as in Fig. 2.

Table 2 Sample execution of LBP in the proposed model

<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td style="padding: 5px;">101</td><td style="padding: 5px;">67</td><td style="padding: 5px;">121</td></tr> <tr><td style="padding: 5px;">45</td><td style="padding: 5px;">99</td><td style="padding: 5px;">56</td></tr> <tr><td style="padding: 5px;">12</td><td style="padding: 5px;">132</td><td style="padding: 5px;">88</td></tr> </table>	101	67	121	45	99	56	12	132	88	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td style="padding: 5px;">1</td><td style="padding: 5px;">0</td><td style="padding: 5px;">1</td></tr> <tr><td style="padding: 5px;">0</td><td style="padding: 5px;">X</td><td style="padding: 5px;">0</td></tr> <tr><td style="padding: 5px;">0</td><td style="padding: 5px;">1</td><td style="padding: 5px;">0</td></tr> </table>	1	0	1	0	X	0	0	1	0	
101	67	121																		
45	99	56																		
12	132	88																		
1	0	1																		
0	X	0																		
0	1	0																		
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">$1x2^0$</td> <td style="padding: 5px;">$0x2^1$</td> <td style="padding: 5px;">$1x2^2$</td> </tr> <tr> <td style="padding: 5px;">$0x2^7$</td> <td style="padding: 5px;">X</td> <td style="padding: 5px;">$0x2^3$</td> </tr> <tr> <td style="padding: 5px;">$0x2^6$</td> <td style="padding: 5px;">$1x2^5$</td> <td style="padding: 5px;">$0x2^4$</td> </tr> </table>			$1x2^0$	$0x2^1$	$1x2^2$	$0x2^7$	X	$0x2^3$	$0x2^6$	$1x2^5$	$0x2^4$									
$1x2^0$	$0x2^1$	$1x2^2$																		
$0x2^7$	X	$0x2^3$																		
$0x2^6$	$1x2^5$	$0x2^4$																		

As two-layer architecture is used in the proposed research, therefore, the results obtained from the first layer and second layer are illustrated using two different confusion matrices as Fig. 3 and Fig. 4, respectively.

$$Accuracy = \frac{\sum \text{True Predications}}{\text{Total Instances}} \tag{1}$$

The accuracy of the proposed model can be obtained by using Eq. (1). The first layer was 88.59% and after the second layer, we got 95.15%. The comparison of other animal detection model is shown in Fig. 5. It shows that our proposed model has very high accuracy than any state-of-the-art model. The measure parameters (like precision, recall, and F-score) on the basis of Fig. 4 can be calculated and the same is represented in Table 3 as:

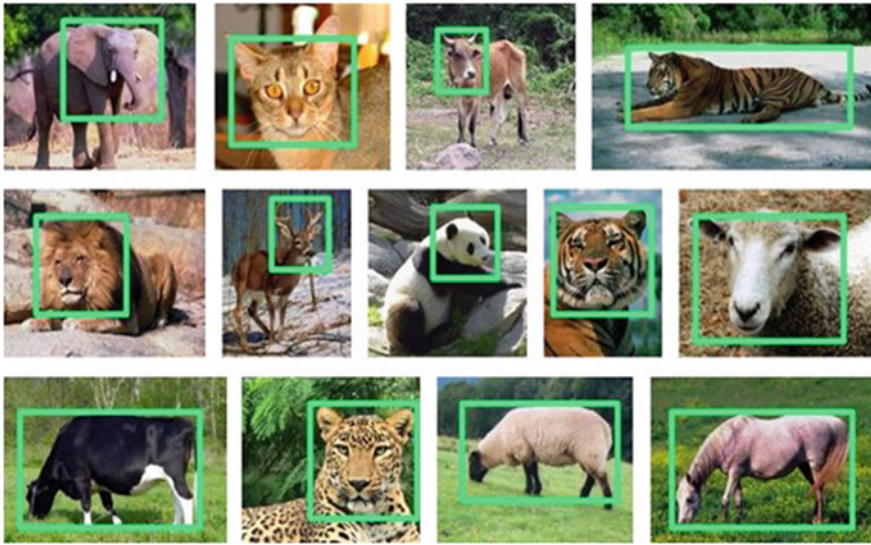


Fig. 2 Some experimental results of animal detection

	Elephant	Horse	Panda	Leopard	Tiger	Lion	Cat	Sheep	Deer	Cow
Elephant	1004	5	7	5	6	4	3	2	2	7
Horse	8	874	4	12	11	8	6	7	14	13
Panda	17	7	863	8	9	9	15	8	7	18
Leopard	10	11	14	752	24	25	17	18	12	13
Tiger	7	12	6	22	868	20	13	8	9	9
Lion	9	8	9	21	23	841	10	8	9	7
Cat	10	14	17	18	19	18	967	35	13	12
Sheep	16	11	15	18	16	17	26	739	13	12
Deer	12	25	11	14	19	15	14	11	845	19
Cow	17	9	12	7	9	8	6	9	19	883

Fig. 3 Confusion matrix after applying the first layer HOG + SVM

5 Conclusion

Recognizing the animals is the basic requirement to save the animals. In this paper, we have proposed two layers machine learning architecture for animal classification using HOG and LBP features. In the first layer, SVM classifier is used to classify the animal and the results are passed to the next layer where gradient boosting classifier is used for better classification. The experiment is performed on 32490 well-labeled animal images and the proposed model works with acceptable accuracy of 95.15%.

	Elephant	Horse	Panda	Leopard	Tiger	Lion	Cat	Sheep	Deer	Cow
Elephant	1026	3	2	0	1	2	0	0	4	7
Horse	2	913	1	3	5	2	2	4	15	10
Panda	5	2	929	3	4	2	5	3	1	7
Leopard	2	7	2	841	13	12	7	6	4	2
Tiger	1	5	3	15	922	10	8	3	5	2
Lion	6	4	1	12	14	895	6	2	2	3
Cat	2	3	6	9	8	5	1058	14	12	6
Sheep	1	2	7	3	4	3	16	815	11	4
Deer	3	6	4	3	5	2	14	11	930	7
Cow	3	4	4	2	3	3	4	5	4	947

Fig. 4 Confusion matrix after applying the second layer LBP + gradient boosting classifier

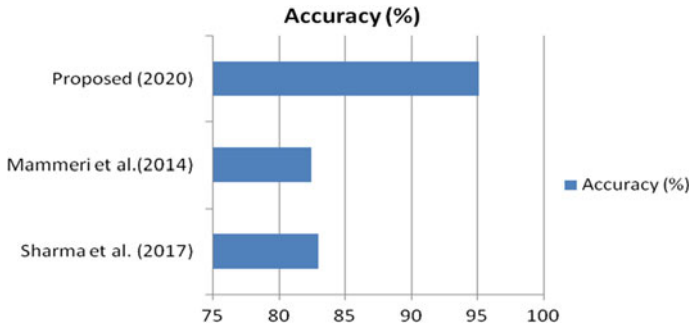


Fig. 5 Accuracy comparison of proposed model with other state-of-the-art models

Table 3 Measure parameter calculations of the proposed model on dataset (LBP + gradient boosting)

Category of animal	Precision	Recall	F-score
Elephant	0.98	0.97	0.97
Horse	0.95	0.96	0.95
Panda	0.96	0.96	0.96
Leopard	0.90	0.94	0.91
Tiger	0.94	0.94	0.94
Lion	0.94	0.95	0.94
Cat	0.94	0.94	0.94
Sheep	0.94	0.94	0.94
Deer	0.94	0.94	0.94
Cow	0.96	0.95	0.95

The size of dataset can be increased in near future to implement the proposed model in real scenario.

References

1. Saad W, Alsayyari A (2019) Loose animal-vehicle accidents mitigation: vision and challenges. 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, Egypt, pp. 359–364
2. Mammeri A, Zhou D, Boukerche A, Almulla M (2014) An efficient animal detection system for smart cars using cascaded classifiers. IEEE ICC 2014
3. Gagnon JW, Dodd NL, Sprague SC, Ogren KS, Loberger CD, Schweinsburg RE (2019) Animal-activated highway crosswalk: longterm impact on elk-vehicle collisions, vehicle speeds, and motorist braking response. *Hum Dimens Wildl* 24(2):132–147. <https://doi.org/10.1080/10871209.2019.1551586>
4. Sibandaa V, Mpofua K, Trimblea J, Zengeni N (2019) Design of an animal detection system for motor vehicle drivers. 29th CIRP Design 2019
5. Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vis* 57:137–154
6. Rowley HA, Baluja S, Kanade T (1998, January) Neural network-based face detection. *IEEE Trans Pattern Anal Mach Intell* 20(1):23–38
7. Felzenszwalb PF (2011) Learning models for object recognition. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2011, Kauai, HI, USA, pp. 1–15
8. Zhang W, Sun J, Tang X (2011) From tiger to panda: Animal head detection. *IEEE Trans Image Process* 20(6):1696–1708
9. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
10. Leibe B, Seemann E, Schiele B (2005) Pedestrian detection in crowded scenes. *Proc CVPR* 1:878–885
11. Dalal N, Triggs B (2010) Histograms of oriented gradients for human detection. *CVPR*, pp. 886–893
12. Papageorgiou C, Poggio T (2000) A trainable system for object detection. *Int. J. Comput. Vis.* 38(1):15–33
13. https://www.csc.kth.se/~heydarma/datasets/animal_database.tar.gz (access dated 21 March 2020)
14. Ahonen T, Hadid A, Pietikainen M (2006, December) Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(12):2037–2041

Machine Learning-Based Detection and Grading of Varieties of Apples and Mangoes



Anuja Bhargava and Atul Bansal

Abstract Computer vision is a consistent and advanced technique for image processing with the propitious outcome an enormous potential. A computer vision has been strongly adopted in the heterogeneous domain. It is also applied to the various domains of agriculture that improve the quality of automation, growth of the economy, and the productivity of the nation. Fruits and vegetable quality highly affects the evaluation of the quality and export market. Recently, automatic visual inspection becomes very important for grading of fruits applications. In this paper, multiple features with support vector machine classifier-based automatic detection and grading of apple and mango are done. Firstly preprocessing is done using histogram equalization to smooth the image. Then, fuzzy c-means clustering is used for segmenting the defected region. Secondly, the combination of statistical, textural, and geometrical features is used to extract the information. Finally, the detection and grading are done using the SVM classifier and achieve accuracy with 98.48 and 95.72%. The agriculture industry achieves the direction of research and support technology for the detection and grading of fruits using multiple features.

Keywords Apple · Mango · Fresh · Rotten · SVM · Features

1 Introduction

Fruit detection and quality grading always endure a hot topic in the agriculture research field. Traditionally, detection and grading are done by labor manually [1]. This will lead to scarcity of consistency and due to shortage of labor results in research for solutions automatically. One of the most difficult processes is to detect and grade fruits visually [2]. Computer vision and image processing are some of the most important techniques used for the identification of features in many agricultural

A. Bhargava (✉) · A. Bansal
Department of Electronics and Communication, GLA, University, Mathura 281406, India
e-mail: anuja1012@gmail.com

A. Bansal
e-mail: atul.bansal@gla.ac.in

products [3, 4]. Different algorithms have been tested and developed to detect and grade fruits automatically [5]. Moradi et al. [6] present the detection of defects in apples using a fuzzy c-means algorithm and achieves 91.00% accuracy. Razak et al. [7] proposed digital fuzzy image processing, content predicted analysis, and statistical analysis to grade mango and achieves 80.00% accuracy. Ashok and Vinod [8] presented quality grading of apples with 83.33% accuracy for probabilistic neural networks. Nandi et al. [9] propose mango grading using a multi-attribute decision theory with a prediction accuracy of 96.00% accuracy. Nandi et al. [10] presented the maturity and quality detection of mangoes using fuzzy and achieve an 87.00% recognition rate. Sahu and Potdar [11] identify defect a maturity of mango fruit using Image Processing Toolbox. Jawale and Deshmukh [12] proposed real-time rotten apple detection using ANN. Naik and Patel [13] presented a fuzzy classifier to grade mango ($L * a * b * \text{color space}$) that achieves 89.00% accuracy. Khan et al. [14] classify apple disease using a genetic algorithm and achieve a 98.10% recognition rate. Nousseir [15] presented the classification and identification of multiple rotten fruits using linear SVM and achieves 96.00% accuracy. Singh and Singh [16] presented a classification of good and rotten apples using texture features with a 98.90% recognition rate. Bhargava and Bansal [17, 18] proposed a quality grading and detection of multiple fruits by machine learning.

This study aims to analyze an algorithm for the detection and grading of fruit and also to identify the performance of fruit algorithms. Furthermore, this study aims to ease meaningful and reliable conclusions. To that end, the proposed method aims to develop an algorithm that detects fruit variety among apple and mangoes and grades under two categories: fresh and rotten. This algorithm is based on a combination of features to attain a successful recognition rate. This work is the first to present a method to detect a variety of fruits in the vision system of agriculture.

2 Proposed Methodology

The purpose of this presented approach is to build a system that detects the fruit then classifies the variety of apples and mangoes among multiple varieties and assesses the sorting into two groups, i.e., fresh and rotten. Figure 1 presents an outline of the proposed method for the detection and categorization of fruits. These steps are explained in the following section.

2.1 Image Acquisition

The algorithm proposed uses two distinct varieties of fruits: apple and mango with five varieties of each fruit. Each set consists of fresh and rotten fruits. The characteristics of the dataset used are shown in Table 1. Figure 2 shows different varieties of fruits used.

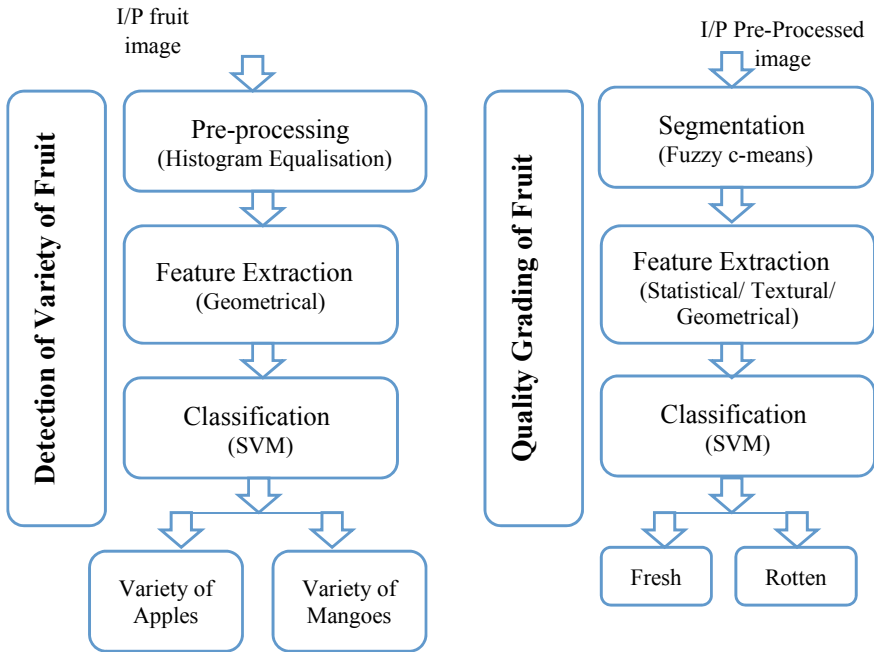


Fig. 1 Basic process for detection and grading of a variety of apples

Table 1 Dataset attributes

S. no.	Authors	Fruit variety	Quantity fresh fruit	Quantity rotten fruit
1	Blasco et al. [19]	Golden delicious	74	26
2	Unay and Gosselian [20]	Jonagold	1120	984
3	Blasco et al. [19]	Multigolden	97	80
4	Purdue Univ. [21]	Fuji	100	60
5	Purdue Univ. [21]	York	100	60
6	Naik [13]	Kesar	40	20
7	Naik [13]	Langdo	40	10
8	Naik [13]	Rajapuri	40	20
9	Naik [13]	Totapuri	40	25
10	Naik [13]	Madrashi aafush	30	30

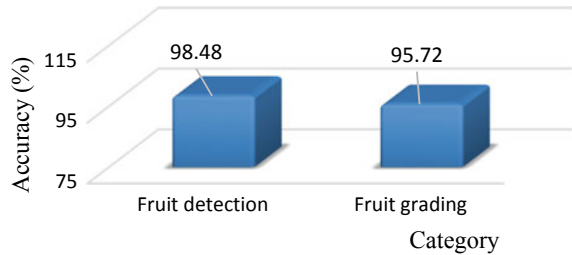
2.2 Preprocessing

Image preprocessing is carried out before the actual analysis [22] to extract particular information. Preprocessing refers to data enhancement for the reduction of distortion



Fig. 2 Sample of a fresh and rotten database used **a** Golden delicious apple, **b** Kesar mango

Fig. 3 Fruit detection and grading accuracy



and noise illumination for the correction of degraded data. It also includes binarization, grayscale conversion, and filtering, smoothing, detection of edges, etc., used for the enhancement of the image.

In this paper, histogram equalization is done using grayscale imaging which results in the whitening of the image. Mathematically, the histogram equation is expressed as:

$$h(v) = \left(\frac{CDF(v) - CDF_{min}}{(A \times B) - CDF_{min}} X(L - 1) \right) \tag{1}$$

where CDF_{min} is cumulative distribution function minimum value, $A \times B$ is some pixels in an image, L is the gray level of the image.

2.3 Segmentation

Image segmentation separates a particular set of pixels into multiple segments from the digital image. The object evaluation is done by separating the background area from the foreground image. Various segmentation methods are utilized such as Otsu segmentation, k-means clustering, fuzzy c-means clustering, and color segmentation. Among all segmentation methods, the important segmentation technique is color segmentation due to which the spot of every disease has a different color.

In this paper, fuzzy clustering is done because it is effective for segmenting images in a controlled environment. In this technique, data points are partitioned into a specific number of clusters. It minimizes the objective function for cluster centroids and given partitioned fuzzy data, n [23].

Mathematically, the membership function,

$$F = \mu_{cd} = 1; \text{ c belongs to d}$$

$$F = \mu_{cd} = 0; \text{ c does not belong to d}$$

The condition to insure that set is exclusive and exhaustive:

$$\sum_{d=1}^k \mu_{cd} = 1, \quad 1 \leq c \leq n \quad (2)$$

$$\sum_{c=1}^n \mu_{cd} > 0, \quad 1 \leq c \leq k \quad (3)$$

$$\mu_{cd} \in \{0,1\}, \quad 1 \leq c \leq n; \quad 1 \leq c \leq k \quad (4)$$

The generalize objective function is given as

$$J(F, Z) = \sum_{c=1}^n \sum_{d=1}^k \mu_{cd}^{\varphi} d_{cd}^2 \quad (5)$$

where

k is number of clusters

n is data point number.

2.4 Feature Extraction

Segmentation develops separated pixels shapes with distinct sizes. The determination of the fruit category depends upon the pixels taken together or independently. Our experiment shows that 13 statistical (mean, RMS, variance, standard deviation, smoothness, skewness, inverse difference moment, kurtosis) and textural (contrast, correlation, energy, homogeneity, and entropy), 14 geometric features (area, eccentricity, major axis length, minor axis length, centroid, bounding box, eccentricity, orientation, convex hull, convex area, solidity, extrema, diameter, extent) [24] are used for best performance in the grading system.

Table 2 Detection and grading accuracy using SVM

S. no.	Category	Performance (%)			Execution time (s)
		Accuracy	Sensitivity	Specificity	
1	Fruit detection	98.48	97.41	99.38	75.34
2	Fruit grading	95.72	93.87	97.02	112.56

2.5 Classification

The extracted features from training images are the input to the classifier. The different characteristics of multiple varieties of apple and mango fruit are learned by the classifier. In this experiment, we use a statistical classifier known as SVM. It is a type of learning system which uses hypothesis space in higher dimensional space of linear function which implements statistical learning theory. It consists of two parts: linearly separable and nonlinear separable. “SVM is a supervised learning method that is based on the minimization procedure of structural risk” [25].

3 Results

The standards of commission European [26] for fruits describes one dismiss and three acceptable conditions. However, ample literature abides of fresh/rotten categorization because of the adversity of the compilation of database and sorting processes. A number of training and testing images used for detection and grading are 1780 and 1216, respectively, in SVM classifier. In the pursuit of detection and grading of apple and mango two category sorting, we have inspected 27 features while SVM is used for classification. The ten datasets of apples as mentioned in Table 1 have been trained with all features. Finally, the grading is done by SVM as shown in Table 2 and Fig. 3.

4 Conclusion and Future Work

In this research, a computer vision-based fruit detection and grading are introduced. Firstly preprocessing is done using histogram equalization to smooth the image. Then fuzzy c-means clustering is used for segmenting the defected region. Secondly, the combination of 27 features is extracted. Finally, the detection and grading are done using SVM classifiers and achieve accuracy with 98.48 and 95.72%. Furtherly, the more concluded and powerful system may be generated with enriched performance in the future. The solution proposed in this work must be tested for multiple fruit images to access its generality, robustness, and accuracy and must be established in a real-life detection and sorting domain.

References

1. Ashraf MA, Kondo N, Shiigi T (2011) Use of machine vision to sort tomato seedlings for grafting robot. *Eng Agric Environ Food* 4(4):119–125
2. Iraj MS, Tosinia A (2011) Classification tomatoes on machine vision with fuzzy the Mamdani inference, adaptive neuro fuzzy inference system based (Anfis-Sugeno). *Aust J Basic Appl Sci* 5(11):846–853
3. Zhang B, Huang Z, Li J, Zhao C, Fan S, Wu J, Liu C (2014) Principle, developments and applications of computer vision for external quality inspection of fruits and vegetables: a review. *Food Res Int* 326–343
4. Dubey SR, Jalal AS (2015) Application of image processing in fruits and vegetables analysis: a review. *J Intell Syst* 24(4):405–424
5. Bhargava A, Bansal A (2018) Fruits and vegetables quality evaluation using computer vision: a review. *J King Saud Univ Comput Inform Sci*. <https://doi.org/10.1016/j.jksuci.2018.06.002>
6. Moradi G, Shamsi M, Sedaaghi MH, Moradi S (2011) Apple defect detection using statistical histogram based fuzzy c-means algorithm. *Iran Conf Mach Vision Image Process*
7. Razak TRB, Othman MB, Bakar MNBA, Ahmad KAB, Mansor AB (2012) Mango grading by using fuzzy image analysis. In: International conference on agricultural, environment and biological sciences, pp 18–22
8. Ashok V, Vinod DS (2014) Automatic quality evaluation of fruits using probabilistic neural network approach. In: International conference on contemporary computing and informatics (IC3I). IEEE, pp 308–311
9. Nandi SC, Tudu B, Koley C (2014) Computer vision based mango fruit grading system. In: International conference on innovative engineering technologies
10. Nandi CS, Tudu B, Koley C (2016) A machine vision technique for grading of harvested mangoes based on maturity and quality. *IEEE Sens J* 16:6387–6396
11. Sahu D, Potdar RM (2017) Defect identification and maturity detection of mango fruits using image analysis. *Am J Artif Intell* 5–14
12. Jawale D, Deshmukh M (2017) Real time automatic bruise detection in (apple) fruits using thermal camera. In: International conference on communication and signal processing, pp 1080–1085
13. Naik S, Patel B (2017) Thermal imaging with fuzzy classifier for maturity and size based non-destructive Mango (*Mangifera Indica L.*) grading. In: International conference on emerging trends & innovation in ICT, pp 15–20
14. Khan MA, Lali MIU, Sharif M, Javed K, Aurangzeb K, Haider SI, Altamrah AS, Akram AT (2019) An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection. *IEEE Access* 7:46261–46277
15. Nosseir A (2019) Automatic classification for fruit types and identification of rotten ones using k-NN and SVM. *Int J Online Biomed Eng* 15(03)
16. Singh S, Singh NP (2019) Machine learning based classification of good and rotten apple. *Recent Trends Commun Comput Electron* 377–386
17. Bhargava A, Bansal A (2019) Quality evaluation of mono & bi-colored apples with computer vision and multispectral imaging. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-019-08564-3>
18. Bhargava A, Bansal A (2019) Automatic detection and grading of multiple fruits by machine learning. *Food Anal Methods*. <https://doi.org/10.1007/s12161-019-01690-6>
19. Blasco J, Aleixos N, Molto E (2003) Machine vision system for automatic quality grading of fruit. *Biosys Eng* 85(4):415–423
20. Unay, Gosselin (2005) Artificial neural network-based segmentation and apple grading by machine vision. In: International conference on image processing
21. Purdue University. <https://engineering.purdue.edu/RVL/Database/IFW/database/index.html>
22. Cubero S, Aleixos N, Molto E, Gomez-Sanchis J, Blasco J (2011) Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food Bioprocess Technol* 487–504

23. Ashok V, Vinod DS (2014) Using K-means cluster and fuzzy C means for defect segmentation in fruits. *Int J Comput Eng Technol* 11–19
24. Ou X, Pan W, Xiao P (2014) Vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM). *Int J Pharm* 460(2):28–32
25. Burges (1998) A tutorial on support vector machines for pattern recognition. *Data Mining Knowl Discov* 2:121–127
26. Anonymous (2004) Commission Regulation (EC) No 85/2004 of 15 January 2004 on marketing standards for apples. *Off J Eur Union L* 13:3–18

Rule-Based Approach for Emotion Detection for Kannada Text



C. P. Chandrika, Jagadish S. Kallimani, H. P. Adarsha,
Aparna Nagabhushan, and Appasaheb Chavan

Abstract In today's world, people have conversation or exchange of information through their local language easily. Since most of the consumers express their views on the product in their local languages, it has become important to build an efficient sentiment analysis model for them. Also, in the field of literature, sometimes we find difficulty in understanding the feelings expressed by the author in their biography or emotions involved in the stories written by them. To solve these kinds of problems, we propose a rule-based methodology to categorize the emotions involved in the documents written in Kannada language. The proposed model gives 85% accuracy which is found to be efficient when compared to the bag-of-words approach which gives 67% accuracy.

Keywords Sentiment analysis · Bag of words · Rule-based · Convolutional neural network · Support vector machine

1 Introduction

Sentiment analysis is one of the evolutionary problems and still a challenging one. Many researchers have contributed different solutions in understanding the sentiment behind the text. One of the major applications of SA is analyzing the consumer

C. P. Chandrika (✉) · J. S. Kallimani · H. P. Adarsha · A. Nagabhushan · A. Chavan
Department of Computer Science and Engineering, M S Ramaiah Institute of Technology,
Bangalore 560054, India
e-mail: chandrika@msrit.edu

J. S. Kallimani
e-mail: jagadish.k@msrit.edu

H. P. Adarsha
e-mail: alliswelladarsha0@gmail.com

A. Nagabhushan
e-mail: aparna98jois@gmail.com

A. Chavan
e-mail: rushi743@gmail.com

opinion on a specific product. Nowadays, many consumers will express their feelings in their native language. It is the need of today's technology to understand their consumers.

Another application of SA is in the field of literature. Many poets or authors have major contributions in their local language. It could be better if we have an efficient tool to understand the emotion behind the written text. To solve these kinds of problems, we propose a rule-based methodology to identify the sentiment behind the text. We have considered the regional language Kannada for this proposed work. This work can help in various online platforms to analyze the textual documents, customer reviews, literature, and many other fields.

Kannada language is morphologically rich. It has become difficult to analyze each part of the text document. In Kannada, the same word can give different meaning with respect to the context. Hence, we need to build an efficient and reliable sentiment analyzer for the Kannada language.

- (1) Bag-of-words approach: This is one of the simple approaches to find the sentiment behind the text. It mainly depends on the frequency of sentiment words appearing in the text.
- (2) Rule-based approach: This approach is based on the context of the sentence and not just on the frequency. The details are provided in the implementation section.

2 Related Work

Many research papers and literature reviews have been carried out on SA, which helped in understanding the various approaches and methodologies of other researches. Several articles are already published only about the idea of sentiment analysis in regional languages and the implementation parts are described in many papers using English and few other languages.

Sentiment analysis by using machine learning methods [1] is discussed, and authors summarize that machine learning methods are the most commonly used as these methods are more diverse and they can also classify both subjective and objective emotions classifications and not required to words, grammars, or sentences to dictionary matching level. In this paper, methods such as SVM, NB, and ME are explained in detail.

Authors in [2] focused on identifying selective data in the Web and differentiating them according to polarity. Authors also provide summarized review about other researchers' approaches such as mathematically incorporating social context into the prediction model, usage of bag of words and its limitations, sentiment analysis on Twitter posts.

CNNs and SVM analysis on sentimental text are described in [3]. First processing of the experimental data is done with segregation and filtering of the words to denoise and they have used a pre-trained word vector to give input; CNN helps to learn the

features in the sentences automatically and SVM as the emotional classifier as well as text classifier in the end.

One of the SA applications can be found in screening for perinatal depression in the field of psychology using deep learning [4]. This method is used for the extraction of emotions and analyzing it by modeling the document for sentiment analysis for findings of the Edinburgh Postnatal Depression Scale, which helps in giving the better solution and suggestions for the patient and improving their performances.

A survey of classification of algorithms used in various research papers on sentiment analysis is illustrated in [5]. They have concluded that there is a vast scope of algorithms such as SVM and Naive Bayes methods and more analysis is required in consideration of textual documents such as customer reviews.

SA on Twitter text using machine learning techniques is discussed in [6]. Using knowledge-based approach and machine learning approach, authors have extracted sentiments in the text. First specific features are extracted to add them in the feature vector. Then the classification accuracy is tested using various algorithmic methods.

Authors in [7] have aimed to review how various technologies have been used to build summarization and perform customer review, record, survey responses, and recommended systems to help the users to extract the emotions and expressions using several algorithms and methodologies and discuss past work done on this field.

A comprehensive survey on the emotion detection for the social media data has been carried out in [8]. By comparing and classifying the expressions in the obtained review data, they have made an application to know the customer needs. This paper concludes that various approaches by the researches provide several challenges in developing decision support systems.

The authors in [9] focused on the stepwise method to analyze the variation in stock prices of many multinational firms. The obtained results are gathered with a slight deviation from the current variation in the stock market. This helps to do a quick analysis on them and give the accurate result in an efficient manner.

The analysis of texts posted on Twitter using case-based algorithms to categorize the expressions and sentence level is discussed in [10]. Together with NLP, they included the ML algorithm as well which is more accurate for the sentiment analysis.

The SA module on Kannada is illustrated in [11]. Authors have classified the sentiments and extended the idea to use semantic methods. They have concluded that these methods aid commercial Web sites to classify Kannada sentences without human interventions.

Analysis of sentence-level text classification for Kannada text is explained in [12]. Authors have used a Naive-based and bag-of-words approach to classify reviews, emotions, questions, and other documents. Depending on the essential requirements of the application, the appropriate methods can be chosen to analyze the sentiments involved in those text information.

From the literature survey, it is evident that most of the work has already been done on English language but less on regional language Kannada. Even though researchers have tried to build efficient SA modules for Kannada using bag-of-words and machine learning algorithms, they have not obtained good accuracy. This motivates us to come with a new technique rule-based approach.

3 Proposed Methodology

The proposed SA module is written in Python 3.6.x. It provides rich and diverse features and adds scalability to the development. Transformation of text to Unicode is easy in Python. Machine learning and deep learning are also facilitated by its features and libraries. This proposed SA model to detect the sentiments in the given text data is divided into following modules as shown in Fig. 1.

After that sentiment is identified and predefined probability is assigned to them.

Overall probability is calculated for the sentiments and output is given.

In this paper, we have followed the rule-based classification approach to identify the sentiments.

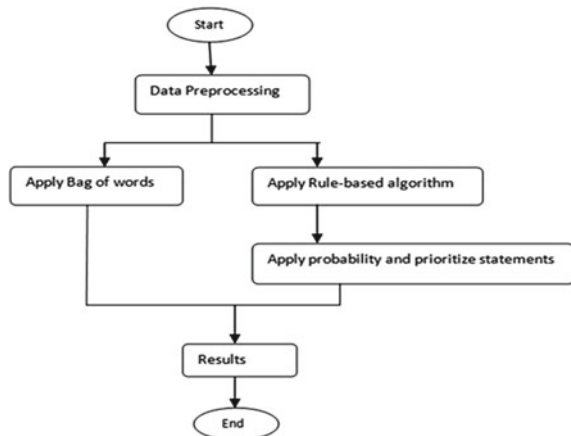
These words are checked whether there is emotion if so their predefined probability is obtained. The obtained probability is changed if the negation for the word is obtained or based on the tense of the word. And then it will go through the predefined rules. These probabilities are stored in a list. At last, based on the probabilities in the list sentiment is calculated.

Explanation of the modules is as follows.

3.1 Data Preprocessing

We have considered a label dataset of 5000 statements which is a mixture of product reviews, article reviews, and people’s emotion. The dataset is cleaned before the processing. Kannada text received from the user is translated into machine level Unicode format. The stop words in the text are removed as it does not help in identifying sentiments. The preprocessed input will be fed to parts of speech tagging module.

Fig. 1 Modules to detect sentiment



Input:ಕಠಿಣವಾದ ಪ್ರಶ್ನೆ ಪತ್ರಿಕೆಯನ್ನು ನೋಡಿ ವಿದ್ಯಾರ್ಥಿಗಳು ಅಸಂತೋಷವಾಗಿರಲಿಲ್ಲ

Input will be fed to preprocessing module, and this module removes stop words like ಈ {Ee}[This], ನೋಡಿ {Nodi}[See], ಆದರೆ {Aadhare} [But],

Output: ['ಕಠಿಣವಾದ{Katinavaada} [Difficult], 'ಪ್ರಶ್ನೆ ಪತ್ರಿಕೆಯನ್ನು {Prashnepatrikeyannu} [Question Paper]', 'ನೋಡಿ'{Nodi}[See], 'ವಿದ್ಯಾರ್ಥಿಗಳು{Vidyarthigalu} [Students], 'ದುಃಖವಾಗಿರಲಿಲ್ಲ{Dukhavagiralilla}[Not sad]].

POS tagging:

It is the process of marking up a word in a text (corpus) as corresponding to a part of speech, based on both its definition and its context. By tagging each word with proper parts of speech, sentiment analyzers can identify words' contribution with respect to context of sentence.

Example:

Input: 'ಕಠಿಣವಾದ', 'ಪ್ರಶ್ನೆ ಪತ್ರಿಕೆಯನ್ನು', 'ನೋಡಿ', 'ವಿದ್ಯಾರ್ಥಿಗಳು', 'ದುಃಖವಾಗಿರಲಿಲ್ಲ'.

The output from the preprocessing module is fed to the parts of speech module. This module will maintain a list of sentiment words with its matching POS tag.

Output: ['ಕಠಿಣವಾದ', 'verb'], ['ಪ್ರಶ್ನೆ ಪತ್ರಿಕೆಯನ್ನು', 'noun'], ['ನೋಡಿ', 'verb'], ['ವಿದ್ಯಾರ್ಥಿಗಳು', 'noun'], ['ದುಃಖವಾಗಿರಲಿಲ್ಲ', 'adjective'].

3.2 Bag of Words

It is the process of identifying specific sentiment words in a given text and assigning probability weightage to that word. In a sentiment text, each word carries different probability weightage, by which overall contribution of that word to the sentiment is calculated.

Example:

Input: ['ಕಠಿಣವಾದ', 'verb'], ['ಪ್ರಶ್ನೆ ಪತ್ರಿಕೆಯನ್ನು', 'noun'], ['ನೋಡಿ', 'verb'], ['ವಿದ್ಯಾರ್ಥಿಗಳು', 'noun'], ['ದುಃಖವಾಗಿರಲಿಲ್ಲ', 'adjective'].

The POS tagged list is forwarded to bag-of-words module. In this module, we have predefined probability values for each sentiment words. These words are identified in the text and each emotion is assigned with its probability weightage between 0 and 1.

Output: ['ದುಃಖ', 'adjective', -0.5], [happy: 0, sad: 0.5, angry: 0, surprise: 0, fear: 0].

Output Sentiment: Sad.

3.3 Rule-Based Algorithm

Rule-based module contains predefined rules for post-processing of sentiment text.

This module divides each word into main word and extension of the word (ಸಂಧಿ).

Sometimes due to association of tenses with sentiment words, the associated word may or may not have the meaning it used to have, hence detecting tense and changing the probability is an important step for getting accurate sentiment results. By using predefined rules for Kannada grammar in rule-based modules, we easily address the above problems.

Example:

Input: ['ಕಠಿಣವಾದ', 'verb'], ['ಪ್ರಶ್ನೆಪತ್ರಿಕೆಯನ್ನು', 'noun'], ['ನೋಡಿ', 'verb'], ['ವಿದ್ಯಾರ್ಥಿಗಳು', 'noun'], ['ದುಃಖವಾಗಿರಲಿಲ್ಲ', 'adjective'].

After dividing word into main word and its extension, rules will be applied. In this example, 'ದುಃಖವಾಗಿರಲಿಲ್ಲ' will be divided into 'ದುಃಖ' and 'ಆಗಿರಲಿಲ್ಲ' {Aagiralilla}[Not at all] as 'ಆಗಿರಲಿಲ್ಲ' comes under negation rule.

Probability of emotion will be downgraded by 40%.

This module contains different rules predefined on the basis of tenses and Kannada grammar. Another example 'ಅತ್ಯಪ್ತಿಯಾಗಿತ್ತು' {Atruptyagittu} [was unhappy] is Kannada word in past tense form, it may or may not indicate the present sentiment of the user. Hence, based on association of tenses with words, we decrease the probability value accordingly (i.e., -0.5 for 'ಅತ್ಯಪ್ತಿಯಾಗಿತ್ತು' will become -0.30 , 40% decrease). After these all five sentiments, probability values are calculated and maintained in a list. The sentiment with highest probability values will be displayed as a result.

['ದುಃಖ' + 'ಆಗಿರಲಿಲ್ಲ', adjective, -0.5].

After applying the rules

Output: ['ದುಃಖ', adjective, -0.2]

Output sentiment: Neutral.

As we did a survey of previous works there was no work which had considered negation while finding sentiment in regional languages. But negation plays a main role as it will reverse the sentiment of the word it is associated with. In our work, we assigned 0.2 to the negated word. The reason we have assigned 0.2 is as a negated word it does not qualify for a normal probability score, so we have kept it at 40% of the original score.

To summarize this, five emotions happy, sad, fear, surprise, and angry are tested in our model. The probability values are kept same for all the categories as 0.5 except for sad it is -0.5 . The negated words will have the value 0.2 and words with past and future tense will have the value 0.35.

To detect negation we are using utf-8 encoding. If word ends with 'ಇಲ್ಲ', then the word is negated. Probability assigned is 0.2 else then the word is not negated then the probability assigned is 0.5.

Unicode for word ends with ಇಲ್ಲ is as follows:

'\xe0\x2\xbf\xe0\x2\x2\xe0\x3\x8d\xe0\x2\x2' (the words ends with ಇಲ್ಲ).

Past and Future Tense:

In a sentiment there may be sentiment in past or future tense and another sentiment in present tense. To give more priority to the sentiment in the present tense, we change the probability of past and future tense to its 70% probability score. So we will assign 0.35 to the word in past or future tense. To detect past and future tense, we are using utf-8 encoding.

If word contains past or future tense then the probability assigned is 0.35.

Else then the word is in present tense Then the probability assigned is 0.5.

Unicode for past and future tense:

[b'\xe0\x2\xa6\xe0\x3\x8d\xe0\x2\xa6\xe0\x3' \times 86',

b'\xe0\x2\xac\xe0\x2\xb9\xe0\x3' \times 81\xe0\x2\xa6\xe0\x3' \times 81',

b'\xe0\x2\xa4\xe0\x3\x8d\xe0\x2\xa4\xe0\x2\xa6\xe0\x3' \times 86'].

An example for this:

ನಿನ್ನೆ {Nenne} [Yesterday], ಆ {Aa} [Due to that],
 ಘಟನೆಯಿಂದಾಗಿ {Gataneyindaagi}[Incident], ನಾನು {Naanu}[I was],
 ತುಂಬಾ {Tumba} [Very much] ದುಃಖಿತನಾಗಿದ್ದೇನೆ {Dukhithanagidde}[Sad],
 ಆದರೆ {Aadhare}[But], ಇಂದು {Indu}[Today], ನಾನು {Naanu}[I am],
 ಸಂತೋಷವಾಗಿದ್ದೇನೆ {Santhoshavagiddene}[Happy].

[0.5, 0.35, 0, 0, 0].

Output: Happy.

The pseudocode for the same is given in Fig. 2.

```

for word in word_list
  find emotion of word
    Assign its probability
    keep count of words in that emotion
    if word is negated
      then reduce its probability score to 20%
    if word is past or future tense
      then reduce its probability score to 70%
  find total number of emotion words in the sentence
  replace
  probability[emotion]=(number_of_words[emotion]*probabil-
ity[emotion])/total_nu
mber_of_words
  output emotion
    maximum probability score and greater than 0.3
  if no emotion with probability score 0.3
    then output sentiment will be neutral

```

Fig. 2 Pseudocode

Some other examples using the same algorithm are given below:

- i. ಮೊಬೈಲ್ {Mobile}[Mobile], ತುಂಬ {Tumba}[Very much], ಸುಂದರವಾಗಿತ್ತು {Sundaravagittu}[Beautiful], ಆಕರ್ಷಣೀಯವಾಗಿತ್ತು {Akarshanavagittu}[Attractive], ಆದರೆ {Aadhare}[But], ಬೆಲೆ {Bele}[Value], ಜಾಸ್ತಿ {Jaasti}[More], ಹೀಗಾಗಿ {Heegagi} [So], ಗ್ರಾಹಕನಿಗೆ {Grahakanige}[Customer] ಖುಷಿಯಾಗಿರಲಿಲ್ಲ {Kushiyagiralilla}[Not happy].

Output:

[ಮೊಬೈಲ್, 'ತುಂಬ', 'ಸುಂದರವಾಗಿತ್ತು', 'ಆಕರ್ಷಣೀಯವಾಗಿತ್ತು', 'ಬೆಲೆ', 'ಜಾಸ್ತಿ', 'ಹೀಗಾಗಿ', 'ಗ್ರಾಹಕನಿಗೆ', 'ಖುಷಿಯಾಗಿರಲಿಲ್ಲ']
[0, 0.2, 0, 0, 0]

Output sentiment: Neutral

Output sentiment is neutral because here ಸುಂದರ and ಆಕರ್ಷಣೀಯ are qualities of a mobile not a sentiment. As the sentence is a negated sentence and there is no evidence of other emotions, the sentence is neutral.

- ii. ಹುಡುಗರು {Hudugaru}[Boys], ಆಟವಾಡುವಾಗ {Aatavaduvaaga}[While playing], ಮಳೆ ಬಂದಿದ್ದರಿಂದ {Male bandiddarinda}[Raining] ನಿರಾಶರಾಗಿರಲಿಲ್ಲ {Nirasharagiralilla}[Was not upset], ಬದಲಿಗೆ {Badalige} [Instead] ಮನೆಗೆ {Manege}[House], ಹೋಗಿ {Hogi}[gone], ಟಿವಿ {T.V}[T.V], ನೋಡಿದರು {Nodidaru} Watched.

Output:

[ಹುಡುಗರು, 'ಆಟವಾಡುವಾಗ', 'ಮಳೆ', 'ಬಂದಿದ್ದರಿಂದ', 'ನಿರಾಶರಾಗಿರಲಿಲ್ಲ', 'ಬದಲಿಗೆ', 'ಮನೆಗೆ', 'ಟಿವಿ', 'ನೋಡಿದರು']
[0, -0.2, 0, 0, 0]

Output sentiment: Neutral

Here, boys stopped playing due to rain, so they started watching television. As they got something to do so they were not sad and instead started doing other things. That is why the sentiment is neutral.

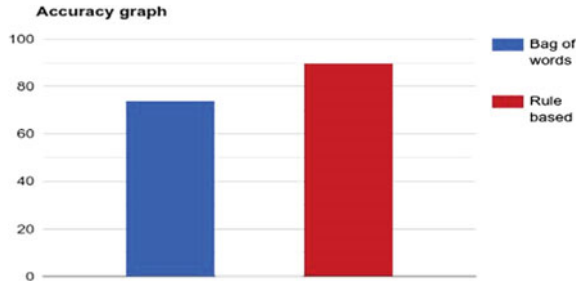
- iii. ಪ್ರಯಾಣ ಮಾಡುವಾಗ {Prayana Maaduvaaga} [While Traveling], ನಿಸರ್ಗದ ಸೌಂದರ್ಯ {Nisargada soundharya}[Nature's beauty], ಅದ್ಭುತವಾಗಿತ್ತು {Adbhuthavaagittu}[Amazing], ಆದರೆ {Aadare}[But], ಆಯಾಸದಿಂದ {Aayasadinda}[due to Fatigue] ನೆಮ್ಮದಿಯಾಗಿರಲಿಲ್ಲ {Nemmadi iralilla}[Not Comfortable].

Output:

[ಪ್ರಯಾಣ, 'ಮಾಡುವಾಗ', 'ನಿಸರ್ಗದ', 'ಸೌಂದರ್ಯ', 'ಅದ್ಭುತವಾಗಿತ್ತು', 'ಆಯಾಸದಿಂದ', 'ನೆಮ್ಮದಿಯಾಗಿರಲಿಲ್ಲ']
[0.2, 0, 0, 0, 0]

Output sentiment: Neutral

At first our rule-based approach was able to find happy sentiment in the statement. But in the end, the passenger was not comfortable due to fatigue which shows sadness, this negated the sentiment. So this will reduce evidence for happiness expressed in the beginning, as we can see both happy and sad emotions at the same time and there is no other sentiment this sentence, the output will be neutral.

Fig. 3 Accuracy graph

4 Obtained Results

We have considered labeled dataset with 5000 statements from 80% of dataset is kept for training and remaining as testing. Once the model is trained and tested with both bag-of-words and rule-based algorithms, we got 67% and 85%, respectively, same is shown in the Fig. 3. From this we can conclude that rule-based approach is an effective one.

If we carefully analyze the literature survey, we noticed that many researchers have tried using different machine learning algorithms, but they have obtained accuracy of 75%. So when we compare our model with the existing models this model works efficiently with sentences that has varied tenses and also extension of nouns and morphemes.

5 Conclusion and Future Work

The proposed technique helps to reduce the complexity in understanding the Kannada sentences and emotions involved in it by a computer system. Rule-based approach proves better than bag of words and even for machine learning model but mainly depends on the dataset. If the dataset is huge then accuracy will increase. Rule-based approach is again based on the user's own opinions, and some users may have different opinion for the above tested sentences. They may want to predict as sad than neutral so it is a debatable one. Future work can be improved with better GUI Web page, much more improved model, involving sarcasm and censored sentences. Also, we can increase the database created, which can help in using the bag-of-words algorithm and extend the technology to other languages with changes in rules based on grammatical structure. We can also put together better rules and sentiments to increase the accuracy of the analyzer.

References

1. Yang P, Chen Y (2017) A survey on sentiment analysis by using machine learning methods. In: IEEE 2nd Information technology, networking, electronic and automation control conference. IEEE, Chengdu, China, pp 117–121. <https://doi.org/10.1109/itnec.2017.8284920>
2. Patil HP, Atique M (2015) Sentiment analysis for social media: a survey. In: 2nd international conference on information science and security. IEEE, Seoul, South Korea, pp 1–4. <https://doi.org/10.1109/ICISSEC.2015.7371033>
3. Chen Y, Zhang Z (2018) Research on text sentiment analysis based on CNN's and SVM. In: 13th IEEE conference on industrial electronics and applications. IEEE, pp 2731–2734. <https://doi.org/10.1109/ICIEA.2018.8398173>, ISBN: 2158-2297
4. Chen Y, Zhou B, Zhang W (2018) Sentiment analysis based on deep learning and its application in screening for perinatal depression. In: IEEE third international conference on data science in cyberspace. IEEE Guangzhou, China, pp 451-456. <https://doi.org/10.1109/DSC.2018.00073>, ISSN: 978-1-5386-4210-8
5. Nidhi HK, Mangat V (2017) A survey of sentiment analysis techniques. In: International conference on I-SMAC. Palladam, India, pp 921–925. <https://doi.org/10.1109/I-SMAC.2017.8058315>
6. Neethu MS, Rajasree R (2013) Sentiment analysis in twitter using machine learning techniques. In: 2013 fourth international conference on computing, communications and networking technologies (ICCCNT). IEEE, Tiruchengode, pp 1–5. <https://doi.org/10.1109/ICCCNT.2013.6726818>
7. Gupta P, Tiwari R, Robert N (2016) Sentiment analysis and text summarization of online reviews. In: International conference on communication and signal processing (ICCSP). IEEE, Melmaruvathur, pp 0241–0245. <https://doi.org/10.1109/ICCSP.2016.7754131>
8. Rajalakshmi S, Asha S, Pazhani Raja N (2017) A comprehensive survey on sentiment analysis. In: Fourth international conference on signal processing, communication and networking (ICSCN). IEEE, Chennai, India, pp 1–5. <https://doi.org/10.1109/ICSCN.2017.8085673>
9. Kaur R, Kaur N, Kaur R, Kaur G (2016) Opinion mining and sentiment analysis. In: 3rd international conference on computing for sustainable global development (INDIACom). IEEE, New Delhi, India, pp 452–455
10. Sarlan A, Nadam C, Basri S (2015) Twitter sentiment analysis. In: Proceedings of the 6th international conference on information technology and multimedia. Putrajaya, Malaysia, pp 212–216. ISBN: 978-1-4799-5423-0. <https://doi.org/10.1109/ICIMU.2014.7066632>
11. Anil Kumar KM, Rajasimha N, Reddy M, Rajanarayana A, Nadgir K (2015) Analysis of users' sentiments from Kannada web documents. In: Eleventh international multi-conference on information. Bangalore, India, pp 247–256. <https://doi.org/10.1016/j.procs.2015.06.029>
12. Jayashree R, Srikanta Murthy K (2011) An analysis of sentence level text classification for the Kannada language. In: International conference of soft computing and pattern recognition. IEEE, Dalian, China. <https://doi.org/10.1109/SoCPaR.2011.6089130>

Gait Recognition Using DWT and DCT Techniques



Shivani and Navdeep Singh

Abstract Gait-based recognition is the new system focused by different researchers. There are two phases of the gait recognition system for successful determination. In the first phase, the features are extracted from the images and also from the video frames. These extracted features are filled into the features vectors. With the help of different classification techniques, features are classified into different classes. The sub-division of these classes will be based on prediction probability. Finally, the prediction outcome of the classifier will be compared to the standard outcome. The proposed system is based on DCT- and DWT-based feature extraction and then KNN, random forest, and decision tree-based classification. The proposed system has been generated the results of accuracy by more than 80%. This shows that the gait-based authentication will be successful for the real-life scenario with further enhanced results.

Keywords DWT · DCT · Decision tree · KNN · Random forest

1 Introduction

In recent times, biometrics is the most usable technology for the person recognition. These biometric-based techniques use different physiological and behavioral traits. These traits are fingerprints, palm prints, face, eyes, voice, etc. Gait-based recognition is new to the field. The major advantage for using gait-based recognition is its ability to recognize the person from distance and with lower resolution [1, 2]. The recognition process will start when human image will be far few pixels to capture for the different stances of the movement. This movement starts with transferring of the weight from left leg to right leg. When person walks, there are 24 different components work [3, 4]. These components can be extracted for the features values.

Shivani (✉) · N. Singh
Computer Science and Engineering, Punjabi University, Patiala, Punjab, India
e-mail: shivausha1975@gmail.com

N. Singh
e-mail: navdeepsony@gmail.com

Measuring few out of 24 will be sufficient for differentiating the persons from each other. This means for the person unique identification for biometrics purpose, gait can be best technique which will be having higher results with less delay [5]. The main advantage of gait over other biometric-based techniques is that it takes less time for recognition and also does not require high-resolution images. It simply checks the instances of the objects and converts the instances to the recognizable objects. There are two approaches for gait recognition. One is the model-based approaches, and second is the model free approaches [6, 7]. In the model-based approaches, the body parts such as joints are used for recognition. These approaches use high-quality images of gait sequence or cycle. It converts the whole human body posture to pre-defined objects. These objects can be ellipse, circle, etc. [8]. Model-free approaches for gait do not considered any of the prior knowledge for interpreting the shape of the object. It simply captures the object with different angle and interpret the object with its physiological and various other features.

2 Literature Survey

Liu et al. [5] has surveyed different ways for the gait recognition. He worked on gait recognition for the video-based gait recognition approach. He defined two ways for the gait recognition. One is the model-based feature extraction, and second is the model-free approaches. Pushpa Rani and Arumugam et al. [9] has proposed a technique for gait-based recognition for biometrics purpose. The video of the persons is subdivided into multiple images. Each image was processed to extract the foreground image and removing the background image. These foreground images are transformed using eigenvector transformation for different feature extraction. These features were later on classified with different classification techniques. The extracted results for the whole gait-based database is in higher band. Liu et al. [10]: has proposed a technique for gait-based recognition by capturing the videos of the persons with multiple views. Later on, Fourier transform technique is applied on the gait energy image. Lower-frequency components adopted for the multiple views of gait-based recognition. The proposed approach is having higher success and also takes less time for evaluation of whole database. Liu et al. [11] has proposed a approach for the gait-based recognition. They have followed a technique with multiple standard steps. In first step, there is transformation of the image into pre-processable features. In second phase, all the features are extracted from the pre-processed image. Those extracted features are transformed into different classes using BP neural network and optimized genetic algorithm. The proposed approach has achieved the accuracy of 96.5%. Kusakunniran et al. [12] has proposed a technique based on two given steps. In the first step, there are various features extracted from the local motion features of the image. These features are filled into the selection vector. Later on, support vector regression (SVR) classification technique is used to classify the features. The proposed approach is for the view-independent gait recognition. Towheed et al. [13] focused on movement of clothes where they examined the both traditional and

western clothes for the person identification. They included 38 individuals, and recognition rate was 88%. Ye et al. [14] proposed a technique in which they used horizontal, vertical, and diagonal directions of the individual identification. The DWT feature extraction method and NN, KNN, and SVM classification methods are used. The experiment was performed on 30 individuals.

3 Feature Extraction and Classification Techniques

The physiological and statistical features extracted using DWT and DCT are classified using different classification techniques for the prediction purpose. Classification techniques classify the features extracted with different attributes. They generate the outcome for the prediction. Compare the prediction of different classification techniques with standard outcome by comparing the success rate.

3.1 *DWT and DCT*

DWT is the discrete wavelet transform used to convert the inputted gait images to interconnected objects and extract the features of each gait image. All the frequency domain is interpreted and extracted for the various features [9]. DCT is the discrete cosine transformation used for the feature extraction from the gait image. It extracts the information for the frequency domain by simply converting the spatial distributed points to sum of the cosine functions oscillating at different frequencies. This is the most simplest approach to convert the signals into measurable features [10].

3.2 *KNN, DT, and RF*

K-nearest neighbor technique is based on supervised machine learning technique. It sorts the distance of each value from the centroid value. K number of values are taken from the top of sorted metrics. This will classify the metrics entries based on distances. Lower distance values of the metrics are put into one class, and others are put into next class [9]. DT is a supervised machine learning technique for classifying the attributes into different classes. Decision tree builds the conditions and represents them as internal nodes. The outcome of the internal nodes is represented as the outer nodes. Each leaf of the tree represents the class for the classification process [9]. RF is the another classification technique based on supervised learning. It classifies the attribute values based on multiple decision trees. It uses the randomness for the features for building the tree. It creates the uncorrelated forest tree, whose prediction percentage is better compared to the regular tree [10].

4 Methodology

The proposed methodology shown in Fig. 1 for the gait recognition is based on two different steps. In first step, features are extracted from the gait-based image. These features are extracted using DWT and DCT. The extracted features are filled into the features vector. In second step, different classifiers like KNN, random forest, and decision tree are applied for the classification of the features.

4.1 Flowchart

Figure 2 shows the DWT- and DCT-based feature extraction for the images taken from the gait database. These features are classified using different classification techniques.

4.2 Algorithm

Gait-based recognition as an alternative for different traditional biometrics techniques is based on physiological and behavioral features. The higher success rate is for gait-based recognition with lower-resolution images.

Fig. 1 Methodology

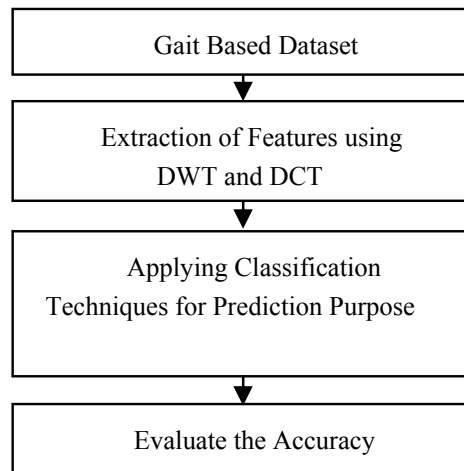
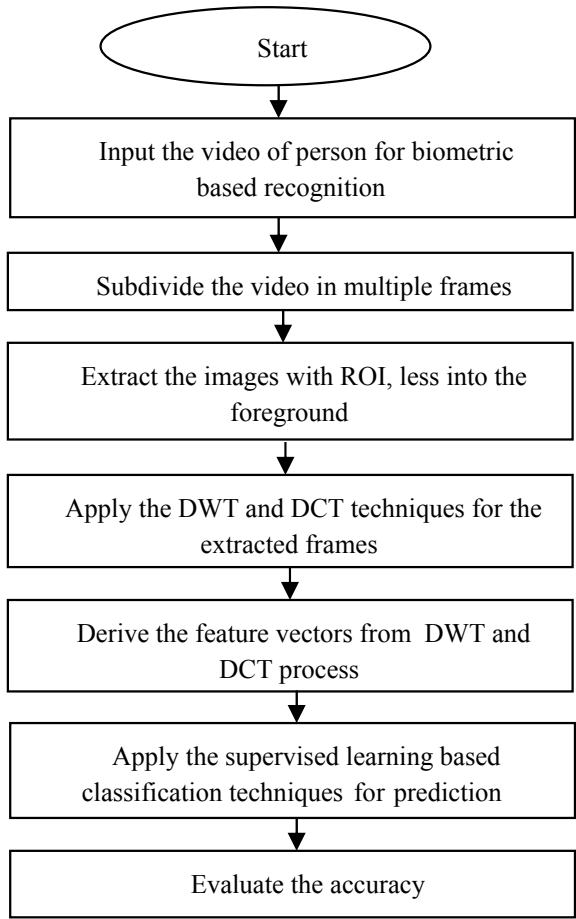


Fig. 2 Flowchart of DWT and DCT feature extraction



The six steps of algorithm are as follows:

- Step 1: Record the videos of the persons moving in forward and backward directions.
- Step 2: Subdivide the video into multiple frames. Each frame will be having single image representing different posture of the person movement.
- Step 3: Pre-process the images to extract the foreground from the image. Background will be removed from the image.
- Step 4: Apply the DCT- and DWT-based techniques for extracting the features vector from the foreground image.
- Step 5: Apply the different supervised classification techniques like KNN, decision tree, and random forest for classifying the features for the prediction of the person.
- Step 6: Evaluate the accuracy for the prediction by comparing the standard outcome

5 Results

Figure 3 shows the comparison for different parameters for DWT with various classification techniques that are KNN, DT, and random forest. The accuracy for the true positive has improved, and both false acceptance rate and false rejection rate have shown reduction. This means for the gait-based biometrics, DWT-based feature extraction and KNN-, DT-, and RF-based classification have shown improvement (Fig. 4).

Table 1 shows the accuracy for the DCT and DWT with KNN, DT, and RF classification techniques for gait-based recognition. This technique shows better accuracy for gait-based recognition authentication. The comparison of existing method accuracy with proposed method accuracy is shown in Table 2. In our proposed technique, we have 99% and 98.9% accuracies, where feature extraction method is DCT

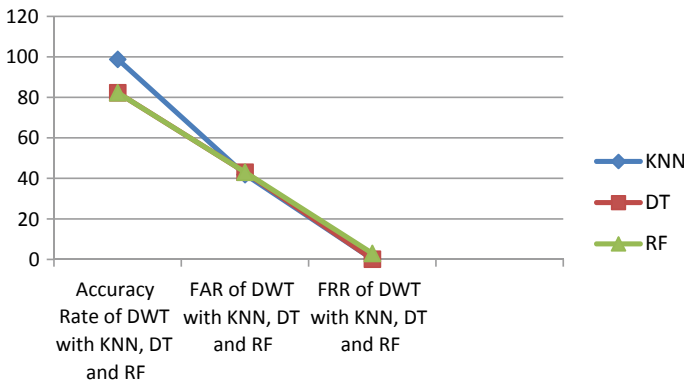


Fig. 3 Comparison of parameters for DWT with classification techniques KNN and DT

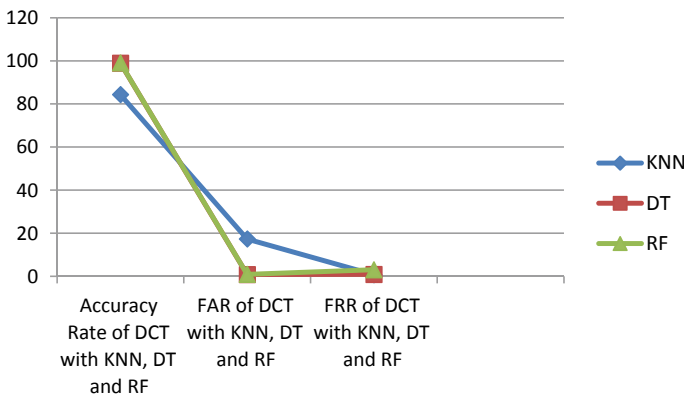


Fig. 4 Comparison of parameters for DCT with classification techniques KNN, DT, and RF

Table 1 Accuracy rate, FAR, FRR of the proposed technique for CASIA-A gait dataset

Feature extraction methods	Evaluation parameter	Classification methods		
		KNN	DT	RF
DWT	Accuracy rate	98.8%	82.32%	82.32%
	FAR	41.91%	43.07%	43.07%
	FRR	0.13%	0.15%	0.15%
DCT	Accuracy rate	84.33%	98.9%	99%
	FAR	17.28%	0.78%	0.89%
	FRR	0.73%	0.86%	0.83%

Table 2 Comparison of existing methods recognition rate with proposed method

Existing methods	Dataset	Accuracy
DCT + LDA (Fan et al. [10])	CASIA	94.4%
DWT + KNN (Rahati et al. [15])	NIST/USF	94%
DCT + PCA (Tahir et al. [16])	CASIA	80%
Proposed method	Dataset	Accuracy
DWT + KNN	CASIA-A	98.80%
DWT + DT	CASIA-A	82.20%
DWT + RF	CASIA-A	82.24%
DCT + KNN	CASIA-A	84.34%
DCT + DT	CASIA-A	98.9%
DCT + RF	CASIA-A	99.00%

and classification methods are RF and DT. And using KNN with DCT, the accuracy is 84.34%. In comparison to existing classification techniques, the proposed classifications methods accuracy are better.

6 Dataset

Gait recognition is the popular pattern recognition problem to be used in various research fields. This gait recognition is used for computer vision, machine learning, biomedical, forensic studying, and robotics. This problem also has great potential in industries such as visual surveillance. Hence, The Intelligent Recognition and Digital Security Group, which was formed in 1998 by Prof. Tieniu Tan at NLPR (National Laboratory of Pattern Recognition), developed this CASIA dataset since December 10, 2001. This CASIA Gait Recognition Dataset currently contains four subsets: Dataset A (standard dataset), Dataset B (multi-view gait dataset), Dataset C (infrared gait dataset), and Dataset D (gait and its corresponding footprint dataset).

Advantages and disadvantages of the dataset:

There are various features of the gait recognition system performed by different researchers with different types of techniques.

Advantages:

- Less time consumption-based authentication: There will be less time taken for the authentication based on the gait recognition. The person whose authentication has to be performed does not have to stop for proving the authentication.
- There is higher success rate of different techniques used for the purpose of recognition.
- There is less overhead with the authentication based on gait-based recognition.

Disadvantages:

- There are various glitches in the techniques when there are unrecognized postures.
- There is higher failure, so still long way to go so that it can be used for real application.

7 Conclusion

In recent times, there is growth in the biometric-based authentication. All the private and public institutions are implementing biometric-based authentication and attendance system. Different researchers are putting up their effort for improving the success rate and also time taken for authentication. This will increase the acceptability of the biometric-based system. In recent times, gait-based biometrics system came into the research circle. This type of biometrics rather than considering physiological or behavioral features considers human's walking postures. The authentication system will take a video for the person while person moves into the campus. This means authentication process starts when person is far away. Gait-based authentication system is having advantage that it can work on the image with poor resolutions but with higher success rate. In proposed system, DWT- and DCT-based techniques are used for feature extraction. The created features vectors in second stage will be having different classifiers like KNN, random forest, and decision tree. The proposed system is showing promising results for the accuracy. Using KNN-based classifiers, the result is almost 99% into the gait-based database.

8 Future Work

The proposed system is for the gait-based biometrics. This type of system is successful for the given image dataset extracted from the video of the person movement. This work can be enhanced further to reduce the false acceptance rate and false rejection rate.

References

1. Liu H, Dai Z, Zheng Y (2019) Research on gait recognition algorithm based on multiinformation perception. *J Phys: Conf Ser*, 1187(4):042058
2. Fanello SR, Gori I, Metta G, Odone F (2013) Keep it simple and sparse: realtime action recognition. *J Mach Learn Res* 14(1):2617–2640
3. Wang D, Tsui K-L (2017) Dynamic bayesian wavelet transform: new methodology for extraction of repetitive transients. *Mech Syst Signal Process* 88:137–144
4. Nogueira SL, Siqueira AA, Inoue RS, Terra MH (2014) Markov jump linear systems-based position estimation for lower limb exoskeletons. *Sensors* 14(1):1835–1849
5. Liu LF, Jia W, Zhu YH (2009) Survey of gait recognition. In: *International conference on intelligent computing*. Springer, pp. 652–659
6. Kusakunniran W, Wu Q, Zhang J, Li H (2010) Support vector regression for multi-view gait recognition based on local motion feature selection. In: *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE 2010, pp. 974–981
7. Yang X, Tian Y (2014) Effective 3d action recognition using eigenjoints. *J Vis Commun Image Represent* 25(1):2–11
8. Chauhan VK, Dahiya K, Sharma A (2019) Problem formulations and solvers in linear svm: a review. *Artif Intell Rev* 52(2):803–855
9. Mostayed A, Mazumder M, Kim S, Park S (2008) Abnormal gait detection using discrete wavelet transform in fourier domain. In: *4th Kuala Lumpur international conference on biomedical engineering 2008*. Springer, pp. 383–386
10. Fan Z, Jiang J, Weng S, He Z, Liu Z (2016) Human gait recognition based on discrete cosine transform and linear discriminant analysis. In: *2016 IEEE international conference on signal processing, communications and computing (ICSPCC)*. IEEE 2016, pp. 1–6
11. Rani MP, Arumugam G (2010) An efficient gait recognition system for human identification using modified ica. *Int J Comput Sci Inf Technol* 2(1):55–67
12. Liu YQ, Wang X (2011) Human gait recognition for multiple views. *Proc Eng* 15, 1832–1836
13. Towheed MA, Kiyani W, Umam M, Shanableh T, Dhou S (2019) Motion-based gait recognition for recognizing people in traditional gulf clothing. In: *2019 IEEE/ACS 16th international conference on computer systems and applications (AICCSA)*. IEEE 2019, pp. 1–6
14. Ye B, Wen YM (2000) Gait recognition based on dwt and svm. In: *2007 international conference on wavelet analysis and pattern recognition, vol 3*. IEEE 2000, pp. 1382–1387
15. Rahati S, Moravejian R, Kazemi FM (2008) Gait recognition using wavelet transform. In: *Fifth international conference on information technology: new generations (itng 2008)*. IEEE 2008, pp. 932–936
16. Yaacob NI, Tahir NM (2012) Feature selection for gait recognition. In: *2012 IEEE symposium on humanities, science and engineering research*. IEEE 2012, pp. 379–383

A Novel Image Captioning Model Based on Morphology and Fisher Vectors



Himanshu Sharma

Abstract Recent image captioning models ignore the text that is present in approximately 50% of images. These texts convey a lot of useful information about the image and its surrounding. The efficiency of an image captioning model can thus be improved by including these texts together with the visual information. In this paper, a novel image captioning model that considers the text existing in an image is proposed. The paper uses the concept of morphology of a word, and thus constructs Fisher Vectors based on the morphology of a word. The proposed model is evaluated on two publically available datasets. The captions produced by the proposed model are comparable to the state-of-the art captioning methods.

Keywords Image captioning · Morphology · Fisher vector

1 Introduction

The applications like image and video processing require the fusion of multimodal features. Human beings do not need detailed explanations for understanding these images and videos. Apart from the visual content contained in an image, textual content may also be available. The datasets like MS-COCO [1] have large number of images containing texts in them. This textual content can be very useful for describing these images. Therefore, it is really essential to propose such models that use this textual information for the better description of these images.

In order to use this textual content together with visual content, a novel captioning model is proposed that uses Pyramidal Histogram of Characters (PHOC) and Fisher Vectors for generating human-like captions. The paper uses model [2] for finding the PHOC of a given scene text. For improving the quality of captions generated, Fisher Vector (FV) [3] are employed which are constructed using the PHOCs computed by model [2].

H. Sharma (✉)

Department of Computer Engineering and Applications, GLA University Mathura, Mathura, India
e-mail: himanshu.sharma@gla.ac.in

It is observed that Fisher Vectors [4–7] can progress the efficiency of various computer vision tasks. By employing fisher vectors in domains like NLP tasks, word embeddings and video action recognition gave improved accuracy over the openly available datasets [5, 8].

In this paper, the morphology of a word is used together with the visual features extracted from an image. The key advantage of the proposed method when compared to previous methods is noticed while handling errors during text recognition and named entities. Every word has a fine-grained explicit class and does not need to recognize them semantically. By fusing both visual and textual content, better image captions are produced as a probability vector is generated by fusing both feature vectors. This work tries to overcome the limitations of [9, 10]. The limitations are related to fixed vocabulary, few words not included in semantic embeddings like word2Vec and GloVe and the error existing in text recognizers. The words having similar morphology are grouped together to construct a descriptor. This descriptor is used to produce fine-grained textual features. In Fig. 1, morphology of similar words is shown.

Overall, the main highlights of the proposed work are:

- The proposed image captioning model utilizes both visual and textual features. The proposed model has the capability to distinguish between the images that almost similar in visual appearance which the previous models fail to distinguish.

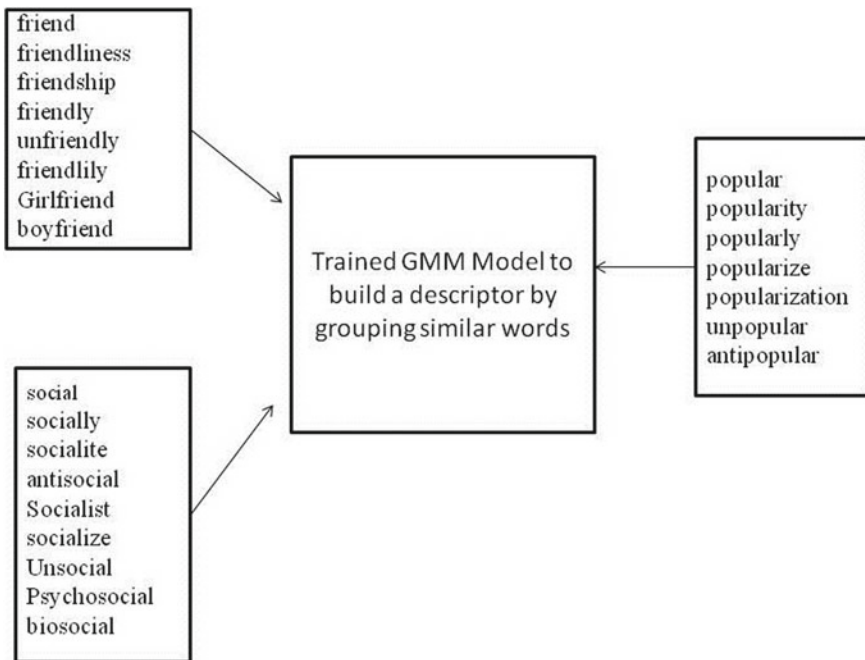


Fig. 1 Words with similar morphology

- The proposed model employs Fisher Vectors generated from scene-text PHOCs and thus creates a descriptor which is both powerful and discriminative in nature.
- The proposed model uses LSTM-Fisher Vector (LSTM-FV) for sentence representation.
- The proposed model is comparable to the previous image captioning works.

The subsequent sections are organized as follows: In Sect. 2, some state-of-the-art captioning methods are presented. In Sect. 3, the proposed model for captioning of images is presented. In Sect. 4, exhaustive experiments on MS-COCO [11] and Flickr30k [12] datasets are conducted. In Sect. 5, some future directions in the task of image captioning are discussed.

2 Related Works

In reference [13], the authors used deep CNN to pull out visual features obtained from its FC-layer and inject them into LSTM-based language model at the first time step. In reference [14], the task of sentence generation is performed using two LSTMs. The last generated word together with image features is given as input to the LSTM unit. Ranking system for generated captions was discussed by reference [15]. They used common sub-space for representing images and sentences when the captions corresponding to an image are produced. Soft and hard attention methods are employed by reference [16] to capture relevant image features in an image. The language structures using neural networks are used by reference [17] to generate image captions. They also applied log bilinear mathematical transforms instead of applying multimodal techniques.

The authors of [18] use both visual and textual information for describing images in natural sentences. Image classification task was performed by employing Fisher Vectors in [19]. RNN-FV was used for the task of action recognition and image annotation by reference [20]. Emotional expressions are used and added to describe images by the authors of [21]. In reference [22], an approach for choosing the best sentences describing an image or video was proposed. In [23], the visual content is fused with entity information and object labels to generate captions. Captions for remote sensing images were proposed by Shi et al. [24]. They employed deep learning and fully convolutional networks to generate captions. Reference [25] used both spatial and semantic attention to find the image regions that match best to a given word. In [26], part-of-speech tagging method is used to describe the images. The authors of [27] applied both attention modules i.e. top-down and bottom-up for the task image captioning and visual question answering (VQA). In [28], supplementary significant information is produced from retrieval or semantic guidance and gives it to the LSTM unit for generating captions. An adaptive novel attention model was proposed by reference [29] which guides whether to capture certain image feature or not. Reference [30] combines features of an image with significant image attributes. Further, they are provided to the hidden state of LSTM for generating the image

captions. CNN-based decoder is used by reference [31] to generate captions. Reference [32] used external knowledge for achieving human-like captions for a given image together with the visual features.

3 Proposed Model

The proposed image captioning model is shown in Fig. 2. Both visual and textual contents are combined together in order to generate more detailed captions. Also, the proposed model uses attention method to remove irrelevant words from the text present in image. The model uses the PHOC of each word present in the scene-text and uses pre-trained Gaussian Mixture Model (GMM) [33] in order to generate a Fisher Vector descriptor. Both visual and textual representations are combined together using LSTM-FV to produce image captions.

3.1 Encoding Visual Features

The model uses a VGGNet [34] to extract the image features from the given input image (img). This model [34] is pre-trained on ImageNet [8]. The proposed model

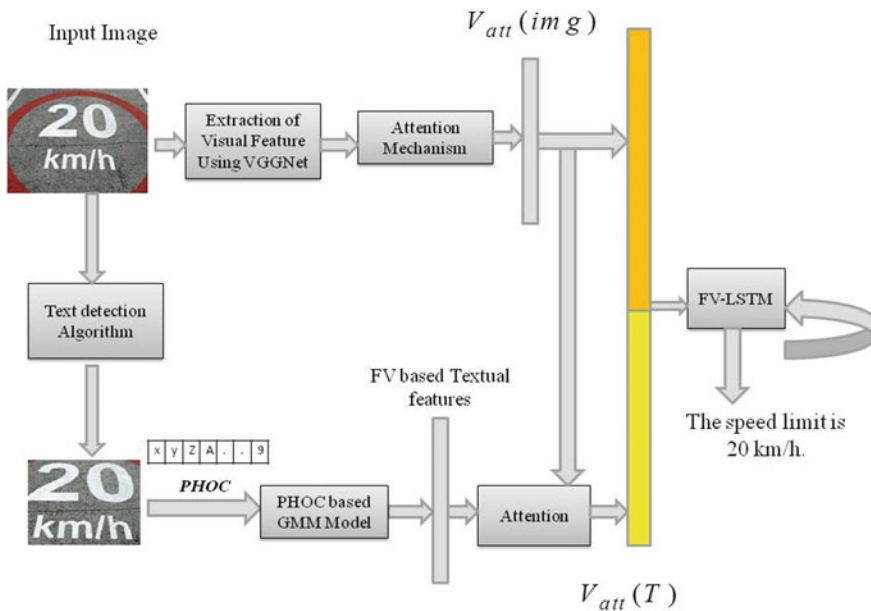


Fig. 2 Block diagram of the proposed model

makes use of the output of the last level/layer of VGGNet for representing the image features represented by $V(\text{img})$. Also, soft attention (att) is applied on $V(\text{img})$. So, the attended visual features computed by the proposed model are given by:

$$V_{\text{att}}(\text{img}) = V(\text{img}) + (V(\text{img}) \times \text{att}) \tag{1}$$

3.2 Encoding Textual Features

The presented model employs GMM and then extracts Fisher Vector-based textual feature representations using GMM model. A TextBoxes [35] algorithm is used to extract the text present in an image. For obtaining differentiating features, the model uses PHOC. In Fig. 3, the PHOC of word “tokens” is obtained. Then, GMM model is employed to obtain the Fisher Vector from these PHOCs of words. The final vector representation of textual features extracted from an image is of size 1×512 dimensions.

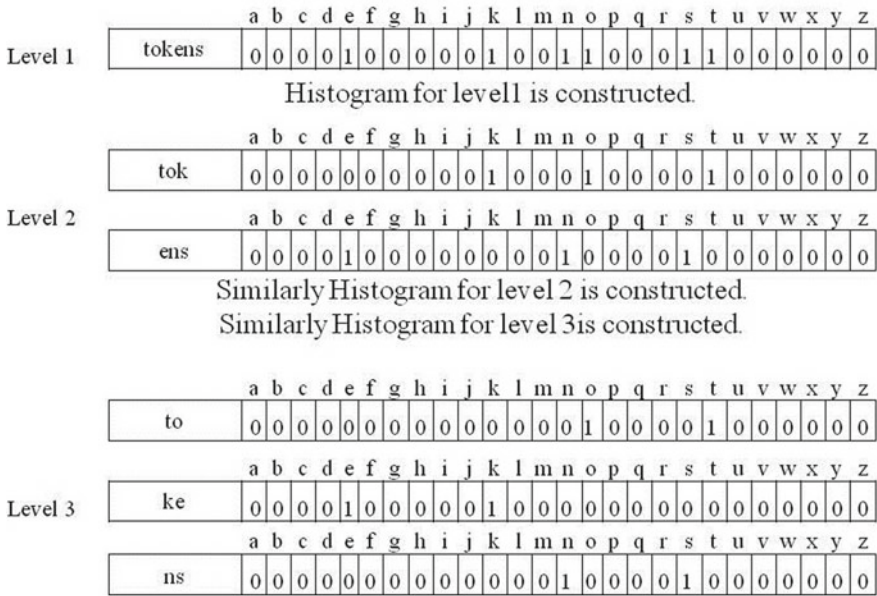


Fig. 3 Pyramidal histogram of characters (PHOC) corresponding to “tokens”

3.3 Applying Attention

The proposed model uses an attention mechanism to extract the most relevant text based on the visual features. By using attention units, the model calculates the tensor of weights denoted by $T_a(w)$. These weights are further used for constructing Fisher Vectors. The model finally computes the textual features as follows:

$$T_a(w) = \text{Soft max}(\tanh(V_{\text{att}}(\text{img})^T \cdot T_w \cdot T_F)) \quad (2)$$

$$V_{\text{att}}(T) = T_a(w) \quad (3)$$

Now, the proposed model concatenated the attended visual features $V_{\text{att}}(\text{img})$ and attended textual features $V_{\text{att}}(T)$ to get a joint feature vector representation F_{comb} given by:

$$F_{\text{comb}} = [V_{\text{att}}(\text{img}), V_{\text{att}}(T)] \quad (4)$$

The obtained vector F_{comb} will be given as input to the FV-LSTM unit to generate the captions.

3.4 Caption Generation Using LSTM

Finally, the obtained feature vector which is the combination of both visual and textual features is given to the FV-LSTM [39] unit. Also, the model employs Fisher Vector based on [40] to produce the captions. On providing a set of words say, $S = (s_1, s_2, \dots, s_{K-1})$ from a sentence having K words $s_1, s_2, s_3, \dots, s_K$, to the FV-LSTM, it generates the next word of the sequence $S = (s_1, s_2, \dots, s_K)$. The captions generated by the proposed model are more close to the captions generated by humans.

4 Discussions

The model is evaluated both quantitatively and qualitatively on two popular datasets: Microsoft COCO [11] and Flickr30k [12]. The proposed model is compared with the popular image captioning models. The details of datasets, implementation, and results are discussed in subsequent sections.

4.1 Datasets Details

Microsoft COCO [11] and Flickr30k [12] are the two datasets which are considered in this paper. MS-COCO dataset has 82,780, 40,500, and 40,770 images for training, validation, and testing tasks, respectively. Each image is labeled with five captions in [11]. Flickr30k [12] includes 31,780 images from Flickr. This dataset contains 158,915 captions for the images available in dataset. In [12], there are five images corresponding to each image.

4.2 Experimental Details

VGGNet is used to obtain visual features from the last Conv-layer. The obtained features are further given as input to the FC-layer of the VGGNet to get the O/P vector of size 1×1024 . The textual features are obtained by applying 15 PHOC proposal for a given input image. Zero padding is applied to finalize the dimension of input features. Further, PCA is applied to lower the dimension of computed PHOCs. Finally, it is converted to a dimension of Tx300.

Adam optimizer is applied in the proposed model with rate of learning set to $5e^{-4}$ by setting batch size to 64 and 40 epochs for model training.

The proposed model employed a pre-trained GMM to attain a Fisher Vector from the PHOCs. The model computes the dimension of FV as $1 \times 38,400$ by applying 64-GMM components. A pre-trained GMM is applied to calculate the Fisher Vector from the above obtained PHOCs. The dimension of the FV is $1 \times 38,400$ by using 64 Gaussian components by trained GMM. The extracted textual features are further input to the FC-layer and thus converted to the output vector of size 1×512 dimensions. After the attention mechanism is performed on both the modalities, the output vector is produced of dimension 1×512 that multiplied the weights to be learned with the scene-text features.

4.3 Metrics for Evaluation

The proposed image captioning model is tested on three metrics which are BLEU [36], Meteor [37], and CIDEr [38]. The main purpose of using these three metrics is to check the correspondence between reference captions and generated captions.

Tests are conducted on both parts of the datasets. The first portion includes only visual features and the second part considers both image and textual information. As demonstrated in Tables 1 and 2, the proposed image captioning model is compared with some of previous captioning methods on the publically available datasets [11] and [12]. The proposed model gives promising results when compared with the state-of-the-art models as indicated in Tables 1 and 2. The reason behind the gain in

Table 1 Evaluation of various models on MS-COCO dataset

Model	B@1	B@2	B@3	B@4	METEOR	CIDEr
Reference model [13]	66.6	45.1	30.4	20.3	—	—
Reference model [14]	69.7	51.9	38	27.8	22.9	83.7
Reference model [15]	62.5	45	32.1	23	19.5	66
Reference model [16] (soft attention)	71.8	50.4	35.7	25	23	—
Reference model [16] (hard attention)	70.7	49.2	34.4	24.3	23.9	—
Reference model [28]	67	49.1	35.8	26.4	22.7	—
Reference model [29]	74.2	58	43.9	33.2	26.6	108.5
Reference model [30]	70.9	53.7	40.2	30.4	24.3	—
Reference model [31]	71.1	53.8	39.4	28.7	24.4	91.2
Proposed model	67.4	53.8	43.3	30.6	23.9	92.5

Table 2 Evaluation of various models on Flickr30k dataset

Model	B@1	B@2	B@3	4	METEOR	CIDEr
Reference model [13]	66.3	42.3	27.7	18.3	—	—
Reference model [14]	57.3	36.9	24	15.7	15.3	24.7
Reference model [15]	66.7	43.4	28.8	19.1	18.5	—
Reference model [16] (soft attention)	66.9	43.9	29.6	19.9	18.5	—
Reference model [16] (hard attention)	60	38	25.4	17.1	16.88	—
Reference model [28]	60	41	28	19	—	—
Reference model [29]	67.7	49.4	35.4	25.1	20.4	53.1
Reference model [30]	64.7	46	32.4	23	18.9	—
Reference model [31]	63.6	45.4	27.3	19.3	19.2	33.5

performance of the proposed model is the fusion of image features with the textual feature. In [40], the results are given for state-of-the-art captioning models. In Figs. 4 and 5, the qualitative results are presented for the model on both the datasets [11] and [12].

5 Conclusion

In this paper, a novel image captioning model is proposed that exploits the textual content jointly with the visual content for generating more comprehensive captions for a given natural scene. To achieve these detailed captions, the proposed model uses morphology of a word and the construct Fisher Vectors based on the morphology of each word present in an image. The results generated by the proposed model are comparable to recent captioning models on publically available datasets such as



Previous Model: A shop in a street

Proposed Model: A man's barbershop in a street.



Previous Model: A bus is standing on a street.

Proposed Model: A purple route bus having 1208 number is standing on a street.

Fig. 4 Results generated by the proposed model versus previous model on Flickr30k dataset



Previous Model: A soft drink can on the table.

Proposed Model: A coca cola can on the table.



Previous Model: Lot of people in front of a building.

Proposed Model: Lot of people standing in front of Adobe company building.

Fig. 5 Results generated by the proposed model versus previous model on MS-COCO dataset

MS-COCO and Flickr30k. In future work, the researchers can include the emotional concepts while generating captions and also develop models that can generate captions related to a given topic.

References

1. Veit A, Matera T, Neumann L, Matas J, Belongie S (2016) Coco-text: dataset and benchmark for text detection and recognition in natural images. [arXiv:1601.07140](https://arxiv.org/abs/1601.07140)
2. Almazán J, Gordo A, Fornés A, Valveny E (2014) Word spotting and recognition with embedded attributes. *IEEE Trans Pattern Anal Mach Intell* 36(12):2552–2566

3. Perronnin F, Dance C (2007) Fisher kernels on visual vocabularies for image categorization. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE, pp 1–8
4. Simonyan K, Parkhi OM, Vedaldi A, Zisserman A (2013) Fisher vector faces in the wild. In: *BMVC* 2(3):4
5. Peng X, Zou C, Qiao Y, Peng Q (2014) Action recognition with stacked fisher vectors. In: *European conference on computer vision*. Springer, Cham, pp 581–595
6. Chatfield K, Lempitsky VS, Vedaldi A, Zisserman A (2011) The devil is in the details: an evaluation of recent feature encoding methods. *BMVC* 2(4):8
7. Perronnin F, Liu Y, Sánchez J, Poirier H (2010) Large-scale image retrieval with compressed fisher vectors. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, pp 3384–3391
8. Wang L, Qiao Y, Tang X (2015) Action recognition with trajectory-pooled deep-convolutional descriptors. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4305–4314
9. Karaoglu S, Tao R, Gevers T, Smeulders AW (2016) Words matter: scene text for image classification and retrieval. *IEEE Trans Multimed* 19(5):1063–1076
10. Bai X, Yang M, Lyu P, Xu Y, Luo J (2018) Integrating scene text and visual appearance for fine-grained image classification. *IEEE Access* 6:6322–6633
11. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Zitnick CL (2014) Microsoft coco: common objects in context. In: *European conference on computer vision*. Springer, Cham, pp 740–755
12. Young P, Lai A, Hodosh M, Hockenmaier J (2014) From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Trans Assoc Comput Linguist* 2:67–78
13. Vinyals O, Toshev A, Bengio S, Erhan D (2016) Show and tell: lessons learned from the 2015 mscoco image captioning challenge. *IEEE Trans Pattern Anal Mach Intell* 39(4):652–663
14. Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T (2015) Long-term recurrent convolutional networks for visual recognition and description. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2625–2634
15. Karpathy A, Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3128–3137
16. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Bengio Y (2015) Show, attend and tell: neural image caption generation with visual attention. In: *International conference on machine learning*, pp 2048–2057
17. Kiros R, Salakhudinov R, Zemel RS (2014) Unifying visual-semantic embeddings with multimodal neural language models. [arXiv:1411.2539](https://arxiv.org/abs/1411.2539)
18. Gupta N, Jalal AS (2019) Integration of textual cues for fine-grained image captioning using deep CNN and LSTM. *Neural Comput Appl* 1–10
19. Maffla A, Dey S, Biten AF, Gomez L, Karatzas D (2020) Fine-grained image classification and retrieval by combining visual and locally pooled textual features. In: *The IEEE winter conference on applications of computer vision*, pp 2950–2959
20. Lev G, Sadeh G, Klein B, Wolf L (2016) Rnn fisher vectors for action recognition and image annotation. In: *European conference on computer vision*. Springer, Cham, pp 833–850
21. Yang J, Sun Y, Liang J, Ren B, Lai SH (2019) Image captioning by incorporating affective concepts learned from both visual and textual components. *Neurocomputing* 328:56–68
22. Dong J, Li X, Snoek CG (2018) Predicting visual features from text for image and video caption retrieval. *IEEE Trans Multimed* 20(12):3377–3388
23. Zhao S, Sharma P, Levinboim T, Soricut R (2019) Informative image captioning with external sources of information. [arXiv:1906.08876](https://arxiv.org/abs/1906.08876)
24. Shi Z, Zou Z (2017) Can a machine generate humanlike language descriptions for a remote sensing image? *IEEE Trans Geosci Remote Sens* 55(6):3623–3634
25. Huang F, Zhang X, Zhao Z, Li Z (2018) Bi-directional spatial-semantic attention networks for image-text matching. *IEEE Trans Image Process* 28(4):2008–2020

26. He X, Shi B, Bai X, Xia GS, Zhang Z, Dong W (2019) Image caption generation with part of speech guidance. *Pattern Recogn Lett* 119:229–237
27. Anderson P, He X, Buehler C, Teney D, Johnson M, Gould S, Zhang L (2018) Bottom-up and top-down attention for image captioning and visual question answering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 6077–6086
28. Jia X, Gavves E, Fernando B, Tuytelaars T (2015) Guiding the long-short term memory model for image caption generation. In: *Proceedings of the IEEE international conference on computer vision*, pp 2407–2415
29. Lu J, Xiong C, Parikh D, Socher R (2017) Knowing when to look: adaptive attention via a visual sentinel for image captioning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 375–383
30. You Q, Jin H, Wang Z, Fang C, Luo J (2016) Image captioning with semantic attention. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4651–4659
31. Aneja J, Deshpande A, Schwing AG (2018) Convolutional image captioning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 5561–5570
32. Sharma H, Jalal AS (2020) Incorporating external knowledge for image captioning using CNN and LSTM. *Modern Phys Lett B* 2050315
33. Gregor J (1969) An algorithm for the decomposition of a distribution into Gaussian components. *Biometrics* 79–93
34. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
35. Liao M, Shi B, Bai X, Wang X, Liu W (2017) TextBoxes: a fast text detector with a single deep neural network AAAI
36. Papineni K, Roukos S, Ward T, Zhu WJ (2002) BLEU: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th annual meeting of the association for computational linguistics*, pp 311–318
37. Denkowski M, Lavie A (2014) Meteor universal: Language specific translation evaluation for any target language. In: *Proceedings of the ninth workshop on statistical machine translation*, pp 376–380
38. Vedantam R, Lawrence Zitnick C, Parikh D (2015) Cider: consensus-based image description evaluation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4566–4575
39. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
40. Sharma H, Agrahari M, Singh SK, Firoj M, Mishra RK (2020) Image captioning: a comprehensive survey. In: *2020 international conference on power electronics & IoT applications in renewable energy and its control (PARC)*. IEEE, pp 325–328

DWT-Based Hand Movement Identification of EMG Signals Using SVM



Vivek Ahlawat, Yogendra Narayan, and Divesh Kumar 

Abstract Electromyogram (EMG) signals are widely used in rehabilitation, medical, engineering, robotic, and industrial fields. For amputee's residual muscles control, EMG signals have been utilized. Multiple-level approximation and detail coefficients of EMG signals were studied here. In this study, EMG signals were extracted from 20 healthy human subjects by multilevel coefficients of wavelet. After the EMG data acquisition, preprocessing was done followed by discrete wavelet transform (DWT) de-noising and feature extraction by db4 wavelet. Three-level decomposition was achieved with DWT technique, and higher classification accuracy was observed by support vector machine (SVM) classifier with principal component analysis. This paper proposed a combination of a highly accurate algorithm for the pattern recognition of EMG signal with different movements. The finding of this paper could be utilized with real-time applications.

Keywords DWT · SVM · ANN · TFD feature · EMG signal

1 Introduction

In the biomedical various systems like exoskeleton robot, wheelchair robot, and also in other medical applications, myoelectric control technique is used [1]. Various studies have been done by which EMG signal acquired from residual muscle, used for the control purpose [2]. The muscle spasm effect for a long time in recording of EMG signal has been demonstrated [3]. Different hand movements are investigated by using

V. Ahlawat

Department of EE, Sri Ram Group of Colleges, Muzaffarnagar, Uttar Pradesh, India
e-mail: vk.ahlawat@gmail.com

Y. Narayan

Department of ECE, Chandigarh University, Chandigarh, Punjab, India
e-mail: Narayan.yogendra1986@gmail.com

D. Kumar (✉)

Department of ECE, GLA University, Mathura, Uttar Pradesh, India
e-mail: kambojdivesh@gmail.com

various pattern recognition techniques [4]. In previous studies, several channels are used to acquire the signal and classification [5, 6]. The complexity of classification for the real-time performance of the system can be by using more number of channels [7]. This study intends to investigate the EMG-based identification of hand movements for the myoelectric prosthetic hand [8]. In this paper, we proposed a two-channel DWT-based EMG classification by using SVM.

EMG can be characterization as the study of muscle functions utilizing electrical signals generated by muscular contractions which may be voluntary or involuntary [9, 10]. Voluntary muscle contractions of EMG activity belong to tension [11]. The functional unit of the muscle contraction is made up of fibers and a single alpha motor neuron called motor unit [12]. If the action potential of the nerve reaches a depolarization threshold, then contraction of muscle fiber occurs, resulting in the generation of an electromagnetic field whose potential is measured in terms of few voltages [13]. Depolarization is nothing but the muscle action potential [14]. EMG signals are based on motor unit action potentials (MUAP) in which motor units are the building blocks of any neuro-muscular system [15, 16]. Analysis of motor units can be used for the study of a neuro-muscular system for diagnosis of muscles diseases classification [17]. The SEMG signals are chosen precisely because of their low amplitude and low signal-to-noise Ratio (SNR) and hence, a challenging task in the clinical research areas and industrial applications [18].

The EMG has been established as an efficient human motion tracking system and has been extensively exploited in the man-machine interface since a few decades [19]. EMG exhibits the neural activity of muscle which produces the electrical signal for elucidating the motion performed by the human subject [20]. The major distention of myoelectric control system is the low invasiveness during signal acquisition in comparison to other bio-signal-based control systems, such as EOG [21].

This paper is divided into five sections: Sect. 2 describes the data acquisition method. Section 3 discusses the feature extraction by DWT with a brief introduction of DWT and SVM. Section 4 shows the experimental results. The conclusion is given in Sect. 5.

2 Materials and Methods

2.1 Data Acquisition Method

The EMG data were acquired by hand movements (flexion and extension) with the age group of 18–24 years of 20 subjects [22]. Two channels are used to acquire the data by using the non-invasive electrodes. The placement of electrodes is accordingly provided by the setup manual [16]. Surface electrodes are widely utilized in kinesiological studies due to its non-invasive character [23]. They are easy to handle but they only detect the surface muscles. To detect the deeper muscles, needle or fine-wire electrodes are used [24]. The preprocessing on signals and data acquisition

is done by Myo-Trace 400 setup. The original acquired EMG signal is shown in Fig. 1.

The EMG data were acquired by hand movements (flexion and extension) with the age group of 18–24 years of 20 subjects [22]. Two channels are used to acquire the data by using the non-invasive electrodes. The placement of electrodes is accordingly provided by the setup manual [16]. Surface electrodes are widely utilized in kinesiological studies due to its non-invasive character [23]. They are easy to handle but they only detect the surface muscles. To detect the deeper muscles, needle or fine-wire electrodes are used [24]. The preprocessing on signals and data acquisition is done Myo-Trace 400 setup. The original acquired EMG signal is shown in Fig. 1.

The process has various steps up to classification. The process flow is shown in Fig. 2.

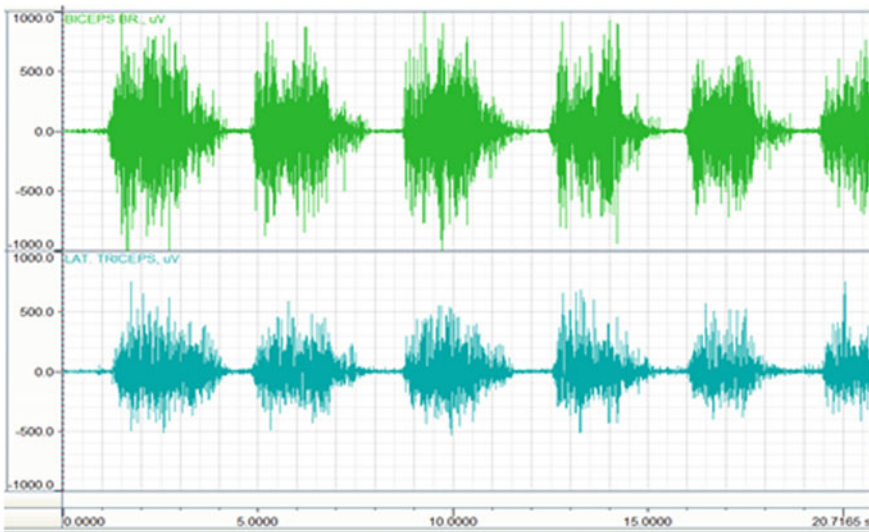


Fig. 1 Original EMG signal acquired with the help of Myo-Trace 400 device

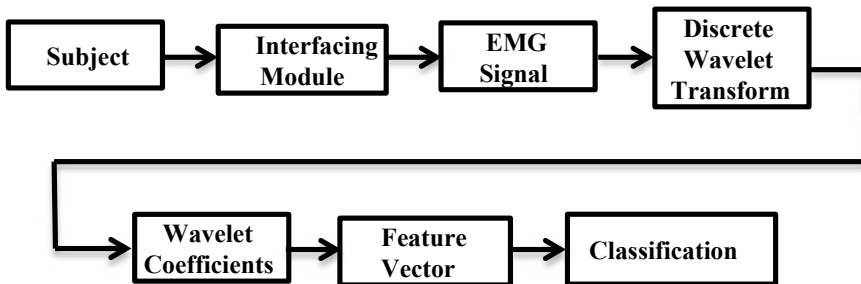


Fig. 2 Flow chart of the EMG signal classification process

3 EMG Feature Extraction

DWT technique was used to decompose the signal. Detail, as well as the approximation coefficients, was utilized for extracting the various features from three levels decomposition [25]. The low-frequency component is termed as approximation coefficients, whereas details consist of high-frequency components as shown in Fig. 3 and Fig. 4. The coefficient values indicate the fitting extent of the signals. The classification accuracy can be enhanced by appropriate feature selection [26].

Three features namely: maximum absolute value (MAV1), root means square (RMS), and mean absolute value (MAV2) were selected for feature vector for the classifier. Third-level detailed and approximation coefficient combination of all three levels was known as AD3 whereas all three levels approximation coefficients are collectively denoted as A3.

3.1 Discrete Wavelet Transform (DWT)

DWT is an emerging technique for feature extraction. Several other wavelet transforms, i.e., wavelet packet transform can be also used for EMG signals [27]. The wavelet filter bank conducts a series of high pass filter and low pass filter connected sequentially. A multi-resolution analysis-based algorithm has been proposed by

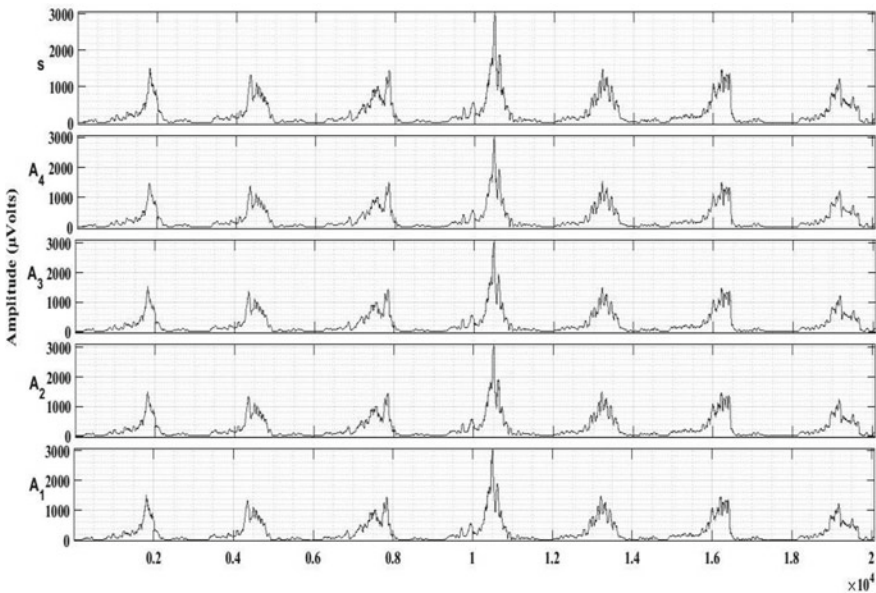


Fig. 3 Signal and approximation(s) coefficient

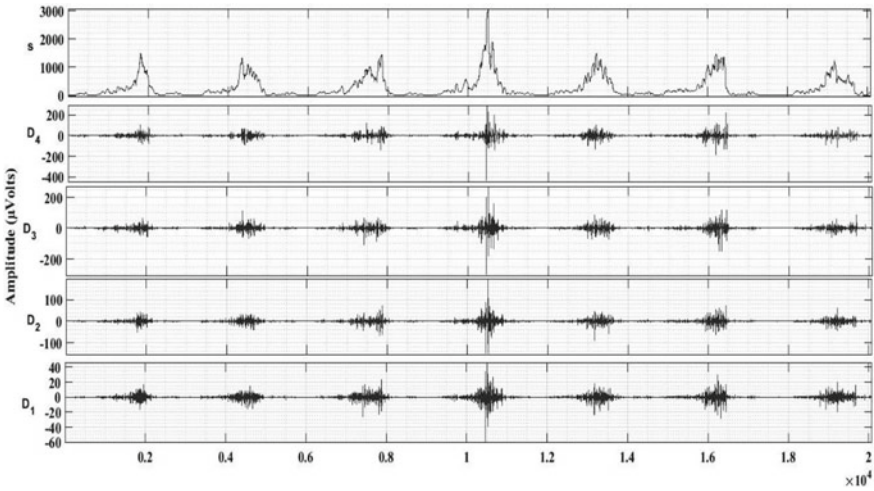


Fig. 4 Signal and detailed(s) coefficient

Mallat [28]. It is representing by Fig. 5. In modern EMG-PR, different time–frequency analysis methods like FFT, DWT, and EMD have been suggested for studying the time-varying properties of SEMG signals [29]. Table 1 gives some wavelet family.

Fig. 5 Mallat algorithm for EMG signal decomposition using wavelet

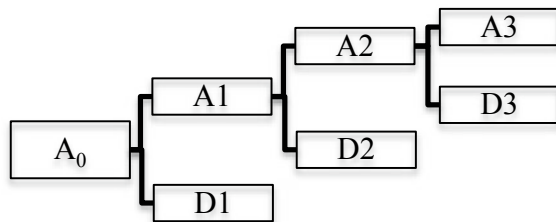


Table 1 Mother wavelet family

S. No.	Wavelet family	Function in MATLAB
1	Daubechies	Db
2	Haar	Haar
3	Symlet	Sym
4	Coiflet	Coif

3.2 Classifiers (SVM)

Support vector machine is a classification method used in machine learning to solve the pattern recognition problems [30]. In this study, the SVM is used for movement classification of sEMG data acquired from different subjects. SVM builds an optimal separating hyperplane based on kernel approach that maximizes the possible margin between points of N different classes as shown in Fig. 6. Nonlinear kernel function like a linear, sigmoid, radial basis function, and polynomial is used for data mapping [31]. It is based on the statistical learning approach and has been widely used in pattern recognition [32]. In a classification problem, a hyperplane distinct the classes of pattern based on the input space sample vector [33]. The vectors near to the hyperplane are known as support vectors.

4 Experimental Results

The main processing steps were as given: (1) EMG signal acquisition; (2) useful feature extraction, followed by (3) classification process. The de-noising process is performed on the EMG signal, for decomposing the signal the mother wavelet namely Coieffet 5 was employed. These three features namely: (1) MAV1; (2) MAV2; (3) RMS, were extracted from all three different levels of detail as well as the approximation coefficients. The feature vectors are formed A3 and AD3. Various fold cross-validations and holdout validations are used for testing the accuracy of classifiers.

Fig. 6 Support vector machine classifier

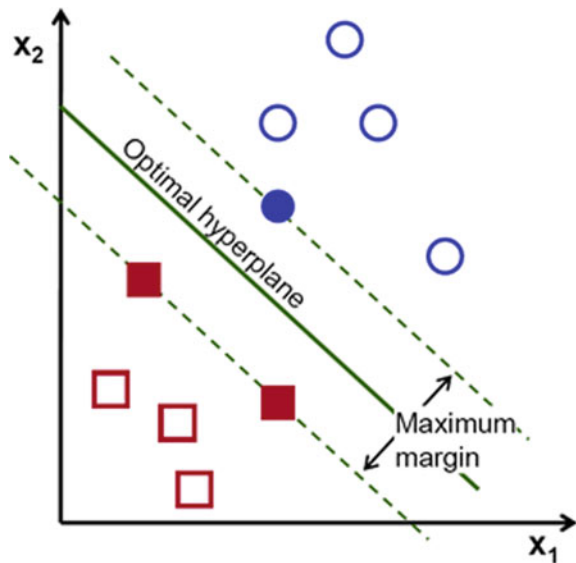


Table 2 SVM classifier accuracy with MAV1

Feature vector	Linear	Quadratic	Cubic	Fine Gaussian
A3	88.14	90.60	96.10	90.72
AD3	90.25	88.50	90.00	90.84
A3 + PCA	87.96	96.34	97.18	94.10
AD3 + PCA	90.42	87.78	88.98	88.68

Table 3 SVM classifier accuracy by using MAV

Feature vector	Linear	Quadratic	Cubic	Fine Gaussian
A3	91.92	92.16	93.54	92.52
AD3	96.1	97.42	98.14	96.28
A3 + PCA	92.1	97	97.72	95.7
AD3 + PCA	98.44	96.16	95.98	97.96

4.1 Maximum Absolute Value (MAV1)

Table 2 depicts that the higher classification accuracy was observed by cubic SVM classifier. The higher accuracy was achieved by applying PCA at the different fold cross-validations.

4.2 Mean Absolute Value (MAV2)

The results show that linear SVM classifier exhibited higher classification accuracy with MAV2 feature extracted from all three levels. By employing the cubic SVM at different fold cross-validations, 98.44% classification accuracy was achieved whereas 98.34% with the different holdout validations as depicted in Table 3.

4.3 Root Mean Square (RMS)

The linear SVM classifier was analyzed with higher classification accuracy rate when the RMS feature was extracted. PCA was applying with ADDD feature set for higher accuracy. Table 4 depicted that higher accuracy of 99.02% is achieved.

Table 4 SVM classifier accuracy by using RMS

Feature vector	Linear	Quadratic	Cubic	Fine Gaussian
A3	93.33	93	92.86	93.05
AD3	96.75	98.61	98.91	96.14
A3 + PCA	93.10	95.32	97	95.03
AD3 + PCA	99.02	97.52	96.33	97.73

4.4 Evaluation by ROC

The ROC curve for the flexion and extension movement is shown in Figs. 7 and 8. The curve shows the relation between the false positive rate and true positive rate. The total area under the curve is 0.98. It is the given by relationship between the FPR of SVM classifier and TPR of the same classifier. In the ROC of extension movement, the positive class is taken as an extension whereas negative class is taken as flexion.

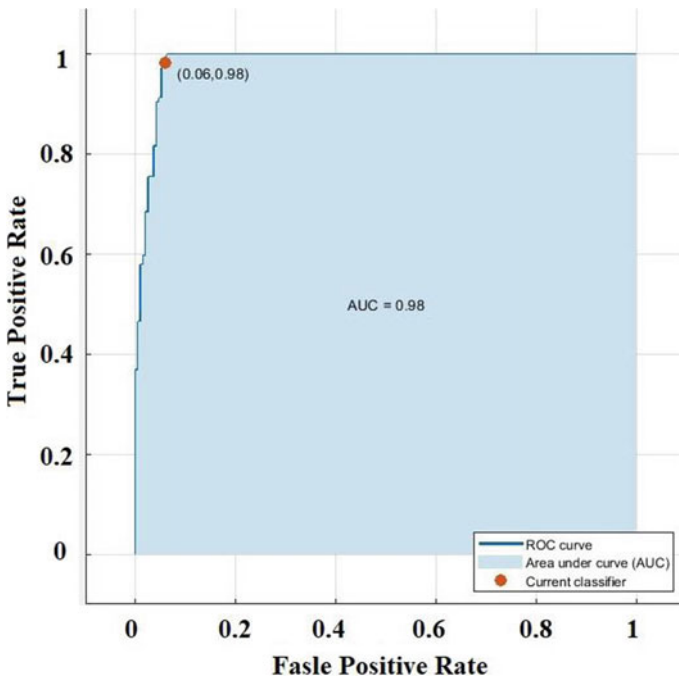


Fig. 7 ROC curve for extension movements

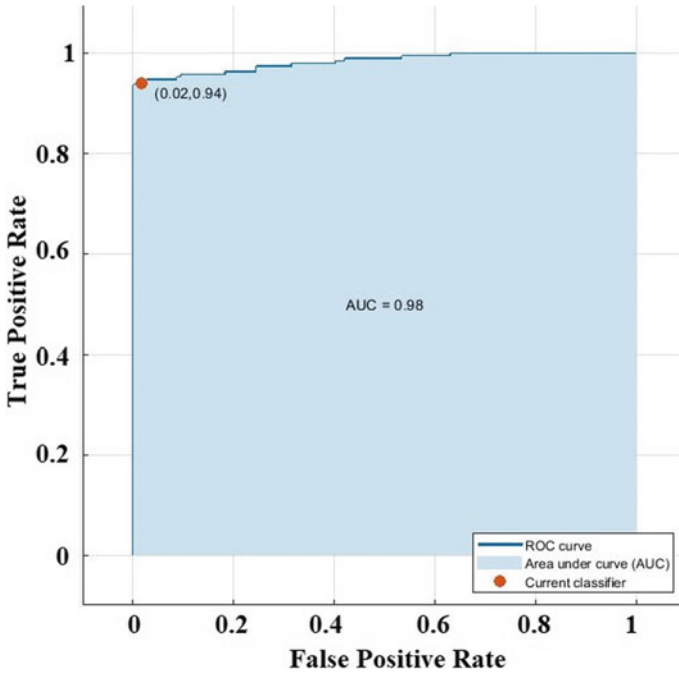


Fig. 8 ROC curve for flexion movements

5 Conclusion

In this study, 20 healthy human subjects participated in two motion classes EMG data acquisition. After the EMG acquisition, it preprocessed and three different features were extracted to the feature vector. Before the EMG feature extraction, DWT denoising and three-level EMG signal decomposition were performed. All three levels approximation and details coefficient were employed for feature extraction process. This study discussed DWT-based feature extraction and classification using two-channel data acquisition. The decomposition of the EMG signal is done up to the third level. Features were computed in the time–frequency domain and feature vector is tested on SVM and ANN classifiers. The result showed that SVM has exhibited excellent performance in terms of speed and accuracy as compared to ANN. The SVM classifier achieved 99.03% classification accuracy that could be considered for real-time application development.

References

1. Côté-Allard U, Campbell E, Phinyomark A, Laviolette F, Gosselin B, Scheme E (2020) Interpreting deep learning features for myoelectric control: a comparison with handcrafted features. *Front Bioeng Biotechnol* 8:1–30
2. Toledo-Perez DC, Rodriguez-Resendiz J, Gomez-Loenzo RA (2020) A study of computing zero crossing methods and an improved proposal for EMG signals. *IEEE Access* 8:8783–8790
3. Rabin N, Kahlon M, Malayev S, Ratnovsky A (2020) Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques. *Expert Syst Appl* 149(7):113281
4. Tuncer T, Dogan S, Subasi A (2020) Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition. *Biomed Signal Process Control* 58:101872
5. Arteaga MV, Castiblanco JC, Mondragon IF, Colorado JD, Alvarado-Rojas C (2020) EMG-driven hand model based on the classification of individual finger movements. *Biomed Signal Process Control* 58:101834
6. Mukhopadhyay AK, Samui S (2020) An experimental study on upper limb position invariant EMG signal classification based on deep neural network. *Biomed Signal Process Control* 55(5):1–8
7. Narayan Y, Ahlawat V, Kumar S (2020) Pattern recognition of sEMG signals using DWT based feature and SVM Classifier. *Int J Adv Sci Technol* 29(10):2243–2256
8. Narayan Y, Kumar D, Kumar S (2020) Comparative analysis of sEMG signal classification using different K-NN algorithms. *Int J Adv Sci Technol* 29(10):2257–2266
9. Ahlawat V, Thakur R, Narayan Y (2018) Support vector machine based classification improvement for EMG signals using principle component analysis. *J Eng Appl Sci* 13(8):6341–6345
10. Narayan Y, Mathew L, Chatterji S (2017) sEMG signal classification using discrete wavelet transform and decision tree classifier. *Int J Control Theory Appl* 10(6):511–517
11. Kumari P, Narayan Y, Ahlawat V, Mathew L (2017) Advance approach towards elbow movement classification using discrete wavelet transform and quadratic support vector machine. In: *Communication and Computing Systems* © 2017 Taylor & Francis Group, London, Advance, pp 839–844. ISBN 978-1-138-02952-1
12. Zhang L, Shi Y, Wang W, Chu Y, Yuan X (2019) Real-time and user-independent feature classification of forearm using EMG signals. *J Soc Inf Disp* 27(2):101–107
13. Turlapaty AC, Gokaraju B (2019) Feature analysis for classification of physical actions using surface EMG data. *IEEE Sens J* 19(24):12196–12204
14. Simão M, Neto P, Gibaru O (2019) EMG-based online classification of gestures with recurrent neural networks. *Pattern Recogn Lett* 128(12):45–51
15. Narayan Y, Singh RM, Mathew L, Chatterji S (2019) Surface EMG signal classification using ensemble algorithm, PCA and DWT for robot control. In: *International conference on advanced informatics for computing research*, vol 10. Springer, Singapore, pp 424–434
16. Virdi P, Narayan Y, Kumari P, Mathew L (2017) Discrete wavelet packet based elbow movement classification using fine Gaussian SVM. In: *1st IEEE international conference on power electronics, intelligent control and energy systems, ICPEICES 2016*, pp 1–5
17. Narayan Y, Kumari P, Garima, Mathew L (2017) Elbow movement classification of a robotic arm using wavelet packet and cubic SVM. In: *Communication and computing systems*, pp 605–610
18. Babita PK, Narayan Y, Mathew L (2017) Binary movement classification of sEMG signal using linear SVM and wavelet packet transform. In: *1st IEEE international conference on power electronics, intelligent control and energy systems, ICPEICES 2016*
19. Narayan Y, Mathew L, Chatterji S (2018) SEMG signal classification with novel feature extraction using different machine learning approaches. *J Intell Fuzzy Syst* 35(5):5099–5109
20. Sui X, Wan K, Zhang Y (2019) Pattern recognition of SEMG based on wavelet packet transform and improved SVM. *Optik (Stuttg)* 176:228–235

21. Li C, Li G, Jiang G, Chen D, Liu H (2018) Surface EMG data aggregation processing for intelligent prosthetic action recognition. *Neural Comput Appl* 0123456789(8):1–12
22. Waris A et al (2018) The effect of time on EMG classification of hand motions in able-bodied and transradial amputees. *J Electromyogr Kinesiol* 40(4):72–80
23. Garg C, Narayan Y, Mathew L (2015) Development of a software module for feature extraction and classification of EMG signals. In: 2015 communication, control and intelligent systems (CCIS), vol 1, pp 250–254
24. Oskoei MA, Hu HHH (2008) Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE Trans Biomed Eng* 55(8):1956–1965
25. Phinyomark A, Nuidod A, Phukpattaranont P, Limsakul C (2012) Feature extraction and reduction of wavelet transform coefficients for EMG pattern classification. *Elektron Ir Elektrotechnika* 122(6):27–32
26. Phinyomark A, Limsakul C, Phukpattaranont P (2011) Application of wavelet analysis in EMG feature extraction for pattern classification. *Meas Sci Rev* 11(2):45–52
27. Samuel OW et al (2017) Pattern recognition of electromyography signals based on novel time domain features for amputees' limb motion. *Comput Electr Eng* 67(10):646–655
28. Subasi A (2012) Classification of EMG signals using combined features and soft computing techniques. *Appl Soft Comput J* 12(8):2188–2198
29. Phinyomark A, Phukpattaranont P, Limsakul C (2012) The usefulness of wavelet transform to reduce noise in the SEMG signal. *EMG Methods Eval Muscle Nerve Func*, no. February 2016, pp 107–132
30. Hortal E et al (2015) SVM-based brain-machine interface for controlling a robot arm through four mental tasks. *Neurocomputing* 151(10):116–121
31. Subasi A (2013) Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Comput Biol Med* 43(5):576–586
32. Alomari F, Liu G (2015) Novel hybrid soft computing pattern recognition system SVM-GAPSO for classification of eight different hand motions. *Optik (Stuttg)* 126(23):4757–4762
33. Alkan A, Günay M (2012) Identification of EMG signals using discriminant analysis and SVM classifier. *Expert Syst Appl* 39(1):44–47

A Comparative Analysis on Hyperspectral Imaging-Based Early Drought Stress Detection for Precision Agriculture in Indian Context



Gajanan H. Chavhan, Yogesh H. Dandawate, and Mangesh S. Deshpande

Abstract Drought stress can be defined as the absence of rainfall or irrigation for a period of time sufficient to deplete soil moisture and cause dehydration in plant tissues. The early detection of drought stress can help in saving the reduction in crop production and dilution in quality of the produced farm yields. The visual inspection of the farm plats is done by farmers to understand the drought stresses in plants. This limited vision analysis of human eyes cannot detect the stresses at early stages. The stress detected at later stage leads depletion of the plant health and reduces the production of the plant yield. Hyperspectral imaging is widely used for analysing the land statistics, vegetation, forest plots, water bodies, etc. The airborne imaging and satellite-based remote sensing data collection helps in monitoring farm yields and projecting national farm product index. The most common reflectance parameters used for detection of the stress parameters in plants are considered by many researchers, and major of them are summarized in this paper. Common water indices as PWI-Plant Water Index, SRWI-Simple Ratio Water Index, NDWI-Normalized Difference Water Index, SIWSI-Shortwave Infrared Water Stress Index, CAI-Cellulose Adsorption Index, etc. are widely accepted as associated parameters for water stress coefficients.

Keywords Precision agriculture · PWI · SRWI · NDWI · SIWSI · CAI

1 Introduction

Farming is the main earning source for over 58% of the residents of rural Indians. As many as every second person is engaged in this activity and follows it as their family business and contributes to 20.4% of the gross value added (GVA) during 2016–2017 and 15% of national gross domestic product (GDP). The environmental

G. H. Chavhan (✉) · M. S. Deshpande

Department of Electronics and Telecommunication Engineering, VIT, Pune, Maharashtra, India
e-mail: ghchavhan@gmail.com

Y. H. Dandawate

Department of Electronics and Telecommunication Engineering, VIIT, Pune, Maharashtra, India

diversity in different parts of the country is the major deciding factors of Indian farm products. The current state of Indian farming is on the basis of geographical locations of the farm and local environmental conditions only. Almost all farming is done on the basis of past experiences of the farmers and with almost no scientific basis for decision making. On the contrary, developed countries like Israel uses maximum of technology developments, facts, and statistics for crop-related decisions making, which produces quality products and high yields.

The national status in Hyperspectral imaging is very nascent in India, and very few hyperspectral imaging data is available for analysis. ISRO collected (HySI) vegetation data by using IMS-1 satellite imagery (2008–2012). Currently, ISRO–NASA initiated the airborne hyperspectral images of 100 Indian cities under ISRO–NASA AVIRIS Science project 2018–2019; AVIRIS-NG Project phase 2 will be the first attempt of producing Hyperspectral images of 6 m pixel resolution at 5 nm band interval with wide range (380–2510 nm) of spectral bandwidth. This data will lead to the development of statistical tools and framework development for supporting decision making in farming.

In India, hyperspectral imaging-based drought stress identification is an open-ended problem and has thrust areas which can lead the next revolution in precision agriculture-based Indian farming. With the advent of latest imagery techniques developed recently and available worldwide, there is an immediate need of attempting this research problem in the Indian context of socio-economical effects. The proposed literature work studied drought stress in crop plants which will ease implementation of futuristic precision farming methodology for Indian farms.

India is an agriculture-based country, and more than 55% of active working population is actively or passively involved in farming activity. The farm filed information is very crucial, and time of availability of data is rather more critical. Many government organisations strive hard to collect authenticate and share this farm field data with the farmers through many available platforms. The Indian Space Research Organization (ISRO) also collects this data through satellites and share RADAR and OPTICAL imagery with the farmers. In monsoon season, radar imagery is very useful for predicting and analysing the monsoon advancement and area coverages. The convergence of all this data can lead to help farmers for decision making and data-centric farming approach.

This strengthening of agriculture sector can lead to sustainable economic as well as human development. By using the technology and hyperspectral imaging analysis, the development of agriculture will help to elevate the farmer's life and will lead to increase the overall earnings of the farmers and farm labourers. Technology-based agriculture development is the need of the hour and should be treated without most priority by country like India where we have second-largest population of the world. This will not only elevate the living standards of the farmers and related industry stake holders but also can contribute a huge enhancement in country's GDP.

This literature revived in the area of precision agriculture so as to present the state of the art of agriculture through geographic information systems (GIS) and hyperspectral imaging. This study can lead to the development of advisory system for Indian farmers on the basis of available resources and research outcomes. Such

system can help farmers to make sensible and appropriate decision against the natural primitive knowledge gained over the time. The scientific mode of farming can definitely enhance the lifestyle of the farmers and can get more yields.

The precise use of water, pesticide and resources will help farmers to produce better quality farm products at comparatively lower cost. The efficient use of water resources will help in increasing the crops production and also help in depletion of the water bodies unnecessarily.

2 Literature Survey

The hyperspectral imaging technique uses a wide spectral band (400–2500 nm) which detects vivid plant-related parameters and stress (physiological, nutrient and water-related stresses) which human eyes cannot detect. A number of researchers have tried with different spectral bands and proposed different algorithms to detect water stress using hyperspectral imagery. The literature is reviewed and classified on the basis of approach used by the researcher, different databases used, and the classifier used to address the issue. Table 1 shows the summery of the various studies about water stress detection.

As it is clearly evident from above table, majority of researchers use following water indices in the algorithms developed for water stress detection. The major water indices are NDWI, SRWI, SIWSI, PWI, CAI, etc. (Table 2).

3 Conclusion

The use of current space-borne and airborne hyperspectral imaging technology for capturing the water stress parameters and development of water stresses in plants at early stages of the crop development need to be studied rigorously. The narrow band hyperspectral images can be collected by using satellite-based multispectral imaging sensors which will help in precision agriculture for Indian farmers. A platform can be then developed for analysing the collected images for detection and a system can be thus developed for helping the farmers. As seen in the referred literatures and the work done by the researchers clearly indicate the scope of the study in India is very high and can lead to provide suitable and accurate system support for the farmers. Many countries have already mastered the precision farming methodologies and enjoying the fruits of the same now.

Table 1 Summary of water stress detection

Researcher	Approach	Database	Classifier	Remarks
Whiting [1]	Narrow band index technique, equivalent water thickness, NDVI	Multispectral lab data, SPOT imagery, AVIRIS–NASA, AVNIR	PCA, vegetation index	0.98 μm
Roberts [2]	LAI–leaf area index, NIR to RED index, simple ration SR, canopy water absorption	IDEAS	RMS, MAE, LAI prediction algorithm	Moisture stress index, water band index
Bauriegel [3]	Fusarium head blight (FHB) detection in wheat, chlorophyll fluorescence imaging	AVIRIS	SAM classification method, Min. distance parallelepiped, Max. likelihood, binary encoding, SAM classification method, Mahalanobis distribution	
Fahlgren [4]	Phenotyping plant traits, countrified phenotyping	Phytomorph.wise.edu, PlantCV	NIR, SWIR, TIR/LWIR	900–1700 nm analysis
Behmann [5]	Detection of plant stress from HIS, drought stress prediction system performances are major factors of logarithmic prediction algorithms	Barlay dataset, surface optics	Linear ordinal SVM, one verses one support vector machine, one verses all support vector machine, SVR, SVORIM	Best analysis tool—linear ordinal SVM
Kizil [6]	Water stress detection, spectral data and ANN can predict water stress, lettuce leaf studied, Spectral VI	Controlled environment database is created, ISMAIL HAKKI–design	R2 ANN, NDVI simple ration, CL chlorophyll, BPNN (Feed forward Backward Propagation ANN)	33% water treatment, three plots of 4000 cm ² area in controlled environment
Kurz [7]	Airborne imaging, UAV-based imaging of plots, Visible and NIR spectral range, Close range HIS	HySpex SWIR	Reflectance coefficient, vegetation indices	Step-by-step flow development for stress detection

(continued)

Table 1 (continued)

Researcher	Approach	Database	Classifier	Remarks
Baran [8]	Spectral imaging with special location, separation between material and vegetation, intrinsic dimension estimation ID	AVIRIS data	Virtual dimensionality VD, PCA, ID, Neyman–Pearson theory	224 channels, LSMA, HFC

Table 2 Summary of parameters used by researchers to specify water contents for performance evaluation of the different algorithms

Title	Structural vegetation indexes	NDWI	SRWI	SIWSI	PWI	CAI
Whiting [1]	✓	✓	✓	✓	✓	
Liua [9]		✓	✓			
Roberts [2]	✓	✓		✓		✓
Singh [10]		✓	✓	✓		
Ge [6]	✓	✓			✓	✓
Kurz [8]		✓	✓	✓	✓	
Singh [11]	✓				✓	
Hestir [12]	✓	✓	✓			
Liua [9]	✓	✓			✓	
Näsia [13]	✓				✓	✓
Ge [14]		✓			✓	✓
Schmitter [35]		✓		✓	✓	✓
Cavender [15]	✓	✓				✓
Fahlgren [4]	✓			✓		✓
Chang [16]	✓	✓	✓		✓	
Baranoskia [17]		✓		✓		✓
Behmann [5]	✓			✓	✓	
Collingwood [18]				✓		✓
Lü [19]	✓	✓		✓		
Schut [20]	✓		✓			

References

1. Whiting ML, Ustin SL, Zarco-Tejada P (2006) Hyperspectral mapping of crop and soils for precision agriculture. *Remote Sens Model Ecosyst Sustain Proc SPIE* 6298:62980B
2. Roberts DA, Roth KL, Perroy RL 14 hyperspectral vegetation indices. In: Chapter 14, hyperspectral remote sensing of vegetation, CRC Press

3. Bauriegel E, Herppich WB (2014) Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to *Fusarium* spec. Infections on wheat. *J Agric* 32:57
4. Fahlgren N, Gehan MA, Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *J Current Opin Plant Biol* 24:93–99
5. J. Behmann, P. Schmitter, J. Steinrücken, L. Plümer “Ordinal classification for efficient plant stress prediction in hyperspectral data,” ISPRS Technical Commission VII Symposium, 29 September – 2 October 2014.
6. Ünalkızıl, levent gençİ, melisinalpulat, Duyguşapolyo, Mustafa mirik Lettuce yield prediction under water stress using artificial neural network (ANN) model and vegetation indices. *J Agric* 99:409–418
7. Kurz TH, Buckley SJ A review of hyperspectral Imaging in close range applications, vol XXIII ISPRS Congress, 12–19 July 2016
8. Baran D, Apostolescu N (2014) A virtual dimensionality method for hyperspectral imagery. In: 25th DAAAM international symposium on intelligent manufacturing and automation, DAAAM
9. Näsia R, Honkavaara E, Blomqvist M, Lyytikäinen-Saarenmaa P, Hakala T, Viljanen N, Kantola T, Holopainen M (2018) Remote sensing of bark beetle damage in urban forests at individual tree level using a novel hyperspectral camera from UAV and aircraft. *J Urban Forest Urban Green* 30:72–83
10. Singh A, Ganapathysubramanian B, Singh AK, Sarkar S (2016) Machine learning for high-throughput stress phenotyping in plants. *J Trends Plant Sci* 21(2)
11. Hestir EL, Brando VE, Bresciani M, Giardino C, Matta E, Villa P, Dekker AG (2015) Measuring freshwater aquatic ecosystems: the need for a hyperspectral global mapping satellite mission. *J Remote Sens Environ* 167:181–195
12. Doneus M, Verhoeven G, Atzberger C, Wess M, Rus M (2014) A new way to extract archaeological information from hyperspectral pixels. *J Archaeol Sci* 52:84–96
13. Lavagnino Z, Dwight J, Ustione A, Nguyen T-U, Tkaczyk TS, Piston DW (2016) Snapshot hyperspectral light-sheet imaging of signal transduction in live pancreatic islets. *Biophys J* 111:409–417
14. Zarco-Tejada PJ, Hornerob A, Hernández-Clemente R, Beck PSA (2018) Understanding the temporal dimension of the red-edge spectral region for forest decline detection using high-resolution hyperspectral and Sentinel-2a imagery. *ISPRS J Photogramm Remote Sens* 137:134–148
15. Cavender-Bares J (2004) From leaves to ecosystems: using chlorophyll fluorescence to assess photosynthesis and plant function in ecological studies. In: *Chlorophyll fluorescence: a signature of photosynthesis*. Kluwer Academic Publishers, pp 737–755
16. Chang C-W, Mausbach MJ, Hurburgh CR Jr Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties, Agricultural and Biosystems Engineering IOWA University
17. Gladimir V. G. Baranoskia, Tenn F. Chena, Bradley W. Kimmela, Erik Mirandaa and Daniel Yima “On the high-fidelity monitoring of C3 and C4 crops under nutrient and water stress,” Proceedings Volume 8524, Land Surface Remote Sensing; 85240W, 2012.
18. Collingwood JF, Adams F (2017) Chemical imaging analysis of the brain with X-ray methods. *J Spectrochimica Acta Part B* 130:101–118
19. Lü Q, Tang M (2012) Detection of hidden bruise on Kiwi fruit using hyperspectral imaging and parallelepiped classification. *Proc Environ Sci* 12:1172–1179
20. Schut AGT, Ketelaars JJM (2003) Early detection of drought stress in grass swards with imaging spectroscopy. *NJAS-Wageningen J Life Sci*
21. WahLiew O, Chong PCJ, Li B, Asundi AK (2008) Signature optical cues: emerging technologies for monitoring plant health. *J Sens* 3205–3239
22. Ge Y, Bai G, Stoerger V, Schnable JC (2016) Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *J Comput Electron Agric* 625–632

23. Stratoulis D, Balzter H, Ziinszky A, Tótha VR (2015) Assessment of ecophysiology of lake shore reed vegetation based on chlorophyll fluorescence, field spectroscopy and hyperspectral airborne imagery. *J Remote Sens Environ* 157:72–84
24. Arellano P, Tansey K, Balzter H, Boyd DS (2015) Detecting the effects of hydrocarbon pollution in the Amazon forest using hyperspectral satellite images. *J Environ Pollut* 205:225–239
25. Duana L, Huang C, Chend G, Xiong L, Liub Q, Yanga W (2015) Determination of rice panicle numbers during heading by multi-angle imaging. *Crop J* 3:211–219
26. Cruz JA, Savage LJ, Zegarac R, Kovac WK, Chen J, Kramer DM (2016) Dynamic environmental photosynthetic imaging reveals emergent phenotypes. *J Cell Syst* 2:365–377
27. Liu K, Zhou Q, Wu W, Xia T, Tang H (2016) Estimating the crop leaf area index using hyperspectral remote sensing. *J Integrative Agric* 15:475–491
28. Holzinger A, Allen MC, Deheyn DD (2016) Hyperspectral imaging of snow algae and green algae from aero terrestrial habitats. *J Photochem Photobiol B* 162:412–420
29. Schut AGT, Ketelaars JJMH (2003) Imaging spectroscopy for early detection of nitrogen deficiency in grass swards. *J NJAS*, 51–53
30. Carasa T, Hedley J, Karnieli A (2016) Implications of sensor design for coral reef detection: upscaling ground hyperspectral imagery in spatial and spectral scales. *Int J Appl Earth Obs Geoinform* 63:68–77
31. Dall’Ara E, Boudiffaa M, Taylorc, Schugd, Fiegled, Kennerleye AJ, Damianouf C, Tozera GM, Kiessling F, Müllerc R (2016) Longitudinal imaging of the ageing mouse. *J Mech Ageing Devel* 160:93–116
32. Singh A, Ganapathysubramanian B, Singh AK, Sarkar S (2016) Machine learning for high-throughput stress phenotyping in plants. *J Trends Plant Sci* 21(2)
33. Liua W, Zhanga HF (2016) Photoacoustic imaging of the eye: a mini review. *J Photoacoust* 4:112–123
34. Ge Y, Bai G, Stoerger V, Schnable JC (2016) Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *J Comput Electron Agric* 127:625–632
35. Schmitter P, Steinrücken J, Römer C, Ballvora A, Léon J, Rascher U, Plümer L (2018) A unsupervised domain adaptation for early detection of drought stress in hyperspectral images. *ISPRS J Photogramm Remote Sens* 131:65–76

Soft and Cloud Computing

Proposing Perturbation Application Tool for Portable Data Security in Cloud Computing



Amit Kumar Chaturvedi, Meetendra Singh Chahar, and Kalpana Sharma

Abstract Cloud computing is an emerging field of computing in a shared resources environment, where multiple clients share the computing resources. The data will be stored on the shared third-party cloud servers and hence it will be in other's hand. There may be the chance of mishandling the data, if the data is placed in original form. Perturbation is a technique that will not only reduce the understanding level of the data but may convert the data sufficiently in such a form that will not be in the preview of any language parser. The main feature of this perturbation technique is that it will provide the facility to implement the perturbation in an innovative form. This paper proposes an innovative "Perturbation Application Tool (PAT)" that provides the facility to implement noise addition algorithm with encryption. We have also illustrated the implementation and outcomes of the PAT tool and this tool is working successfully for securing the portable data.

Keywords Cloud computing · Server · Security · Perturbation · Encryption · Noise addition

1 Introduction

Cloud computing is a now a popular and useful means of computing for creating variety of application in different service areas. For infrastructure providers, it is good computing platform to share its valuable resources to the potential developers on pay per use basis and is a good means to get the best value for their resources. From developers point of view, it is good platform for the development because

A. K. Chaturvedi (✉)
MCA Department, Govt. Engineering College, Ajmer, India
e-mail: amit0581@gmail.com

M. S. Chahar · K. Sharma
CS Department, Bhagwant University, Ajmer, India
e-mail: meetendra26@gmail.com

K. Sharma
e-mail: kalpanasharma56@gmail.com

if anyone in the field of IT and have an innovative idea for the computing world, then it can develop application and start the work without worrying of infrastructure purchasing or establishing a communication system between provider and customers. The applications can be started at anytime from anywhere, and all the requirement can be arranged online on pay per use basis. This is potential of the cloud computing.

When we discuss the potential benefits of the cloud computing, it is also required to discuss the challenges and disadvantages of the system. Every system has some drawback, and hence, the cloud computing also has some drawbacks. Cloud computing uses five-layered governance structure of IT organization on which all the cloud computing based application executes. The five layers are: Network, Storage, Server, Services, and Applications.

Applications

Services

Server

Storage

Network

These all the service if owned by the organization then it is called as on-premise services, but these all the services application developer may hire on pay per use basis. As much as among these services are hired by the developer, the control will be more shifted toward the vendor providing these services [1–4]. But according to the nature of the cloud computing, all the services in network area are hired from the other vendors; i.e., we use already established networks from the network service providers. In the same manner, there are storage servers providers which provide online storage services on pay per use basis. There are various pre-configured servers like Apache Server, Web Server, etc. and the developer can hire the services of these servers directly by online paying for it. In the similar way, variety of web services and applications are also available of rental bases in the cloud computing on hired bases. Every type of solution is available for the customer today online through cloud computing solution.

But security of data is a very important issue still in cloud computing. Because when your application is residing and executing on a server, but your database or data resides on another server then the security of data is very insecure [5, 6]. You have the application codes and you can maintain it also by yourself but what about the data of customers, which are regularly working online in a 24×7 basis, if it not disturbed but passed to the other who might misuse it. Then the question arises who is responsible and how the responsibility will be fixed. The customer is only the bearer in this case. So, the security of the data is very important issue. This is motivation for writing this paper for us.

2 Perturbation Technique

Perturbation technique is a new technique in the data security dimension of cloud computing [7–12]. This technique applies a transformation on customer’s data, which reduces the usefulness of the underlying data. This transformation is also called noise that means this technique adds some noise to the data that reduces the usefulness of the underlying data. There are different ways to add noise in the data, each method of noise addition has different degree of noise. The degree may be measured on the basis that how much changes will occur in a unit of data [13–15]. As the degree of noise increases, the security level will be increased, but the delay will also be increased and hence the performance will be decreased [16–18]. So, the degree of noise should be limit by a threshold value, say k is a threshold value.

$$\text{data security} \propto \text{degree of noise} \quad (1)$$

$$\text{Performance} \propto \frac{1}{\text{degree of noise}} \quad (2)$$

$$\text{degree of noise} < k, \text{ where } k \text{ is a threshold value of noise} \quad (3)$$

Using encryption and decryption is another very important method to transform the data from its original form to another form and it definitely reduces the usefulness of the underlying data and helps in securing the data [19–22]. There are different encryption and decryption algorithms like RSA, DSA, AES, DES, etc. Each encryption and decryption algorithm has some advantages and disadvantages [23–25]. So, we may choose any algorithm according to the required feature in it.

3 Proposed Tool’s Model and Methodology

In this paper, we are proposing Perturbation Application Tool (PAT) for adding noise and then applying DES encryption algorithm for securing the data before sending it to the third-party cloud server.

This PAT tool mainly contributes in two major areas of perturbation: (i) Noise addition, i.e., improving the security by maintaining the degree of noise and (ii) application of DES encryption algorithm. Before presenting the working of tool, let us first discuss the model used to describe the steps of this PAT tool. Figure 1 shows the perturbation application model.

As shown in Fig. 1a, b that there are two applications of proposed PAT tool. The first application is shown in Fig. 1a, in which the “Plain.txt” file containing the original contents will be converted into a file that contains the data of plain.txt file but with noise. It means the understanding level of the data of “Plain.txt” is reduced in such a way that the contents will lose their meaning and it not fit in the syntax

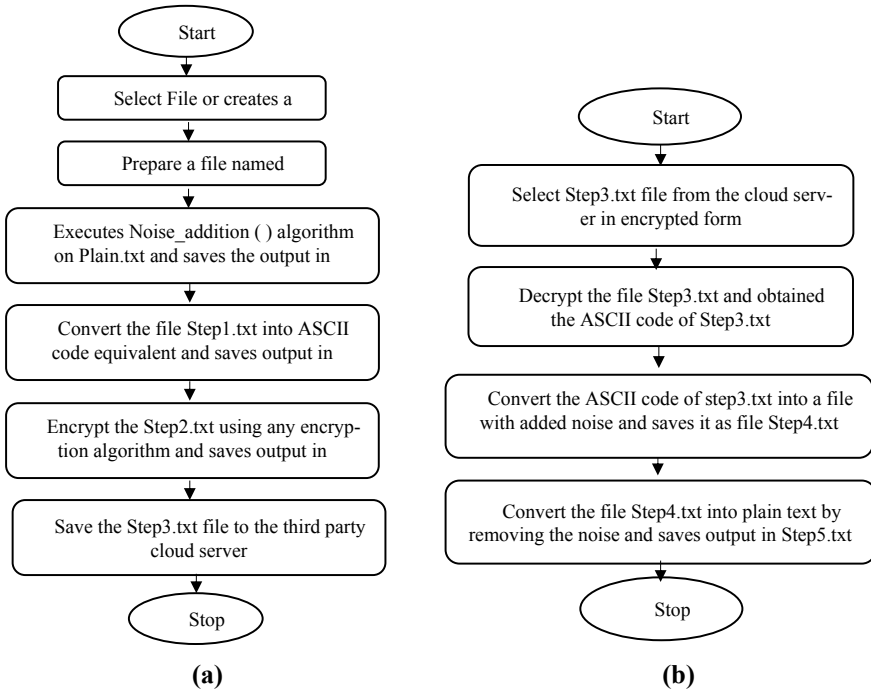


Fig. 1 Application of PAT tool for converting. Application of PAT tool for converting encrypted noisy file into plain text file

of any language parser. The data hackers are basically the servers, when pick this data and analyze it, this noisy data itself defeat their purposes. Hence, this is main contribution of our proposed PAT tool, i.e., Noise_addition() algorithm.

In addition to that, we also increase the level of security in this proposed PAT tool by converting this noisy data, which is available in “Step1.txt” file, into ASCII form and saves it into the “Step2.txt” file. Now, as the data is now in the ASCII form, the tool will apply the DES encryption and converted this ASCII form of data into encrypted form and saves the output in the “Step3.txt” file. Now, this file is very much secured than before and is totally safe to upload on the cloud server for further processing.

The second application of this PAT tool is shown in Fig. 1b and is used when the user wants to recover original form of the data from the encrypted form, i.e., “Step3.txt”. It will first download or select the file Stept3.txt from the cloud server. Then in the first step, the PAT tool covert the decrypt the contents of the “Step3.txt” file and the output of this process will be the ASCII form of the “Step3.txt” file’s data and it is the “Step4.txt”. In the second phase of this process, the PAT tool will convert this ASCII form of the data, i.e., “Step4.txt” into the noisy form. This file will have the noisy form of data and is not understandable. In the third phase, the PAT tool will convert this noisy data into its original form and saves it as “Step5.txt”.

4 Perturbation Application Tool [PAT] and Results

The Perturbation Application Tool [PAT] is prepared in Java and shown in Fig. 2.

Let there is a file “Plain.txt” as shown in Fig. 3 with some data that is to be secured before saving this data on the third-party cloud server.

The file “Plain.txt” is having the original content and we have to save this file “Plain.txt” on the third-party cloud server. So, before saving it on the cloud server, the Perturbation Application Tool (PAT) will be executed to perform the first step of code conversion to secure this file contents, i.e., the Noise_addition() algorithm and this step will create a file “Step1.txt”, that will reduce the understanding level of the

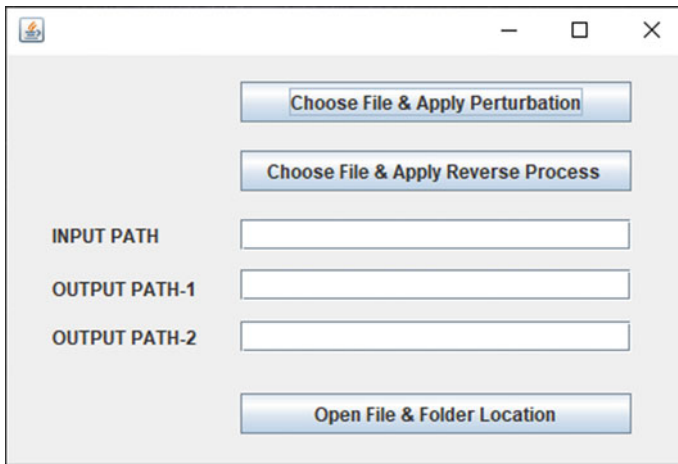


Fig. 2 Perturbation application tool [PAT]

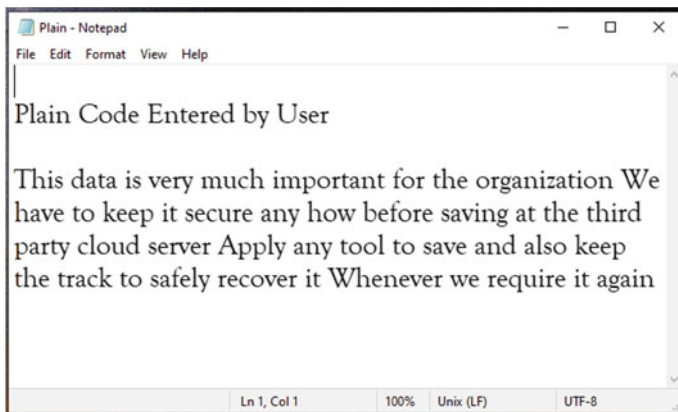


Fig. 3 Plain.txt file

contents in “Plain.txt” file. This “Step1.txt” file is shown in Fig. 4. It is very clear from Fig. 4 that the Noise_addition() algorithm sufficiently reduces the understanding level of the data of the “Plain.txt” file. No language parser will work on this data.

The proposed PAT tool in addition to this will also work on increasing the security level of the data and it will perform the code conversion from text to ASCII in the second step. The outcome of this step is shown in Fig. 5 and saves the output in the system by the file name Step2.txt.

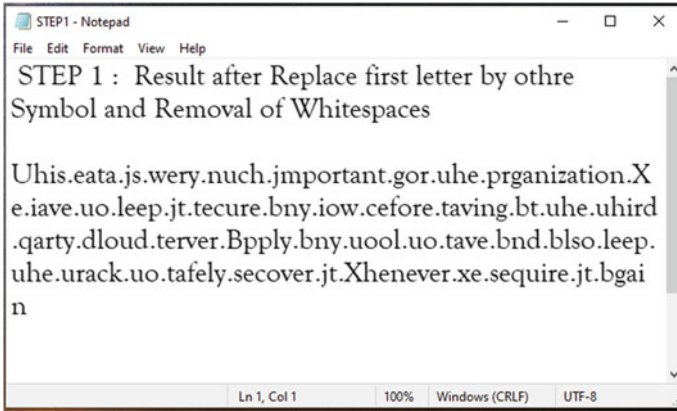


Fig. 4 Step1.txt

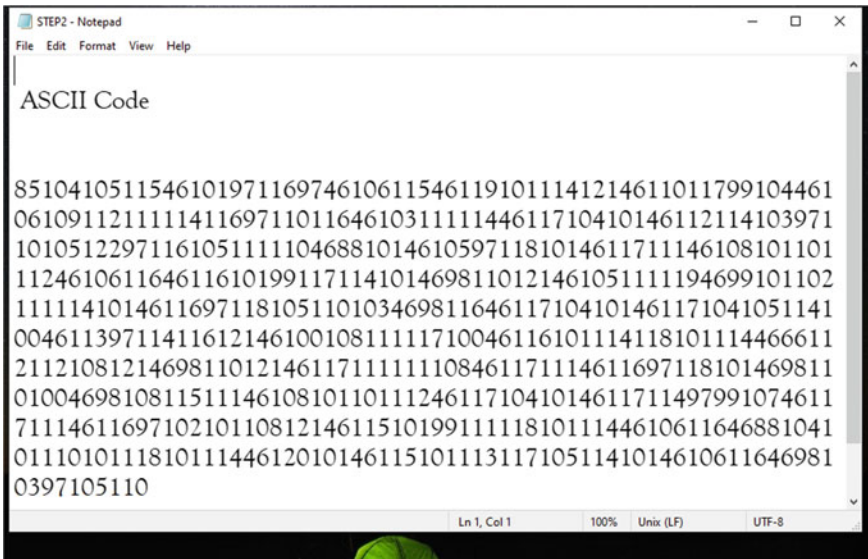


Fig. 5 Step2.txt

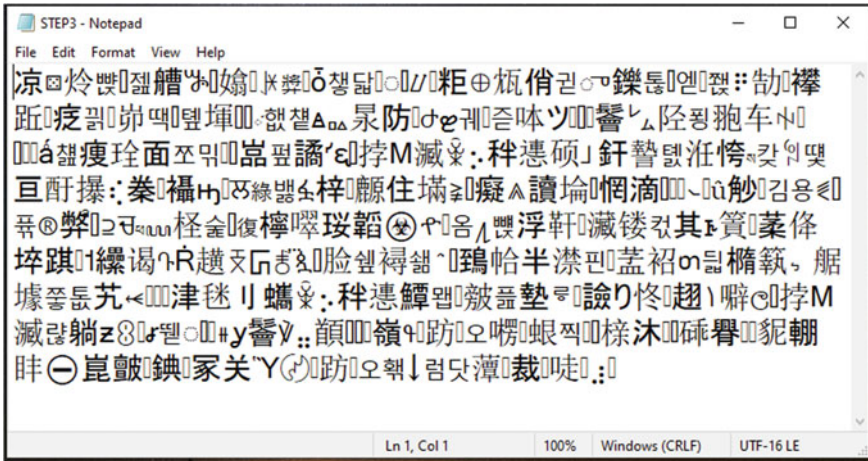


Fig. 6 Step3.txt

Proposed tool now applies DES encryption algorithm for encrypting the contents of “Step2.txt” file and generates the final form of the original contents, i.e., “Step3.txt”. Encrypted form of the Plain.txt file is shown in Fig. 6. Now, this output file, i.e., “Step3.txt” is very safe and secure to sand it on the third-party cloud server.

The above-discussed PAT tool has two modes of application, in the first mode of application it performs three steps of code conversion method and converts the Plain.txt contents to encrypted text form as shown in “Step3.txt” and it is already discussed above.

Second mode of application of this PAT tool works in reverse direction, i.e., to generate the original contents from the encrypted form. “Step3.txt” file having the encrypted form of data and now let us understand the steps of this reverse process of code conversion. The first step is to covert the encrypted form of the data present in “Step3.txt” file into ASCII code form and save it in the Step4.txt file. The output of this step is shown as file “Step4.txt” in Fig. 7.

The second step of this reverse process of code conversion is to convert this Step4.txt into noisy form in first phase and then in the plain text as it is in the start of this whole process. This is performed and saved in the “Step5.txt” file and is shown in Fig. 8.

Hence, the PAT tool has successfully performed the given task. Another important point to be considered here is that the size of the files remains under limit. In many perturbation techniques, the size increases in multiples but with the PAT tool it is well maintained. The sizes of the files used in this whole process are listed below:

S. No.	Name of the file	Size of the file (KB)
1	Plain.txt	1
2	Step1.txt	1

(continued)

(continued)

S. No.	Name of the file	Size of the file (KB)
3	Step2.txt	1
4	Step3.txt	1
5	Step4.txt	1
6	Step5.txt	1

5 Conclusion

Data security is an important aspect when we work in a distributed and open architecture like cloud computing. In cloud computing, the applications, services, and computing resources are hired on pay per use basis. The data is stored, processed, and served by third-party-controlled computing resources. The data can be confidential or sharable. The confidential data should be kept in such a way that no one can use it for malicious purposes. So, for securing such confidential data, perturbation technique may be applied.

Perturbation technique provides the opportunity to the developer to set the noise level according to the requirement. Proposed PAT tool utilizes the noise level in such an efficient way that the size of the files is under control and not increased so much. The innovative Noise_addition() algorithm is unique and sufficiently reduces the understanding level of the data and keeps the data reserved from the reach of language parsers. Another feature of PAT tool is the application of DES encryption algorithm, which definitely increases the security of the data and improves the security level.

Hence, the proposed PAT tool has successfully implemented the perturbation technique, the Noise_addition() algorithm is unique and successfully serving its purpose with properly limiting the file sizes. So, PAT tool is an innovative tool for sharing the data secretly in a distributed and open architecture computing environment like cloud computing.

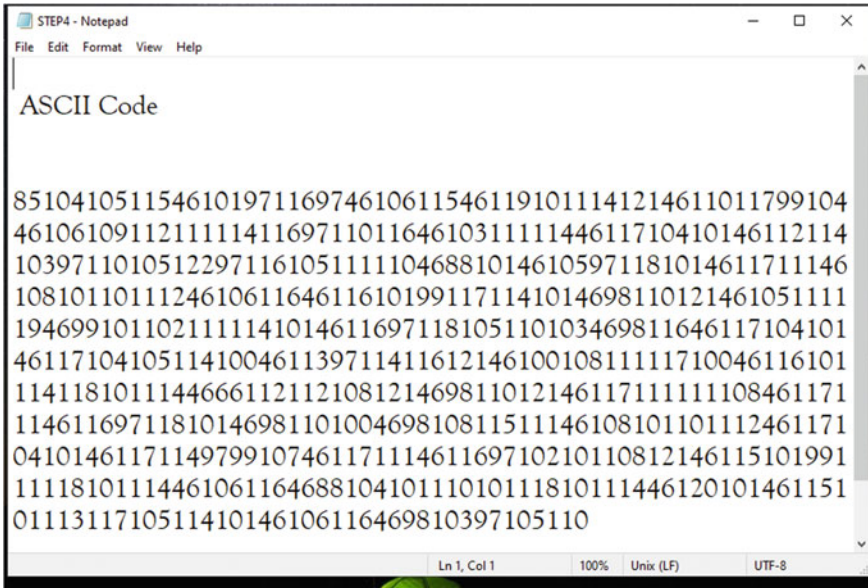


Fig. 7 Step4.txt

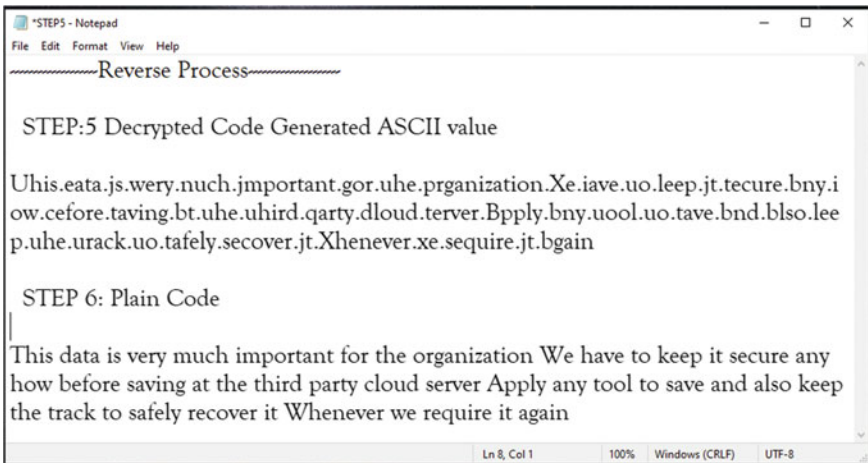


Fig. 8 Step5.txt

References

1. Gurevich A, Gudes E (2006) Privacy preserving data mining algorithms without the use of secure computation or perturbation. In: 10th international database engineering and applications symposium (IDEAS '06).<https://doi.org/10.1109/ideas.2006.37>

2. Li L, Kantarcioglu M, Thuraisingham B (2009) Privacy preserving decision tree mining from perturbed data. In: 42nd Hawaii international conference on system sciences. <https://doi.org/10.1109/hicss.2009.353>
3. Reddy VS, Rao BT (2017) A combined clustering and geometric data perturbation approach for enriching privacy preservation of healthcare data in hybrid clouds. Department of Computer Science and Engineering, Koneru Lakshmaiah University, Vijayawada, India
4. Shinde A, Saxena K, Mishra A, Sahu SK (2016) Privacy prevention of sensitive rules and values using perturbation technique. IEEE, pp 577–581. ISSN: 978-9-3805-4421-2/16/\$31.00_c
5. Kaur M, Jain A, Verma A (2017) Optimized cloud storage capacity using data hashes with genetically modified SHA3 algorithm. In: International conference on energy, communication, data analytics and soft computing (ICECDS), pp 2980–2984. <https://doi.org/10.1109/icecds.2017.8390002>
6. Tangwongsan S, Itthisombat V (2014) A highly effective security model for privacy preserving on cloud storage. In: IEEE 3rd international conference on cloud computing and intelligence systems, pp 505–509. <https://doi.org/10.1109/ccis.2014.7175788>
7. Kamakshi P (2014) A survey on privacy issues and privacy preservation in spatial data mining. In: 2014, international conference on circuits, power and computing technologies [ICCPCT-2014]. <https://doi.org/10.1109/iccpct.2014.7054961>
8. Siddhpura A, Vekariya DV (2018) An approach of privacy preserving data mining using perturbation & cryptography technique. Int J Future Revol Comput Sci Commun Eng 4(4):255–259. ISSN: 2454-4248
9. Kaur A (2017) A hybrid approach of privacy preserving data mining using suppression and perturbation techniques. In: International conference on innovative mechanisms for industry applications (ICIMIA). <https://doi.org/10.1109/icimia.2017.7975625>
10. Kaur A, Sofat S (2016) A proposed hybrid approach for privacy preserving data mining. In: International conference on inventive computation technologies (ICICT). <https://doi.org/10.1109/inventive.2016.7823283>
11. Stant O, Sirdey R, Gouy-Pailler C, Blanchart P, BenHamida A, Zayani M-H Privacy-preserving tax calculations in smart cities by means of inner-product functional encryption. In: 2018, 2nd cyber security in networking conference (CSNet). <https://doi.org/10.1109/csnet.2018.8602714>
12. Pooja HP, Nagarathna N (2015) Privacy preserving issues and their solutions in cloud computing: a survey. Int J Comput Sci Inform Technol (IJCSIT) 6(2):1588–1592
13. Dhiman EV, Himakshi E, Kaur EA, Kumar M (2017) Pragmatic approach to conquer security perturbation in cloud computing using level classification. In: 2nd international conference for convergence in technology (I2CT). <https://doi.org/10.1109/i2ct.2017.8226119>
14. Yonezawa C, Takeuchi S (2003) Perturbation caused by cloud in ERS SAR interferogram. In: IGARSS-2003, IEEE international geoscience and remote sensing symposium. Proceedings (IEEE Cat. No.03CH37477), pp 4365–4367. <https://doi.org/10.1109/igarss.2003.1295517>
15. Yan K, Du Y, Ren Z (2018) MPPT perturbation optimization of photovoltaic power systems based on solar irradiance data classification. IEEE Trans Sustain Energy 1–8. <https://doi.org/10.1109/tste.2018.2834415>
16. Peng K, Bao F (2010) Trust management in privacy—preserving information system. In: IEEE conference, pp 1–4, 978-1-4244-5540-9/10/\$26.00 ©2010 IEEE
17. Mochizuki Y, Manabe Y (2015) A privacy-preserving collaborative filtering protocol considering updates. In: 10th Asia-Pacific symposium on information and telecommunication technologies (APSITT), pp 142–144. <https://doi.org/10.1109/apsitt.2015.7217100>
18. Drosatos G, Efraimidis PS, Athanasiadis IN, D’Hondt E, Stevens M (2012) A privacy-preserving cloud computing system for creating participatory noise maps. In: IEEE 36th annual computer software and applications conference, pp 581–586. <https://doi.org/10.1109/compsac.2012.78>
19. Pariselvam S, Swarnamukhi M (2019) Encrypted cloud based personal health record management using DES scheme. In: IEEE international conference on system, computation, automation and networking (ICSCAN). <https://doi.org/10.1109/icscan.2019.8878773>

20. Kale PV, Welekar R (2017) A survey on different techniques for encrypted cloud data. In: International conference on intelligent computing and control systems (ICICCS), pp 245–247. <https://doi.org/10.1109/iccons.2017.8250718>
21. Shen J, Liu D, Shen J, Tan H, He D (2015) Privacy preserving search schemes over encrypted cloud data: a comparative survey. In: First international conference on computational intelligence theory, systems and applications (CCITSA), pp 197–202. <https://doi.org/10.1109/ccitsa.2015.46>
22. Wang H, Zheng Z, Wang Y (2013) A new privacy-preserving broadcast encryption scheme from DPVS. In: 5th international conference on intelligent networking and collaborative systems, pp 329–334. <https://doi.org/10.1109/incos.2013.61>
23. Saxena VK, Pushkar S (2014) Privacy preserving model in cloud environment. In: Conference on IT in business, industry and government (CSIBIG). <https://doi.org/10.1109/csibig.2014.7056953>
24. Gui Q, Cheng X (2009) A privacy-preserving distributed method for mining association rules. In: International conference on artificial intelligence and computational intelligence, pp 294–297. <https://doi.org/10.1109/aici.2009.486>
25. Biswal B (2009) Privacy preserving data communication model. In: Second international conference on emerging trends in engineering & technology, pp 333–336. <https://doi.org/10.1109/icetec.2009.185>

Horizontal Dynamic Resource Scaling by Measuring the Impacts of Scheduling Interval in Cloud Computing



Amit Kumar Chaturvedi, Praveen Sengar, and Kalpana Sharma

Abstract Cloud servers or Host machines are having a sufficient number of computing resources, and these resources will be shared among its client applications as per the requirement. The traffic on the cloud servers or Host machines is unpredictable, and the demand may increase or decrease with a great variations many times. Even there are lots of reasons of such variations in traffic, here we are not going to discuss such reasons. During the COVID-19 pandemic, most of the meetings, teaching–learning, documents exchanging, etc. are performed using online applications. It is noticed that the both data and processing load are increased in multiple times during the last six months on the social networking sites, and clients faced the problem of low connectivity and slow processing of applications. These cases show that there is very great requirement of load balancing of computing resources on cloud servers to maintain the smooth processing of tasks. In this paper, our main focus is to measure the impacts of scheduling interval on dynamic resource allocation and to calculate the demand of resources dynamically on the basis of dynamic creation or deletion of virtual machines to serve the cloudlet load efficiently. Dynamic resource scaling can be done either horizontally or vertically. In horizontal dynamic resource scaling, the Host machines are equipped with the more number of resources rather than increasing the number of Host machines. In cloud computing, the service provider’s intention is to provide a cost-effective and efficient service support to its clients. The resources and services in cloud computing are on the pay-per-use basis, and hence, the horizontal dynamic resource scaling provides a good cost-effective solution for the resource requirement.

A. K. Chaturvedi (✉) · K. Sharma
MCA Department, Govt. Engineering College, Ajmer, India
e-mail: amit0581@gmail.com

K. Sharma
e-mail: kalpanasharma56@gmail.com

P. Sengar
CS Department, Bhagwant University, Ajmer, India
e-mail: psengar14@gmail.com

Keywords Cloud computing · Elasticity · Dynamic · Resource scaling · Horizontal Scaling · Scheduling interval

1 Introduction

Cloud computing is not a new concept even it is in practice in many organizations to serve their computing requirements. This dimension of computing provides such an efficient and flexible platform where the users need not to worry about arranging the various computing resources [1–3]. It runs on the solutions provided by the various third-party solution providers. The resources and solutions are divided into three service layers: (1) IaaS (Infrastructure as a Service), (2) PaaS (Platform as a Service) and (3) SaaS (Software as a Service). Every service layer deals in a different area of service and hence provides different types of resources, service tools or solutions. All the infrastructure facilities, service tools, platforms, security solutions, data storage solutions, etc. are available on pay-per-use model, i.e., on rental basis. Any service of this useful computing environment can be started and stopped at any time by just paying for that. Many aggregator and extender deal in the area of providing easy access to these services [4–6]. Users with less computing knowledge may also access the benefits of these computing resources, and these services may be accessed with the help of these aggregators and extenders.

The main motivation behind writing this research paper is to understand and illustrate the process of horizontal dynamic resource scaling in cloud computing with elasticity. As the organizations by using cloud computing resources are performing very well in their application areas, such as mailing services, accounting, social networking, etc. So, it is obvious that their user's base will increase in exponential manner and hence their computing resource demand will also increase accordingly [7–9]. There are the load balancers, which maintain the balance between the demand and allocation of resources. But these load balancers will only balance between the demand and available resource allocation balance. So, there will be the need of dynamic resource scaling for proper work execution and maintain the customer's satisfaction in the system, i.e., it will hire more resources on requirement and then do the allocation of these resources by maintaining the load balance.

The manner through which the dynamic resource scaling is done within the limited time frame is called *Elasticity* [10–12]. The *Elasticity* comprises the handling of multiple parameters for providing the services during execution, means for adding extra resources or handling the scheduled tasks with the existing infrastructure resources. The most important parameter is the *scheduling interval*. It is always required that the resource requirements should be fulfilled within the limited time frame that will be called as *rapidity*, or *rapid elasticity*. In this paper, we will discuss the available dynamic resource scaling solutions with *rapid elasticity* and its applications to perform it.

2 Dynamic Resource Scaling

One of the important factors of migrating from one cloud to another cloud is *Scaling* or *elasticity in resource scaling*. Because as we have discussed in the introduction part that we have to pay for everything that is hired in cloud computing. In the beginning of any service or organization, when started working with cloud computing solutions, they hired only limited resources as much as required at that time. With the increase of users and work load, more resources are required and hired accordingly to make a cost-effective solution for better performance [13–16]. So, scalability provides this flexibility to a particular cloud server to maintain the load balance in resource allocation. Dynamic resource scaling is an advanced feature of scaling, in which the demand is estimated by the load balancer at the Data Center. To meet the requirement of the resources, the server dynamically increases or decreases the resources allocation [17, 18]. This is also called as *Auto-scaling*.

This auto-scaling may be done *horizontally* or *vertically*. In *horizontal resource scaling*, resources like Hosts, VMs, CPU, RAM, PEs, bandwidth are increased as much as the resources are there at the Data Center that means number of Hosts may be increased, simultaneously number of VMs, PEs, or other computing resources may be increased to perform the given task or successfully executing the cloudlets without sacrificing on the ground of computing resources or according the infrastructure capabilities to meet the requirement of required performance without increasing the servers. For example, we may understand that in a computer game a warrior is equipped with the more weapons and capabilities to face the enemy not adding more warriors.

In the *vertical resource scaling*, the number of servers or Data Center is increased rather than increasing the underlying hardware, to meet the requirement. If we try to understand it from the above given example of computer game, in this case, new warriors are added with the previous one to handle the execution load or to handle the requirements of increased scaling.

As the computing resources like Hosts, VMs, etc. will increase proportionally with the increase in the execution load or cloudlet load on the server, similarly these resources will be reduced when the execution load or cloudlet load on the server decreases. It means that the cloudlet load on VMs or Hosts are regularly measured or monitored on these created Hosts, VMs by the load balancers or listeners and when it is found that the Hosts/ VMs, or other computing resources become idle; then after spending a pre-specified time, these will be destroyed also. This process releases the occupancy of these computing resources, and hence, the provider or developer will not be burdened to pay for them, as the cloud computing is a pay-per-use model. So, this makes the services more cost-effective as well as provides the *elasticity* to the resource allocation and deallocation process.

Here in this paper, we are more focused toward applying horizontal dynamic resource scaling and measuring the impacts of scheduling interval on horizontal dynamic resource scaling. Load balancers balance load by dynamically creating VMs, according to the arrival of Cloudlets. Cloudlets are dynamically created and

submitted to the broker at specific time intervals. A horizontal VM scaling is set to each list of scalable VMs initially created, that will be checked after a specific time intervals, by checking that the VM is overloaded or not and then requesting the creation of a new VMs to serve the arriving Cloudlets load.

The Data Center Broker is accountable to perform horizontal downscaling. The downscaling is enabled by setting a function using the Data Center Broker's set VM destruction delay function. This function defines the time the broker has to wait to destroy a VM after it becomes idle. If no function is set, the broker just destroys VMs after all running Cloudlets are finished, and there is no Cloudlet waiting to be created.

3 Experimental Setup and Results

The *interval or scheduling interval* is one in which the Data Center will schedule events. As lower is this *scheduling interval*, sooner the processing of VMs and Cloudlets is updated and you will get more notifications about the simulation execution. However, this scheduling interval also affects the simulation performance. A large schedule interval, such as 15, will make that just at every 15 s the processing of VMs is updated. If a VM is overloaded, just after this time the creation of a new one will be requested by the VM's horizontal scaling mechanism. If this interval is defined using a small value, you may get more dynamically created VMs than expected. Accordingly, this value has to be trade-off.

It is also required to define the VM Destruction Delay, so that the broker will wait 10 s before destroying an idle VM. By this time, no down scaling will be performed and idle VMs will be destroyed just after all running Cloudlets are finished and there is no waiting Cloudlet. Another important parameter is the *cloudlets creation interval*, it creates new Cloudlets at every *cloudlets creation interval* (in seconds), up to the 50th simulation second. The method is called every time the simulation clock advances.

The broker will create a list of initial VMs in which each VM is able to scale horizontally, and when it is overloaded, the number of VMs parameter defines a number of VMs to create and return the list of scalable VMs. This is called as horizontal VM scaling.

In this experimental setup, the data is collected with setting number of Hosts parameter at (10, 20, 30, 40, 50) values and checked this with the values of scheduling intervals (5, 10, 15) s. The results collected are displayed below in Table 1.

Now let us examine the results obtained from the experimental setup and analyze it. In Table 1, results are shown with three values of scheduling interval 5, 10 and 15 s with flexibility to scale number of Hosts up to (10, 20, 30, 40, 50). From the above shown results, Fig. 1 shows the reduction in load with the increase in the number of Hosts with constant scheduling interval, i.e., at S.I. = 5 s. It is also observed from the results that the scheduling interval is an important parameter to maintain the performance against the work load.

Table 1 Experimental results

No. of Host preconfigured	DC	No of host actually created	Host PEs CPU cores	No of VMs	Scheduling interval	No of cloudlets	Total exec time	TET/Host	TET/VM	TET/TCLL	TCLL (total cloudlet load)
Host 10	1	10	32	31	5	74	1289	128.9	41.58	1.33	972,000
	1	10	32	12	10	34	738	73.8	61.5	1.56	472,000
	1	10	32	12	15	82	5704	570.4	475.33	5.41	1,054,000
Host 20	1	20	32	39	5	90	1551	77.55	39.77	1.32	1,178,000
	1	10	32	12	10	34	738	73.8	61.5	1.56	472,000
	1	10	32	12	15	82	5704	570.4	475.33	5.41	1,054,000
Host 30	1	30	32	39	5	90	1551	51.7	39.77	1.32	1,178,000
	1	10	32	12	10	34	738	73.8	61.5	1.56	472,000
	1	10	32	12	15	82	5704	570.4	475.33	5.41	1,054,000
Host 40	1	37	32	39	5	90	1551	41.91	39.77	1.32	1,178,000
	1	10	32	12	10	34	738	73.8	61.5	1.56	472,000
	1	10	32	12	15	82	5672	567.2	472.67	5.46	1,038,000
Host 50	1	37	32	38	5	89	1547	41.81	40.71	1.31	1,178,000
	1	10	32	12	10	34	738	73.8	61.5	1.56	472,000
	1	10	32	12	15	82	5672	567.2	472.67	5.46	1,038,000

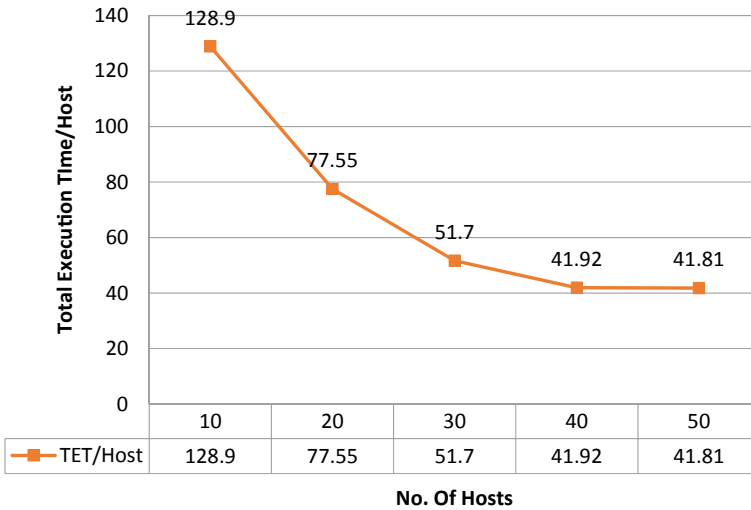


Fig. 1 Total execution time per host at scheduling interval 5

As it is discussed above, the scheduling interval is the time duration after which the load balancer checks the status of the VM’s load. If the Hosts/VMs are idle and not having job, load of that Host/VM will be destroyed, and if the existing Hosts/VMs are overloaded, then new Hosts/VM’s will be created according to the load measurement.

It is also observed by setting values to the key parameters that as shorter as we set the value to scheduling interval will result in the better performance of the system. Because what we are expecting with the system here is the **rapid elasticity** in the dynamic resource scaling. Now, it is point of discussion here that what elasticity’s role here.

So, **elasticity** is the limit to increase or decrease in the number of resources, i.e., Hosts, VMs, PEs, RAM, bandwidth, etc. Then what does the meaning of **rapid** here. So, the **rapidness** is the time interval, after which load is being checked. As less as this time interval, the system is called more rapid, and if this time interval is greater it will be said that the system is less rapid. This is directly proportional to the system performance. So, the **rapid elasticity** in dynamic resource with horizontal scaling refers to measurement of existing Hosts/ VMs load regularly, with a short interval, and accordingly creating new Hosts / VMs and allocating new computing resources dynamically, or deleting existing Hosts/VMs dynamically because that are idle and as now they don’t have job load.

In Fig. 1, the outcome of the facts is displayed well. In this figure, a graph is representing the total execution time (TET) per Host at the scheduling interval 5, and this result is collected by keeping the creation of Host’s limit as (10, 20, 30, 40, and 50). This is the case of horizontal dynamic resource scaling, and results clearly show

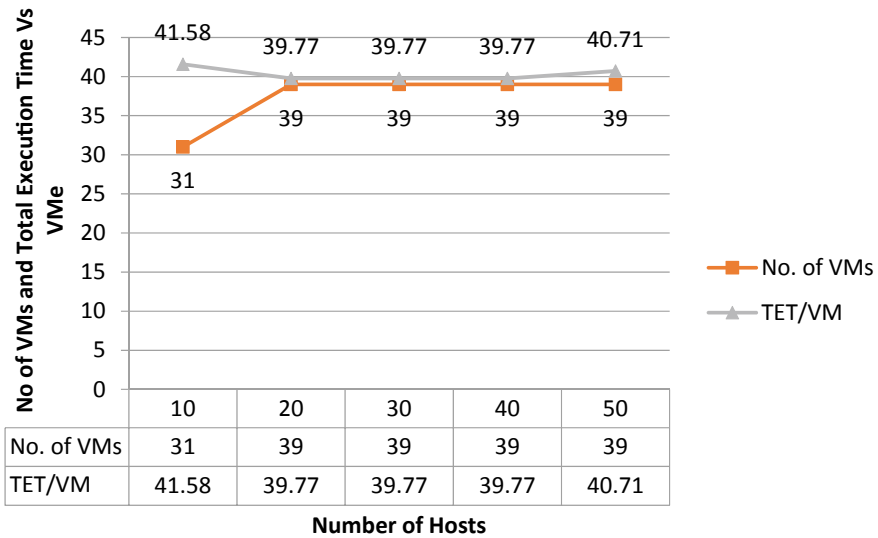


Fig. 2 Total execution time versus number of VMs actually created

that as much as we have greater limit of creating Host, there will be less execution load per Host.

Figure 2 shows that the total execution time (TET) Vs number of VMs actually created, when the scheduling interval is set at 5 s, and these values are calculated when the dynamically Hosts creation limit is set to 10, 20, 30, 40 and 50 Hosts.

It is analyzed that if the scheduling interval is set at the higher side, the number of Hosts is actually not created as much as created when the scheduling interval set to lower side. Because the time after which Hosts/VMs load are measured on the load bases is high, during this time gap of load measurement most of the jobs are finished and remaining jobs are allotted to the already created VMs and hence the need of creating new Hosts/VMs is not required. The total execution time will be more in this case. If the Hosts/VMs load was measured within the specified time and creation of Hosts/VMs was done within specified time the total execution time would be definitely low.

This results that the less number of Hosts is actually created even there is sufficient limit to create Hosts/VMs; this can be easily understood by the results shown in Table 1. So, in this way we may also as say if the system is less rapid, the performance obtained will be low, and if the system is more rapid, then the performance will be obtained high.

$$\text{Dynamic system or rapid factor} \propto \frac{1}{\text{scheduling interval}} \tag{1}$$

$$\text{System Performance} \propto \text{Dynamic system or rapid factor} \tag{2}$$

Equations (1) and (2) show the relative nature of three important terms, i.e., dynamic system or rapid factor, scheduling interval and system performance.

4 Conclusion

Cloud computing definitely proved that it is a good platform for the developers to come with innovative ideas for serving to the organizations, institutions, society by computing and making their complex service structure easy. There are three major contributions of cloud computing: (i) pay-per-use model, (ii) dynamic scalability of resources and (iii) large shared computing environment. Dynamic resource scaling includes horizontal and vertical scaling of resources. Horizontal scaling is more frequent, and in this paper, we have tried to explore the underlying nature and parameters of horizontal dynamic resource scaling. Samples are taken and analyzed at different values of scheduling interval. The dynamic limits of other resources like number of Hosts or number of VMs are set and results are recorded. The outcome of the results analyses is that the scheduling interval should be kept at its lower side rather higher side and dynamic limits of number of Hosts, and number of VMs should be kept at higher side. This will definitely improve the overall system performance, and resources can be utilized efficiently.

References

1. Zhang S, Qian ZZ, Wu J et al (2015) Service-oriented resource allocation in clouds: pursuing flexibility and efficiency. *J Comput Sci Technol* 30(2):421–436
2. Anastasi GF, Coppola M, Dazzi P, Distefano M (2016) QoS Guarantees for network bandwidth in private clouds. In: CLOUD FORWARD: from distributed to complete computing, CF2016, 18–20 October 2016. Madrid, Spain, *Proc Comput Sci* 97:4–13
3. Barba-Jimenez C, Ramirez-Velarde R, Tchernykh A, Rodríguez-Dagnino R, Nolasco-Flores J, Perez-Cazares R (2016) Cloud based video-on-demand service model ensuring quality of service and scalability. *J Netw Comput Appl*
4. Sareen P, Kumar P, Singh TP Resource allocation strategies in cloud computing. *Int J Comput Sci Commun Netw* 5(6):358–365
5. Rajasekar B, Manigandan SK (2015) An efficient resource allocation strategies in cloud computing. *Int J Innov Res Comput Commun Eng* 3(2):1239–1244
6. Alnori A, Djemame K (2018) A holistic resource management for graphics processing units in cloud computing. *Electron Notes Theoret Comput Sci* 340:3–22
7. Gmach D, Rolia J, Cherkasova L, Kemper A (2007) Capacity management and demand prediction for next generation data Centers. In: Published in the international conference on web services (ICWS '2007), 9–13 July 2007. Salt Lake City, Utah
8. Lim H, Babu S, Chase J Automated control for elastic storage. In: Proceedings ICAC '10, June 7–11, 2010. Washington, DC
9. Urgaonkar B, Pacifici MSG, Shenoy PJ, Tantawi AN (2005) An analytical model for multi-tier internet services and its applications. In: Proceedings SIGMETRICS, June 6–10, 2005. Banff, Alberta, Canada, ACM 1-59593-022-1

10. Toosi AN, Son J, Chi Q, Buyya R (2019) ElasticSFC: auto-scaling techniques for elastic service function chaining in network functions virtualization-based clouds. *J Syst Softw*
11. Baruah P, Mohanpurkar A (2015) Impact of elasticity on cloud systems. *Int J Comput Appl* (0975-8887) 120(14):23-28
12. Ghose S (2016) Testing elasticity of cloud platform. *Int J Comput Eng Appl* X(VI):58-66
13. Rehman Z, Hussain OK, Hussain FK, Chang E, Dillon T (2015) User-side QoS forecasting and management of cloud services. Received: 26 August 2014/Revised: 26 November 2014, Springer Science Business Media, New York
14. Singh J, Agarwal S, Mishra J (2015) A review: towards quality of service in cloud computing. *Int J Sci Res (IJSR)*, 555-561. ISSN (Online): 2319-7064 Index Copernicus Value, 78.96 Impact Factor 6.391
15. Ghahramani MH, Zhou MC, Hon CT (2017) Toward cloud computing QoS architecture: analysis of cloud systems and cloud services. *IEEE/CAA J Automatica Sinica* 4(1):6-18
16. Rahamath Nazneen S, Kavitha R (2014) Cloud computing integrated with testing to ensure quality. *Int J Adv Res Comput Sci Technol (IJARCST 2014)*, *IJARCST* 2(1):196-199. ISSN: 2347-8446 (Online), ISSN: 2347-9817 (Print)
17. Kivity A, Kamay Y, Laor D, Lublin U, Liguori A (2007) kvm: the Linux virtual machine monitor. In: *Proceedings of the Ottawa Linux symposium*, vol 1, pp 225-230
18. Choi J, Ahn Y, Kim S, Kim Y, Choi J (2015) VM auto-scaling methods for high throughput computing on hybrid infrastructure. *Clust Comput* 18:1063-1073

Recapitulation of Research in Artificial Intelligence: A Bibliometric Analysis



Utkal Khandelwal  and Trilok Pratap Singh 

Abstract “Artificial intelligence” is the ability to understand and adjust technological advances; computer thinking theory and actions have become a vital task for successful operational advancement in the most modern organizational environment. Artificial intelligence technology provides a significant competitive advantage in the status of society. This article aims to analyze essential artificial intelligence work and discusses the concept of artificial intelligence by incorporating bibliometric research.

Keywords Artificial intelligence · Operational management · Bibliometric Analysis · Technological advances

1 Introduction

The term “artificial intelligence” was coined in 1956 and becomes more popular today. Artificial intelligence is needed to convince most of today’s organizational transformation environments [1]. The primary market activity of firms was implicitly artificial intelligence (equivalent or superior to human thought, action, and technical execution theories). Artificial intelligence is now at the heart of several institutions created and formally published by the scientist John McCarthy at the Dartmouth Conference [2]. Artificial intelligence is primarily concerned with imitating the human brain and making decisions in a variety of situations and conditions, such as humans [3]. Computational power and control, data, and genetic algorithms have been developed and greatly enhance the performance of the processor. Artificial intelligence business is projected to reach \$3.9 million in 2022, up 70% from \$1 billion in 2017 [4]. Artificial intelligence machinery has gained enormous popularity in the agricultural and food industries for proper product classification and selection; rapid; accurate labeling has made this digital era a groundbreaking shift [5]. However, the field of artificial intelligence still needs to be extended to include further analysis and case-based studies to expand the knowledge base further. This research is an attempt

U. Khandelwal (✉) · T. P. Singh
GLA University, Mathura, India
e-mail: utkal.khandelwal@gla.ac.in

to explore the bibliometric picture of artificial intelligence research during the period from 2015 to 2019. In bibliometric analysis, this research presents year-wise publication, most cited papers, most prolific author in the field of artificial intelligence. This study also identified most prolific countries and institutions and the most frequently appearing keywords.

2 Research Methods

Many research areas use bibliometric methods for three purposes: expanding their area of operation, assessing the influence of a research group, or determining the effect of a specific study. Some describe it as an instrument throughout the production of scientific literature that enables the state of scientific and technological progress to be examined at a particular level. It's a way to locate a nation with a world an organization with a community, and even individual researchers with their societies [6]. This research work is an attempt to examine the dispersion of artificial intelligence papers, including the time of publication, author's nations, publications, reviews, and numbers of citations. In addition, keyword analysis also has been done.

This study examines the artificial intelligence research papers listed in the Scopus Index journals as the credible sources for the database over the 05-year period (2015–2019) to achieve the objective. Data were collected from the Scopus database for this document. Paper details from the selected database have been extracted on April 16, 2020 covering the period from 2015 to 2019, comprising 35,851 documents. The research request composed of all papers in their name, description or keywords with the term “Artificial Intelligence.” Each registration document included the year of publication, the period of publication, the name(s) of the author, the definition of the source and keywords, the outline and the references. In spite of the fact that this investigation was not comprehensive, we trust it will contribute as an intensive reason for an elucidation of artificial intelligence developing patterns.

3 Results

3.1 *Classification of the Publication Year*

Figure 1 shows the spread of yearly distributing yield from 2000 to 2019. The year of publication frequency Fig. 1 shows the sequential movement of the amount of artificial intelligence papers distributed.

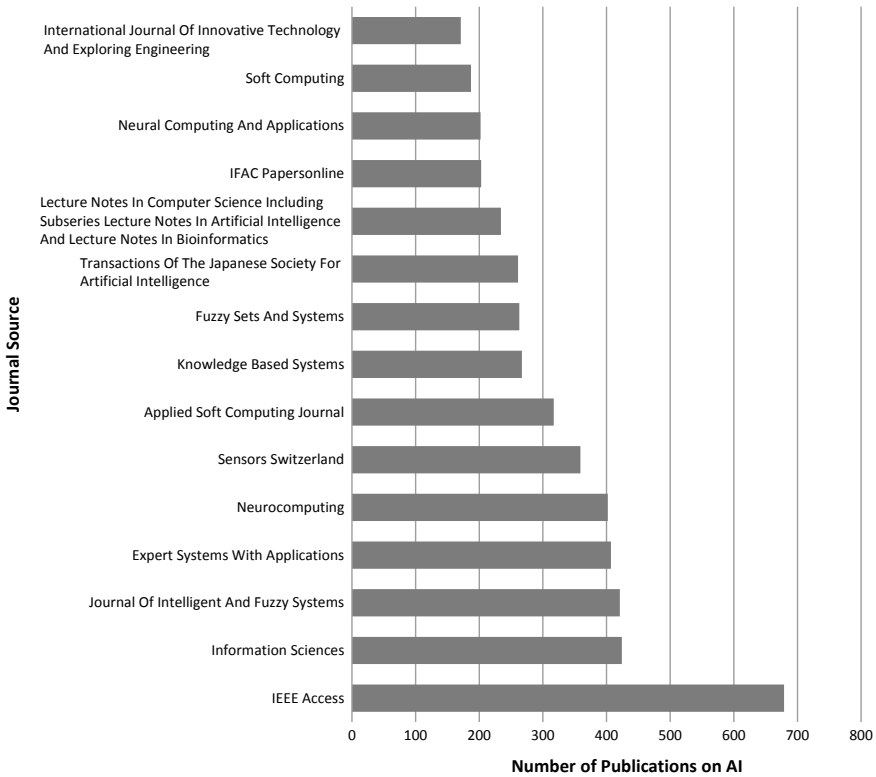


Fig. 1 Top 15 journals with maximum publication on artificial intelligence

3.2 Last Five-Year Scopus Publication on Artificial Intelligence

Figure 2 indicates the allocation of artificial intelligence publications by years from 2015–2019, published by all researchers in different discipline. Based on the associations with the discipline, from 2015 to 2019, the analysis indicates that the maximum number of publication on m-banking publications is in the year 2019.

3.3 Artificial Intelligence Cited Times Report

Citation rates indicate the effect on academics and other fields of artificial intelligence. The artificial intelligence articles, widely cited in other journals, are given in Table 1. The research paper which was most widely cited was “Mastering the game of Go without human knowledge “which was published in 2017, and this paper was

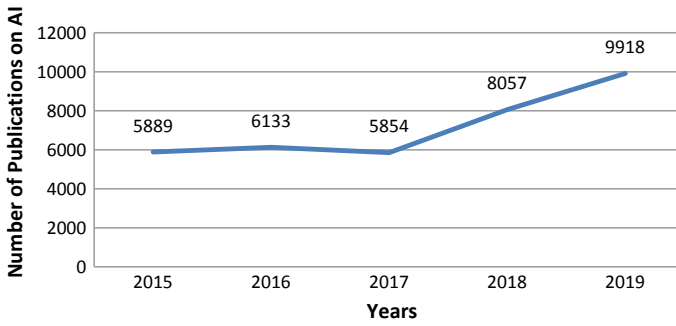


Fig. 2 Last five-year Scopus publication on artificial intelligence

Table 1 Most cited papers

Authors	Title	Year	Cited by
Silver et al.	“Mastering the game of Go without human knowledge”	2017	1381
Greff et al.	“LSTM: A Search Space Odyssey”	2017	813
Sze et al.	“Efficient Processing of Deep Neural Networks: A Tutorial and Survey”	2017	399
Kermany et al.	“Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning”	2018	347
Van et al.	“Computational radiomics system to decode the radiographic phenotype”	2017	341
Butler et al.	“Machine learning for molecular and materials science”	2018	299
Topol et al.	“High-performance medicine: the convergence of human and artificial intelligence”	2019	280
De et al.	“Clinically applicable deep learning for diagnosis and referral in retinal disease”	2018	273
Davies et al.	“Loihi: A Neuromorphic Manycore Processor with On-Chip Learning”	2018	265
Hua et al.	“Skin-inspired highly stretchable and conformable matrix networks for multifunctional sensing”	2018	260

cited 1381 time in the last three years; the authors of this research paper are Silver et al. [7].

3.4 Distribution of Artificial Intelligence Publications by Author

Figure 3 shows that some well-known authors have published artificial intelligence research papers. Artificial intelligence concept attracted many well-known authors,

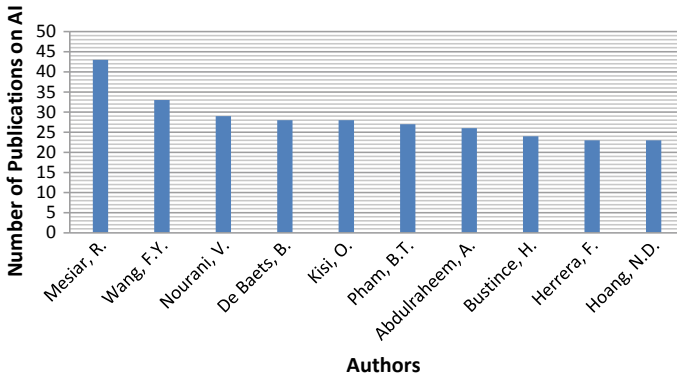


Fig. 3 Most prolific authors

who published their papers in various journals (see Fig. 3). The artificial intelligence research is carried out by the researchers in variety of disciplines. Figure 3 shows the contribution of various authors on the topic artificial intelligence.

3.5 *Distribution of Artificial Intelligence Publications by Institutions*

Figure 4 shows that some well-known institutions have published artificial intelligence research papers. Artificial intelligence concept attracted many well-known institutions, who published their papers in various journals (see Fig. 4).

3.6 *Artificial Intelligence Publications Distribution by Countries*

Figure 5 indicates the allocation of the 35,851 artificial intelligence papers by researchers’ nations. Based on associations with the researcher, from 2015 to 2019, the analysis listed nations generating the most one—artificial intelligence publications.

3.7 *Artificial Intelligence Important Keywords*

Keywords’ occurrence indicates the effect on academics and other fields of artificial intelligence. The artificial intelligence articles, widely searched in other journals,

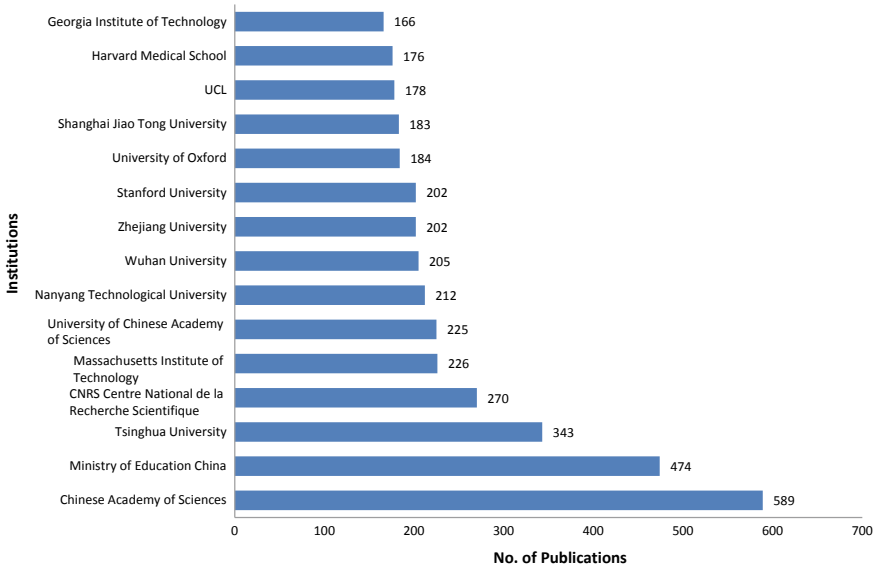


Fig. 4 Most prolific institutions

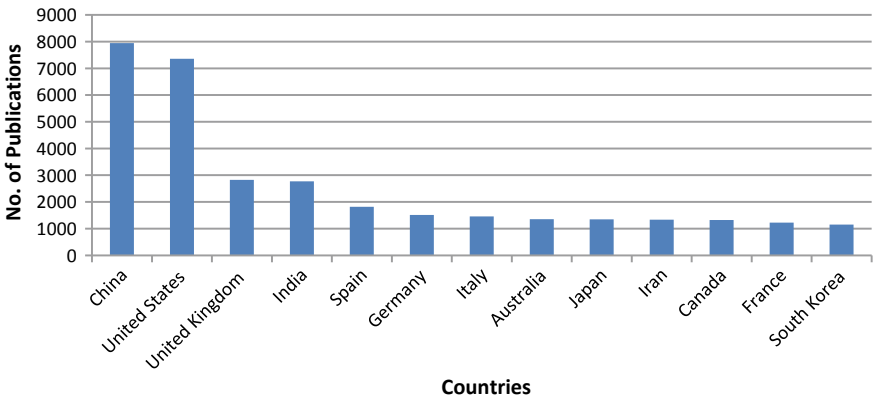


Fig. 5 Most prolific countries who have published more than 1000 papers on AI in last five years

are given in Table 2. The keyword which was most widely searched was artificial intelligence which was searched 27,986 times in the last five years.

Table 2 Top 50 keywords' occurrence in last five-year publications of AI

#	Keyword	Occurrence	#	Keyword	Occurrence
1	Artificial intelligence	27,986	26	Female	1325
2	Learning systems	9562	27	Genetic algorithms	1316
3	Machine learning	7245	28	Prediction	1228
4	Human	5391	29	Decision support system	1214
5	Article	4355	30	Big data	1210
6	Humans	3692	31	Male	1207
7	Learning algorithms	3573	32	Controlled study	1168
8	Algorithms	3513	33	Classification	1121
9	Optimization	2999	34	Software engineering	1075
10	Algorithm	2846	35	Adult	1063
11	Decision support systems	2843	36	Image processing	1061
12	Neural networks	2809	37	Review	1038
13	Forecasting	2533	38	Automation	1027
14	Priority Journal	2409	39	Machine learning techniques	1024
15	Decision making	2239	40	Regression analysis	1024
16	Classification (of information)	2216	41	Particle swarm optimization (PSO)	906
17	Data mining	2144	42	Pattern recognition	844
18	Procedures	2141	43	Diagnosis	794
19	Artificial neural network	2099	44	Robotics	791
20	Deep learning	1799	45	Evolutionary algorithms	782
21	Support vector machines	1645	46	Semantics	771
22	Decision trees	1533	47	Sensitivity	765
23	Optimization	1449	48	Fuzzy logic	763
24	Support vector machine	1424	49	Swarm intelligence	762
25	Feature extraction	1368	50	Supervised learning	747

4 Basic Functionalities of Artificial Intelligence

4.1 Transformations

It might be easier to say how important it is to our everyday lives, business activities, and culture, what kind of modern civilization has not influenced artificial intelligence

(AI). Intelligence machines have an impact on virtually every aspect of our lives in order to optimize productivity and improve human power. AI is intertwined with everything we do; life is hard to imagine without it [8]. AI is the central tenet of transformative transformations in the fourth industrial transformation. It is a transformation that will test our theories about what society means and could be broader than any other technological change before us. Computers can understand and make similar human decisions with artificial intelligence [9]. There are many examples of artificial intelligence, such as machine learning, in which machines can perceive, interpret, and learn more from data and error than they can from our human brains. This technology has had an impact on consumer products and has made significant advances in health and engineering, as well as in various industries, including manufacturing, banking, and retail [10]. The vast amount of data generated by us on a daily basis and the available computing resources have led to the explosion of artificial intelligence.

4.2 AI in Different Business Dimensions

Technology is shifting from finance to manufacturing, as market leaders and innovators strive to deliver on the promise of artificial intelligence to gain competitive advantage and cost and time savings. Companies like Heineken, 150 years old, use AI, the vast amount of data they collect, and the Internet of Things to direct business decisions and campaigns, enhance customer interaction and events [11]. Artificial intelligence enables organizations across all sizes and across all industries to improve effectiveness and, ultimately, from goods production to sales and customer support accounts at every stage of the company's life cycle, from controlling global supply chains to maximizing delivery. It helps businesses create, manufacture, and deliver products and services more effectively than ever before [12].

AI is turning our healthcare system from its operations to those provided by these organizations, by identifying customized doses of drugs to better diagnostic methods, and even robots to assist in surgery. In addition to our healthcare systems, AI is potentially very involved in dealing with the environmental problems we face as a result of global warming [13]. AI will increase human capital to address our global challenges by promoting smart infrastructure to develop our natural resources and intelligent agriculture to feed the growing population. This is also an integral part of our information and security networks. Experiments in AI compose original pieces from books to music, make recipes based on what's in the cupboard, and even work of art expands the influence of AI on our lives. While the latest innovations and improvements contribute to the extended and refined implementation of AI, it is expected that this technology will be intertwined in our daily lives, workplaces, and society [14].

5 Conclusion and Future Direction

This study presents the main trends in artificial intelligence publications in the bibliometric analysis. The work is based on a bibliometric review. Scopus is one of the most widely used and widely accepted research databases [15]. Our findings show some of the implications for marketing managers in the design of artificial intelligence services. Finally, the article contributes to significant research on artificial intelligence knowledge. This article deals with papers published over the last five years due to rapid technological growth. Scopus search articles are making inferences. There is a good reason for scientists, industry, administrators, policymakers, and other relevant groups to use this information for further research and development in AI. The Fourth Industrial Revolution and the Sustainability Challenge present new global challenges and opportunities for rapid technological advancement [16]. Industry 4.0 technologies, like comprehensive data, are essential for productive operations [17]. Artificial intelligence is not rare, especially when used in manufacturing. However, artificial intelligence has a wide potential to control production processes [18]. Something wise is making sound choices. Machine learning involves analyzing and summarizing real-world data through a mathematical decision-making process. That is artificial intelligence, thinking you can do a smart thing. Artificial intelligence uses a variety of approaches, from simple methods of data analysis to advanced technologies such as deep learning. The analysis of data is artificial intelligence, mathematically representing the best decisions. This work has provided new ideas for potential researchers.

Artificial intelligence (AI) is a ground-breaking technology that allows us to rethink how knowledge is processed, interpreted, and used to enhance decision making. Every area of existence is already changing. Therefore, further studies include analyzing practical issues in product development and proposing models designed to incorporate artificial intelligence into the operating system to solve problems. Future research would also address critical issues related to artificial intelligence for different industries and sectors. Data privacy and security concerns are a significant concern when implementing artificial intelligence, and future researchers need to focus on preventing project failure.

This study reveals some findings that can help guide researchers in the field of artificial intelligence. The most important limitation of this is that particularly only the Scopus search results served as a source for the dataset used in this study. Therefore, in the future studies, data can be collected from all such sources too, and the analyzed results can be compared with the results of this study. Future researchers may also this kind of literature review with different functional areas to present the practical applications of AI. In conclusion, it is emphasized that this research endeavor illuminates the AI literature in a new fashion. It is expected that this research will act as a plentiful source of information for the researcher community interested in conducting future studies in this area.

References

1. Aldasoro U, Merino M, Pérez G (2019) Time consistent expected mean-variance in multistage stochastic quadratic optimization: a model and a matheuristic. *Ann Oper Res* 280(1–2):151–187
2. McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine* 27(4):12–12
3. Kumar S (2019) Artificial intelligence divulges effective tactics of top management institutes of India. *Benchmark Int J* 26(7):2188–2204
4. Richards G, Yeoh W, Chong AYL, Popovic A (2019) Business intelligence effectiveness and corporate performance management: an empirical analysis. *J Comput Inform Syst* 59(2):188–196
5. Wauters T, Kinable J, Smet P, Vancroonenburg W, Berghe GV, Verstichel J (2016) The multi-mode resource-constrained multi-project scheduling problem. *J Sched* 19(3):271–283
6. Vinkler P (2010) Indicators are the essence of scientometrics and bibliometrics. *Scientometrics* 85(3):861–866
7. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A et al (2017) Mastering the game of go without human knowledge. *Nature* 550(7676):354–359
8. Scherer MU (2015) Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. *Harv JL Tech* 29:353
9. Paschen J, Pitt L, Kietzmann JH (2019) Emerging technologies and value creation in business and industrial marketing. *J Bus Indus Market* 34(2)
10. Ghahramani Z (2015) Probabilistic machine learning and artificial intelligence. *Nature* 521(7553):452–459
11. Davis E, Gary M (2015) Commonsense reasoning and commonsense knowledge in artificial intelligence. *Commun ACM* 58(9):92–103
12. Chau KW (2006) A review on integration of artificial intelligence into water quality modelling. *Mar Pollut Bull* 52(7):726–733
13. Hanson CW, Marshall BE (2001) Artificial intelligence applications in the intensive care unit. *Crit Care Med* 29(2):427–435
14. Gonzalez LF, Montes GA, Puig E, Johnson S, Mengersen K, Gaston KJ (2016) Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors* 16(1):97
15. Van Nunen E, Verhaegh J, Silvas E, Semsar-Kazerooni E, van de Wouw N (2017) Robust model predictive cooperative adaptive cruise control subject to V2V impairments. In: 2017 IEEE 20th international conference on intelligent transportation systems (ITSC). IEEE, pp 1–8
16. Telukdarie A, Buhulaiga E, Bag S, Gupta S, Luo Z (2018) Industry 4.0 implementation for multinationals. *Process Saf Environ Prot* 118:316–329
17. Waltman L, Van Eck NJ, Noyons EC (2010) A unified approach to mapping and clustering of bibliometric networks. *J Informetr* 4(4):629–663
18. Dwivedi YK, Rana NP, Jeyaraj A, Clement M, Williams MD (2019) Re-examining the unified theory of acceptance and use of technology (UTAUT): towards a revised theoretical model. *Inform Syst Frontiers* 21(3):719–734

A Study on Identification of Issues and Challenges Encountered in Software Testing



Omdev Dahiya  and Kamna Solanki 

Abstract Comprehensive studies play an essential role in the field of engineering and technology. They aim to make us understand why and how things work. This enables budding practitioners and researchers to use this understanding to bring innovations that will shape the modern world. Software testing plays a fundamental role in the development of the software. Software is subjected to testing to discover faults so that necessary amendments can be made in it. This overall increases its quality, reliability, and robustness. Researchers have even used many techniques in an integrated form to increase the effectiveness of testing. This study aims to discuss various issues that arise during the software testing process. Addressing these issues will result in the development of a fault-free, reliable, and robust software product. For this process, this study has laid points that should be taken care of so that delivery of the right quality product to its intended users is ensured. The objective of this study is to bring the attention of emerging practitioners and testers toward the areas where there is a need for inspection and investigation.

Keywords Software Testing · Software Quality · Issues and Challenges

1 Introduction

An empirical study is an investigation and a test to compare what we believe with the one which we have observed. In software engineering research, it is deduced that the most prominent challenges were not in planning and leading individual investigations. It lies in conceptualizing and sorting out a collection of work that could be depended on as the reason for changing an organization's practices followed by it from a prolonged period for the development process [1, 2]. The testing of software

O. Dahiya (✉) · K. Solanki
University Institute of Engineering and Technology, Maharshi Dayanand University Rohtak,
Rohtak, India
e-mail: Omdahiya21792@gmail.com

K. Solanki
e-mail: Kamna.mdurohtak@gmail.com

is concerned with the development of good-quality software systems. It helps in ensuring whether the developed or developing product matches the outcome, which is expected from it [3–5]. In the present study, we have identified a set of issues and challenges that the practitioners and researchers have to face in the testing field. For this, various published research works and available online materials in the form of blogs and articles have been searched. The motive behind this is to bring the attention of budding practitioners and researchers toward the areas where there is a need for inspection and investigation. The organization of the paper is as follows. Section 2 presents the background of the software testing methodology. Section 3 provides the identified issues and challenges encountered in software testing. Section 4 presents the conclusion and future scope.

2 Background

In the development process of a software product, its reliability can only be ensured by implementing it through sufficient testing. Despite this fact, software testing is termed as a time consuming and expensive process [6, 7]. There are many techniques available for testing the software, but which technique should be applied so that all factors are balanced, remains an area of research always. Researchers keep on developing several techniques to improvise the testing process [8, 9]. Practitioners and researchers keep on working to find an optimal technique so that they may answer “which testing technique to use and when to use it?”

In the study performed by Belay [10], the author has worked on finding out the challenges and addressing the issues while developing the industry level software projects. The study has shown how the quality and reliability of the software are affected by the number of issues. A systematic literature review is conducted to find out the worthwhile research published, and they are thus analyzed and documented accordingly. It is concluded how testing tools, organizational structure, methods, artificial intelligence, machine learning, security, and maintenance of the software project matters a lot in developing these vast systems for industry.

In the study conducted by Klotins et al. [11], the authors have shown how the startups are contributing to the development of software and bringing innovations in this field. They have shown which practices of software engineering are followed. They discussed that due to the improper implementation of testing techniques, many startups fail to deliver a quality product. They have concluded that this study may be used by the researchers to further explore this area by learning from others’ experiences.

In the study conducted by Szopinski et al. [12], the authors have presented a brief description of the software tools to be used for different perspectives and environments. They said that there is a need for a study where there is a collection of different software tools along with their guide for usage and which tools can be used for which purpose. They have addressed how the quality of the software is ensured by rigorously testing it via different tools and techniques.

A study was conducted by Mansor et al. [13], in which the authors have performed an analysis of the existing researches and documents to find out the significant issues and challenges in testing the software. They have discussed the solutions to the general issues and stressed that addressing these will help in developing a framework for testing the software, which will eventually increase its quality.

In the study conducted by Ghanam et al. [14], various challenges which can be met by a software developing organization, while transitioning on the different platforms of software were discussed. They have tried to address almost all the identified issues from the data collected by them. They have concluded that this study will provide useful insight to the practitioners and researchers in making decisions and setting the expectations which are close to ground reality.

Researchers have integrated different techniques to get better results. Still, many issues and challenges are faced by the testers during the testing phase. This study has documented some of them so that upcoming practitioners and researchers may get an insight into the possible hurdles they may face, and a road map for tackling them can be developed. Below, Fig. 1 provides an overall layout of the testing process.

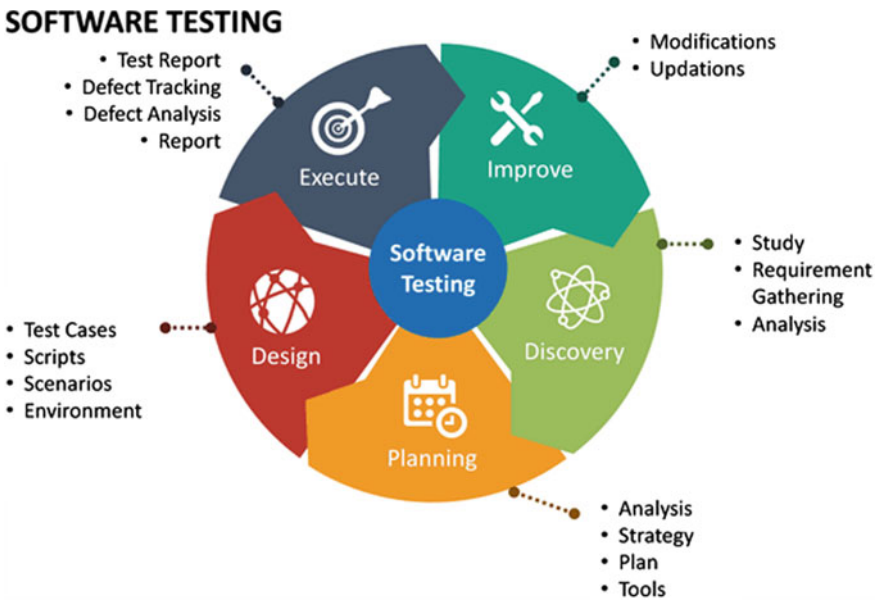


Fig. 1 Picture depicting an overall layout of the testing process. *Img. Source* <https://www.sketchbubble.com/en/presentation-software-testing.html> [15]

3 Issues and Challenges

This section discusses many issues and challenges faced by the testers during the testing phase of the software. After discussing the problem, its respective solution is also provided. Following points provide a useful insight into them.

3.1 *Communication Issues*

Problem During the development process of software, issues in communication are persistent. It may either between teams working on the project or between clients and project managers. There are many reasons for this, such as time zones are different for developers and clients or personnel engaged on a project are working in different shifts and many more. This gap in communication can severely affect the quality of a product [16].

Solution During the development phase, there should be complete collaboration in the development, testing, and implementation phase. A well-defined and transparent process should be laid down for that. When these steps are followed in an integrated manner, the clients will get on-time delivery and within the specified budget.

3.2 *Testing Issues Are Not Addressed Thoroughly*

Problem There are a million lines of codes in the software developed nowadays. For assuring the delivered product to be fault free and reliable, each section of code must be subjected through a comprehensive testing process. However, due to a surge in demand for delivering a software product in a shorter period, the developing organization faces difficulty in executing the whole test cycle. There is excessive pressure on the developers and testers from the project managers and clients. This is another reason for not executing a complete test cycle [17].

Solution The developing organizations should make a list according to the priority of the requirements. Due to this, it will be easy to identify those portions which need testing first. This will ensure that the essential portions of the software are executed and verified first before approaching to next stages.

3.3 *Improper Documentation*

Problem Many organizations make changes in a developing software product just based on verbal communication. A record of non-functional and functional scopes

of a product is not adequately documented. Many vital pieces of information can be missed in case of verbal communication as testers are not able to gather requirements accurately and may face problems later [18].

Solution A comprehensive requirement analysis is required to be documented. This will facilitate the developing and testing staff in modeling and shaping the product according to the need. The testers will be able to generate required test cases by analyzing the specifications and objectives. It will aid them in minimizing potential defects.

3.4 Environment for Testing is not Stable

Problem Software development at times is done at a fast pace, and sometimes specifications of requirements are changed in this cycle of developing and delivering the software. To add new functions or to fix the issues identified while testing the software, developers sometimes change the environment of the test. It becomes quite cumbersome to keep the record of changes made in the product by an individual member when there are many testers engaged in the process of testing a single product. The development process is disrupted when the quality assurance testing team does not have an update of the changes made in the system. It is quite challenging to perform testing of a product whose information is not complete [19].

Solution A controlled and organized testing environment is required to be created by the project managers and team members so that the delivery of the right quality product is ensured. Proper documentation of the changes made in the specifications of the desired product will be beneficial to the testing team. It assists the testing and quality assurance team to work in a coordinated manner as they do not have to repeat the process again and again to gather requirements for testing. Besides this, there are numerous challenges also which are required to be tackled to ensure the delivery of a good-quality product. Some of them are skilled testers who are not available, complexity in an application, and inability to set a proper testing environment.

3.5 Testing is not Adequately Scheduled

Problem Testing consumes much time. Still, it is mandatory to be performed to find discrepancies in the developed system. Testing is required to be done systematically, along with the development phase. This will help the developer to make desired changes to remove the identified errors. Generally, testing is performed at the later stages of the development process. Due to this, there is little or no room left to perform final testing, and an inadequate process is scheduled [20].

Solution Managers must schedule the testing and development process in such a manner that both runs simultaneously. This will lead the testers to borrow enough time to look at the vulnerabilities and systematic adequacies comprehensively.

3.6 Tools and Environment for Testing Are not Sufficient

Problem For performing proper software testing, tools and the environment play a vital role. However, testing is performed manually in an adequate environment. There is a possibility of inadvertent errors in the environmental components. There is a lack of adequate control over the software to be tested, test data, and the environment for testing [21].

Solution To identify errors that the potential clients may face at their end, test managers must set the testing environment in such a way that it resembles the environment close enough to be of the clients. To execute testing in a fast manner, test managers may deploy automated tools for testing so that resources can be used judiciously, and more errors can be found out in the developed system.

3.7 Testing Mindset is Wrong

Problem Instead of finding the faults in the systems, testing team members often revolves around exploring its functionality. Due to this, testers are distracted from finding faults in the software. The same type of problems is repeated in projects after projects [22].

Solution Testers should work on uncovering faults in the system by executing it in a different environment and conditions, instead of just that it is working. The management team must ensure that there is proper and careful documentation of every lesson learned in the earlier executed products. They should be implemented in the upcoming projects.

3.8 Engineering and Testing Processes Are Poorly Integrated

Problem There should be proper integration in the engineering and testing processes. If they are not, then proper testing cannot be ensured. Software modules are tested for errors even before they are mature enough to be tested using all parameters [23].

Solution There is a requirement to look into the specific needs of the projects. “One size fit all” formula is not applicable in all the spheres of the testing. It is need of the hour where all the teams should work in close integration with each other.

4 Conclusion and Future Scope

Software products today are increasingly unpredictable to deal with. The persistently changing requirement specification and fast advancement cycle assume a significant job in making a problematic test condition. From social occasion fundamental venture components to doing mechanized experiments, everything should be smoothed out for fruitful testing results.

Toward this, a dependable quality-oriented programming testing organization, including experts, engineers, analyzers, and task administrators, brings the entire application advancement process in one spot. Therefore, it gets conceivable to boost profitability and discharge the ideal form of the software product. It is best said that unlike a single person, being a team is smarter, being a group is more astute. In this way, the best exhortation is to get together and fly high. Huge endeavor and customers ought not to invest an excess of energy representing blunders which come up. Following a decent and methodical practice for your quality assurance needs will give quality to your group. To overcome the limitations in the existing studies, one way is to move from small scale studies to projects which have a broad scope. It involves eminent experts and researchers in the testing field. Comparing other studies of this field, which are based upon well-defined contexts, leads to better valid results.

References

1. Perry DE, Porter AA, Votta LG (2000) Empirical studies of software engineering: a roadmap. In: Proceedings of the conference on the future of software engineering, pp 345–355
2. Dahiya O, Solanki K (2018) A systematic literature study of regression test case prioritization approaches. *Int J Eng Technol* 7(4):2184–2191
3. Alrmuny DZ (2014) Open problems in software test coverage. *Lect Notes Softw Eng* 2(1):121
4. Briand L, Labiche Y (2005) Empirical studies of software testing techniques: Challenges, practical strategies, and future research. *ACM SIGSOFT Softw Eng Notes* 29(5):1–3
5. Briand L (2007) A critical analysis of empirical research in software testing. In: First international symposium on empirical software engineering and measurement
6. Khatibsyarbini M, Isa MA, Jawawi DN, Tumeng R (2018) Test case prioritization approaches in regression testing: a systematic literature review. *Inf Softw Technol* 93:74–93
7. Yaseen M, Ibrahim N, Mustapha A (2019) Requirements prioritization and using iteration model for successful implementation of requirements. *Int J Adv Comput Sci Appl* 10(1):121–127
8. Jarzabek S, Liszewski K, Boldak C (2020) Inferring hints for defect fixing order from requirements-to-test-case mappings. In: Integrating research and practice in software engineering. Springer, Cham, pp 43–51

9. Nayak S, Kumar C, Tripathi S, Majumdar N (2020) An improved approach to enhance the test case prioritization efficiency. In: Proceedings of ICETIT 2019. Springer, Cham, pp 1119–1128
10. Belay E (2020) Challenges of large-scale software testing and the role of quality characteristics: empirical study
11. Klotins E, Unterkalmsteiner M, Gorschek T (2019) Software engineering in start-up companies: an analysis of 88 experience reports. *Empirical Softw Eng* 24(1):68–102
12. Szopinski D, Schoormann T, John T, Knackstedt R, Kundisch D (2019) Software tools for business model innovation: current state and future challenges. *electronic markets*, pp 1–26.
13. Mansor Z., Ndudi EE (2015) issues, challenges and best practices of software testing activity. In: Proceedings of 14th Conference on Applications Computer Engineering (ACE15), South Korea, pp 42–47
14. Ghanam Y, Maurer F, Abrahamsson P (2012) Making the leap to a software platform strategy: Issues and challenges. *Inf Softw Technol* 54(9):968–984
15. <https://www.sketchbubble.com/en/presentation-software-testing.html#mz-expanded-view-1425308818354>.
16. <https://tweakyourbiz.com/business/software/qa-testing-2>.
17. <https://www.testbytes.net/blog/software-testing-issues/>.
18. https://insights.sei.cmu.edu/sei_blog/2013/04/common-testing-problems-pitfalls-to-prevent-and-mitigate.html.
19. Boland PJ (2002) Challenges in software reliability and testing. In: Third international conference on mathematical methods in reliability methodology and practice
20. Baresi L, Di Nitto E, Ghezzi C (2006) Toward open-world software: issues and challenges. *Computer* 39(10):36–43
21. Dahiya O, Solanki K (2019) Comprehensive cognizance of regression test case prioritization techniques. *Int J Emerging Trends Eng Res* 7(11):638–646
22. Bertolino A (2007) Software testing research: Achievements, challenges, dreams. In: Future of software engineering (FOSE'07). IEEE, pp 85–103
23. Deak A, Stålhane T, Sindre G (2016) Challenges and strategies for motivating software testing personnel. *Inf Softw Technol* 73:1–15

A Detailed Survey Study on Classification and Various Attributes of Fake News on Social Media



G. Sajini and Jagadish S. Kallimani

Abstract Due to the surge in the usage of social media in the recent past, communication between people has undergone a great change. Users interact with each other by sharing a lot of information about the ongoing trends. But most of the information shared recently is misleading with the spread of false news which is known as fake news. A huge volume of such fake news being spread can lead to many complications. The research field has been concentrating on the inception, spread and consequences more nowadays. Detecting the truthfulness of the news is of great concern. It can face many technical difficulties on many grounds. Usage of online tools has made the generation of content easy and is expanded fast, which can lead to a huge amount of data for analyzing. The online content is divergent, which deals with numerous fields, contributing to the job complication. Only computers are not able to evaluate the truthfulness and purpose; therefore, it is dependent on the person–computer interaction. Sometimes, the information that might be termed as fake by a specialist in the field might delude people. Such availability is in restricted quantity but could be a base for the mutual endeavor. Here, a broad summary of the discoveries pertaining to the news which are false is given. The unfavorable effects of ongoing work on methods to detect such news are demonstrated. The readily available datasets are studied which are used to classify fake news. An assuring solution is recommended to analyze online fake news.

Keywords Social media · Online Fake News · Fake News Detection · Target-based · Social context based · News Content-based · Creator-based Fake News

G. Sajini

Department of Computer Science and Engineering, M S Ramaiah Institute of Technology,
Bangalore, India

e-mail: sajini.narayana@gmail.com

J. S. Kallimani (✉)

Department of Computer Science and Engineering, Visvesvaraya Technological University,
Belagavi, Karnataka, India

e-mail: jagadish.k@msrit.edu

1 Introduction

Due to the invention of Internet, there is a tremendous change in the manner in which human communication is carried out. Social media apps are helpful in reaching out to people throughout the world. Because of less money, easy handling, sudden increase, it is a big place for communicating and transfer of data. With respect to volume, velocity and veracity, we can characterize the fake news. Usually, fake news are in big volume, real-time velocity and uncertain veracity. Almost 67% US grown-ups acquire news from social media, and the count is increasing at a large rate.

But because of how popular social media is growing, the net is a proper place to distribute online fake news, like content which misleads, false ratings, fake commercials, gossips, false dialogues made by politicians, etc. The popularity of online fake news is not more from media but is more in social media. In the industrial and academic sector, there is extreme ambiguity because of the facts which are partial. For shopping online, and other ventures, it is affecting heavily.

1.1 Importance of Detecting Fake News

The online fake news is irritating, misleading, intrusion by spreading everywhere. On communities, there is a huge effect. Features of fake news are shown below:

The volume of Fake News

Since to cross-verify, there is no method, anyone can create fake news. Many sites on the web are developed explicitly to generate fake news [1]. These are usually looking like legal sites. Just to generate false claims, improper words, usually for selfish reasons like money and power. Huge amounts of data are all over despite notice of the users.

The Variety of Fake News

In human interaction, major modifications are seen impacting the thinking of home sapiens. Teenagers and the elderly are brainwashed easily because of exponential growth [2]. Mayhem can be caused due to the incidents happening around the globe if the false news is spread a lot. In our day-to-day lives, online fake news creates a big keen extreme deep acute detail.

The Velocity of Fake News

Fake news generators live for less time span. Many fake news concentrate on the ongoing affairs so that the spread is easy. There are complexities to check the number of people related to the current content, the tremendous repercussions. Putting an end to it is extremely grueling and punishing.

2 Fake News Characterization

The broad area is reached in this day and age because of online fake news, and it is obnoxious. Highlighting server varieties of online fake news not leaving the originality, fake ratings and commercials are given [3]. Few essential features of online fake news are Table 1:

Fake news can be viewed as four main types of elements. The elements are discussed below:

Creator: Those who generate fake news are may be people or computers. There might be spelling mistakes, purposefully, done with bad intentions by hackers [2].

Target Victims: These are those who are targeted by the online fake news. These are people who make use of social media and the rest of the environments. They can be academicians, guardians, old, etc.

News Content: It is the main part of the news having things like videos, audios, gifs, animations, and many things.

Social Context: This shows the way in which the news is spread throughout. It contains users analyzing the users and pictures.

Table 1 Various categories of fake news

News content	Non-physical content	Main purpose
		Sentiment
		News topics
	Physical content	Image/Video
		Body text
		Headlines
Social context	Platform	Social media
		Main streaming
	Distribution	Broadcast pattern
		Community of users
Target victims	Platform	Online users
		Main streaming users
	Potential risk analysis	Role-based analysis
		Temporal-based analysis
Creator/Spreader	Real human	Fake news creator
		Benign author & publisher
	Non-human	Cyborg
		Social bots

Many of the fake news available over Facebook by various users can be mapped to the above discussed four types of elements [4].

2.1 Fake News Generators and Publishers

The person who generated and the reason for their idea to create fake news needs to be known, understood, found out and must be explored. Generators can be either people or non-living beings.

Non-living being: Many robots exist to produce fake news. Using different algorithms for showing features like people and generating on their own not making use of people [5]. Few legal data distributing bots play a significant part. Tweeting can be done by cyborg and take part in the activities. Same to same because of the fake news, there can be many issues that might arise like trustworthiness.

People: People are very important to start point of fake news. The non-living things are the transformers of fake news online. Automatically, it is trained to generate and unfurl the news. If it is by a human or computer, irrespective of that, a group creates fake news and rolls out to reach and expand to the maximum number of people. Detecting whether the new is false or real is a really tedious mundane task [6]. Even though we are not the generators, several legit users unknowingly unintentionally become the carrier of such news. Their close ones think it is real news and share it further which spread like wildfire within less time. Since there is anonymity, nobody takes the blame for what is shared, replied, pasted.

2.2 News Content

Every news contains two contents, physical and non-physical. They are described below one by one.

Physical News Content: This shows the headlines, vital part and many components likes' pictures and multimedia. Strong channel to convey information is Facebook, Twitter, etc. Since it is a happening event, it happens at a high speed. URL, #, smileys, multimedia, pictures are part of this class which are the significant factors.

Non-physical News Content: The first-class passes on the content and the second class are suggestions, viewpoints, judgments, advice, feelings, which the generators are in desire to formulate. There are various types, classifications, varieties, kinds of fake news like unreal ratings, unreal commercials, unreal content in the field of politics, etc.

Flipkart, Myntra, Snapdeal, and many more e-commerce sites are being assessed, rated, appraised, and criticized day to day [7]. If there is partiality in such things, it can give a substandard inferior poor unpleasant terrible name to these sites. This will

cause an effect in the way the users think, manipulate their selection choice options and will ruin shatter wreck tear down a company's honor fame.

On the same lines to unreal ratings, unreal and false commercials are posted to deceive the consumers by giving ads which are deluding fooling misguiding content. They two are treacherous and demolish the trust of these sites.

3 Fake News Detection—Practical-Based Approaches

Fake news detection entails the mission to check how true the news is. Resources in detecting fake news are classified; they are of two types: practicality method and current research approach. Here, major discussion is on the former online fake news detection.

3.1 Online Fact-Checking Resources

Media organizations commonly perform fact-checking resources. As ongoing contents are usually mixed of content, 2 class will sometimes tell the whole issue completely. Currently, a lot of performance factors are employed in determining the truthfulness of the content.

Verifying fact is a good way to identify fake news, informing people the trueness, wrong, a combination of the two. In the following, evaluation and comparison of a few famous verifying facts online tools are carried out [8].

Classifynews shows an environment for online to identify fake news with a mindset of building a method for determining trueness for a news item. Machine learning algorithms are used to identify this, basing solely on the textual content labelled news articles are collected from *Open Sources* and everyday training is implemented to these correct, incorrect page examples. To determine, two methods exist, namely "Context only" and "Content only" using two classifiers of machine learning.

FackCheck.org is created with the aim of reducing the amount to fool, ambiguous US politics, and this webpage evaluates the actual trueness of sentences made by big figures. These sentences originate in a multitude of information platforms like television ads, talks, and fresh information. Factmata.com refers to statistically fact verification and detecting.

Hoaxy.iuni.iu.edu are platforms to fetch, find, and analyze online fake news.

Hoax-Slayer.com shows the method for mail spamming and hackers.

Snopes.com is the 1st verification of fact page to verify US tradition.

This shows a detailed comparison of the existing fact-checking resources, in terms of "Topic coverage," "Source of the fake news," "Rating levels," "Dashboard & visualization," "API," "Detection technology" and "Others." Overall, 100% automate

fake news detection is difficult to achieve and has a long way to go. Popular fact-checking websites like Snopes.com, PolitiFact.com, and FactCheck.org are solely dependent on manual detection by professional experts and organizations. However, this process may be time-consuming and expensive and involves a large amount of human involvement to maintain detection systems. It is essential to develop automatic detection approaches to improve the simplification and performance of the existing systems [9].

4 Fake News Detection—Research-Based Approaches

With the growing use of online networking sites and the need to digitalize things, it is very important to have a correct transmission of information on the Internet. A large amount of information is shared, stored on the Internet with the growing rise of the use of the Internet for the smallest things. Our daily activities are done on the Internet like searching for restaurants to order food, or searching a customer care number for your queries, or making payment for a purchase. It is important to ensure that the correct information is available on the Internet and we are not using any wrong information which could lead to exposure of our details on the wrong website. Here we will discuss how we can mitigate the leaking of personal details by detecting any false information on the website [10]. We will look into what are the existing ways of detecting false information, how it is detected what are its limitations, how we can have a system that combines the advantages of all the available methods and mitigate the limitations. Deep learning can be used to process the text to detect if it fake news by analyzing the semantic of the text. There are various categories of fake news detection which are as follows.

4.1 *Component Established Type*

Here we look into the details like who is author of the news, who is the reader, what type of content it is, what country or location the news targets, etc. and then analyze these attributes and find if the news is fake news. Normally, there will be a pattern followed by online fake news authors which can be detected by doing semantic analysis on the content [11]. The content will be such that will have enormous emotions so that it attracts a large audience in a short time. Also, normally this kind of fake news will come from a similar author previously detected in fake news detection, and this can be checked by checking the user ID of the author, location of the author, etc. Also, we analyze the type of audience targeted which can help in checking if it can be fake news.

Another method to detect false information is by having a group of experts to evaluate the information being circulated and validate the authenticity of the information [12]. This requires manual intervention to address the limitations we can have when

we use automated tools to detect FN. Automated tools require correct datasets and can fail if the pattern is not recognized from the previous models. Human validation can address these limitations.

4.2 Data Mining Established Type

Data mining algorithms are widely used these days for this kind of analysis. Fake information can be detected either by a supervised learning model or unsupervised learning model. Former, however, requires large correctly labeled dataset which can be a problem in our case as the fake information data is huge on the Internet, and it becomes difficult to filter these datasets because of various different patterns it follows every time. Hence, latter can help in this case over former as we can still detect fake information without the need of having a large dataset. Latter uses semantic analysis to find patterns in the information to detect fake news.

4.3 Implement Established Type

This involves two types of systems: online detection and offline detection system. Former targets real-time data base and can be beneficial over offline detection system. The models used in the latter can have a disadvantage as it cannot accurately represent current online fake news pattern. Hence, the online system is used to predict effectively the real-time news, and this model can be used to improve the offline methods as well.

Even with all these methods of detecting fake information, we can still see a large amount of fake news are still published and the audience being unaware of the truthfulness. There is a need for a solid model which can help in detecting fake news effectively on the Internet. Research is still in progress in various institutes to achieve this goal but with the growing technology and the use of the Internet and new ways and patterns used by fake authors for distribution fake information, the development of effecting fake information detection system becomes difficult.

Various features which are involved in fake news representations are: creator or user-based features, news content-based features and social context-based features. Each of these features are again involving features such as user profiling, user credibility, behavior-based, linguistic and syntactic, style-based, visual-based, network-based, impact-based, and finally temporal-based features.

5 Future Research and Open Challenges

Along with some promising research in this area, a few challenges and open issues for automatic online hoax news detection are tested. Thus, here we submit a way to construct an effective online fake news detection ecosystem.

5.1 *Unsupervised Learning for Fake News Analysis*

The dataset in real-world available for practical analysis does not come with quality labeling. This is one of the major hurdles in the detection of online fake news. Unsupervised learning refers to a type in the machine learning algorithm. Three types of unsupervised learning models for determining fake news exist. The types are as follows: cluster analysis (CA), outlier analysis (OA), semantic similarity analysis (SSA) and unsupervised news embedding (UNE).

Cluster Analysis: In cluster analysis, the information is classified on the basis of increasing within class commonness, decreasing the between class commonness instead of examining labels. Generation of labels is done for data. In fake news detection, this method is used to analyze and recognize the independent class of content and those who write while discarding the dissimilar ones.

Outlier analysis: Working of OA is based on detection of the uncommon actions. Making use of statistic, distance and weight approaches, outlier analysis algorithm can expose false or forged information and suspicious authors.

Semantic similarity analysis (SSA): Semantic similarity analysis is used for determining almost copied content. Online fake news creators reuse the existing news content because of missing information, comprehension, consciousness, understanding plus awareness. Through semantic similarity analysis, the replica, half replica news information impressionable by crooked writers, an instance, and a false commentator who changes just some phrases for feedback to deceive users can be detected. SSA are essential methods for online fake news detection due to their textual nature.

Unsupervised news embedding: Embedding is an important step in natural language processing, which is a procedure of drawing out decentralized picturization of information. The number pattern is utilized for input to analyze in future. Implanting mechanization is categorized into different types depending on the ways in which they acquire the features of information with another viewpoint. Choice of an embedding method depends on the underlying nature of the news, for successful exposure of faulty information.

5.2 Evaluation of Fake News Detection System

Here to detect fake news, an environment can be given and the performance measures to construct an efficient device to detect online fake news to study. Following are the elements to find out how OFN performs detecting:

Accurate Detection (AD): The sole purpose in a device to detect can be AD of false information. Because of difficulties that the FN figuring process proposes, several effective methods need to be built and employed to increase their efficacy in the systems.

Interactive Visualization: Visualization is an essential element of online fake news detecting and watching systems. They facilitate and eases human understanding by bringing different views and directions to representing data which are time delicate.

Early Warning and Post Intervention (EWPI): As concluded by the above discussions, EWPI has a lot of scope for research when it comes to the finding of online fake news. They are significant features checking which the device for detecting fake news can be unveiled.

Verifying 3rd Party: Involvement of third-party verification eases the detecting device and can make it self-reliant. System to figure out online fake news is discussed here. An effective fake news detection ecosystem contains a combination of an *Alert system*, *Detection system*, and *Intervention system*. It covers analysis, alerting, intervention, and detection of fake news, which form the pillars of an efficient system. A typical comprehensive fake news detection system may involve: intervention, fact checking, fake news detection, suspicious analysis, and potential fake news prediction.

6 Conclusion

On social media, we tend to notice harmful and threatening fake news emanating. This helps malevolent entities to maneuver individual's recourse and pronouncement on salient daily pursuit, to name a few, like education, healthcare and so on. To overcome this complication, the need to have a fake news detection is an extremely magnitude but exacting task across all works of life. To summarize this paper, we have dispensed specific key points as:

- In-depth apprehension of the distinct aspects of fake news, as news content, social context, and news creator to name a few. There can favor an implicating role in social imparting data analysis and anomalous information detection. Through clear characterization of online fake news.
- This survey can help provide knowledgeable and practical convenience for both researchers and participators. Juxtapose to existing detection proposition, existing datasets for training supervised models we provide a perspective for fake news ascertains and

- The researchers are propounded in order to notice the open issues, look into existing assertion frameworks and ameliorate them, and establish a monitoring and detection opportune for online fake news.

References

1. Zhang X, Ghorbani AA (2020) An overview of online fake news: characterization, detection, and discussion. *Inform Process Manage* 57(2)
2. Banerjee R, Feng S, Kang JS, Choi Y (2014) Keystroke patterns as prosody in digital writings: a case study with deceptive reviews and essays. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (2014), pp 1469–1473
3. Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Trans Assoc Computa Linguistics* 5:135–146
4. Bordes A, Chopra, S, Weston J (2014) Question answering with subgraph embeddings. E-print
5. Baldwin BT (2009) Automatic satire detection: Are you having a laugh? In: *Proceedings of the ACL-IJCNLP 2009 conference short papers*. Association for Computational Linguistics, pp 161–164.
6. <https://www.businessinsider.com/here-are-the%20-most-and-least-trusted-news-outlets-in-america-2014-10>. Accessed 19 April 2018
7. <https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/blob/master/data/facebook-fact-check.csv>. Accessed 19 April 2018
8. Cao N, Shi C, Lin S, Lu J, Lin YR, Lin CY (2016) Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Trans Visual Comput Graphics* 22(1):280–289
9. Castillo C, Mendoza M, Poblete P (2011) Information credibility on twitter. In: *Proceedings of the 20th international conference on World Wide Web*. ACM, pp 675–684
10. Castillo C, Mendoza M, Poblete B (2013) Predicting information credibility in time-sensitive social media. *Int Res* 23(5):560–588
11. <https://www.cbc.ca/news/canada/toronto/scarborough-hijab-attack-1.4487716>. Accessed 21 March 2018
12. Cha M, Mislove A, Gummadi KP (2009) A measurement-driven analysis of information propagation in the flickr social network. In: *Proceedings of the 18th international conference on World Wide Web*. ACM, pp 721–730

Recognition and Generation of Logically Related Words for Historical Text Data using Reconstruction of Protowords



G. Sajini and Jagadish S. Kallimani

Abstract This article presents new methods for the study of language evolutions which helps researchers and experts. Initially, a method is used to determine if the words are cognate or not. A linguistic information algorithm is proposed to derive cognates from online dictionaries. Later, a dataset is created of similar terms and machine learning techniques are used to focus on spelling in order to classify the cognates. The aligned subsequences are used to identify standards and guidelines for language change in newly created languages mainly to distinguish between non-cognates and cognates which are used for classification algorithms. Next, discriminating cognates and debts give an insight into a language's history and allow a clearer understanding of the linguistic relationship. The task of reconstruction of protowords is to recreate words from its modern daughter languages in an ancient language. The method is based on the regularity of words and use knowledge from many modern languages to build an ensemble method for protoword reconstruction. This method is applied to multiple datasets to improve from the previous dataset accuracies.

Keywords Cognates · Protoword · Linguistic borrowing · Diachronic linguistics · Etymons · Ensemble Learning

1 Introduction

The natural languages are like symbiotic systems because they constantly change and upgrade over the years. The core concept behind the comparative approach is to compare several sister languages on the basis of properties in order to determine the properties of their mutual ancestor [1]. The comparative reconstruction was a

G. Sajini

Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore, India

e-mail: sajini.narayana@gmail.com

J. S. Kallimani (✉)

Department of Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India

e-mail: jagadish.k@msrit.edu

manual method which took a long time, involving comprehensive research. The first question involves designing strategies to identify cognates. The second issue includes investigating regarding borrowing and studying about the words movement when it comes from other languages. Cognates are nothing but words with the same meaning and shared origin in various languages. Cognates can be studied in many research fields, such as language learning, bilingual words comprehension, corpus linguistics, cross-lingual knowledge retrieval and computer translation, which are not only useful in historical but also in comparative linguistics.

The process through which words enter into another's vocabulary is called as *linguistic borrowing*. There is no such thing as a *pure language*, which means without the need to borrow from any foreign languages. An empty word too is called as borrowing. The term *loanword* is described as lexical object that is borrowed from a different language, which was not originally in part with receiver language vocabulary but taken from different language that became words of the vocabulary of the language that is borrowed [2]. The exceptional connections between languages and the proliferation of interactive technologies have resulted in language amplification by borrowing. The languages are primarily focused on needs and popularity in other languages.

Differentiating cognates from borrowings are critical in language classification problems. While the language classification that identifying cognates and word borrowings from different languages is undoubtedly regarded as important to the history of language borrowing, it is necessary that context of phylogenetical inference may increase the incorrect values to which the cognates are identical [3]. The fact that false cognates can draw wrong conclusions about the relation among different set of languages. One should highlight two directions of the research, which depend on these word-forms: diachronic linguistic analysis, which deals with the production of language over time and reviewing other languages by concentrating on learner during the acquisition of second languages [4]. Cognates can also contribute in poorly written languages with minimal resources to the lexicon generation process.

2 Identification of Words

In the historical linguistics, most research concentrate on recognizing pairs of cognates automatically. In order to classify cognates, there are three critical things commonly examined: semantic, phonetic and orthographic similarities. These were combined and used individually for finding cognate pairs. WordNet is used to determine semantic similarity and String similarity measures of cognate candidates, with phonetic or spelling word types as inputs these were different measures taken to examine and to compare. For cognates on the basis of orthographic and phone forms, algorithms for string alignment should be used [5]. The changes are undergone by words as they were entered from one language to another, and the transition rules they obey were effectively used to identify cognates in a variety of different ways.

Some of the metrics in this area is most widely used is edit size, XDice metric and longer common subsequence ratio. The SpSim is a more complex process to calculate the cognate pair similarities that tolerate learned transitions between words. A newly developed ALINE lines phonetics of words based on various phonetic characteristics and uses complex programming to measure similarity ratings [6]. Recent methods have been focused on neural networks and dictionary definitions to identify cognates reliably. The distinction between cognates and bonds is based on frequent sound changes, which produce frequent phonemic correspondences in cognates, in keeping with the regularity principle.

3 Generation of Related Words

The two main concerns and difficulties faced in linguistics of the diachronic are history and comparative reconstruction. The historical derivation is that new forms of the words are derived from the former. The contrary method is the comparative reconstruction of words that are borrowed from different set of languages will be newly constructed. Language derivation has been a constant interest for researchers [7]. The first attempts to solve this issue were regular sound connections, with a protolanguage, to construct modern forms of words or vice versa. For some early research, multiple alignments for historical comparison and proposal methods for cognitive alignment and recognition are being studied. Most of the previous approaches were based on phonetic.

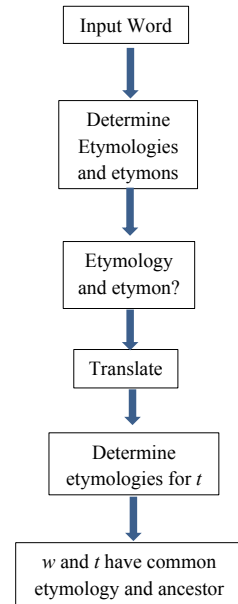
When orthographic and phonetic forms of the words are applied, it was very effective to align the associated words with the extraction of spelling changes from one language to another. Spelling changes have also been used to produce cognates, which have not yet been studied as intensively but are very closely related with the task of identifying cognates [8]. Whereas cognates are identified as a word in source language is used to generate cognates in a target language automatically.

An algorithm [9] developed to focus on editing distancing and recognizing orthographic signals. When the words enter a new language, a method is used to generate cognates that are mainly focused on statistical computer translations and learning orthographic patterns.

4 Constructing Dataset of Related words

A cognate extraction algorithm from electronic dictionaries containing etymological information is created and get a set of corresponding words from dictionaries, from this an automatically develop machine learning methods to identify and produce related words [10].

Fig. 1 Identification of word pairs and cognate pairs



4.1 Steps for Cognate Pair Recognition

- Find a selection of terms for defining the cognate pairs in a certain L1 language and apply the following strategy between L1 and the related language L2.
- Define the etymologies of the terms given then all words without etymology are translated into L2. The pairs of input terms and their representation are considered by cognate candidates.
- Extract the etymological details for the translated terms using electronic dictionaries.
- Compare their etymologies and etymons to classify cognates for each pair of candidates.
- Mark the words as cognates if they fit. Even if different forms of the same word are inflected, presume that the Etymons match. The Figure 1 shows how the word-etymon pairs and cognate pairs are identified.

4.2 Calculation of accuracies for different linguistic datasets

Randomly take 500 input words w from different other background languages such as German, French, Greek, Romanian and Latin, and along with the algorithm accuracies one should manually calculate the cognates in order to test the above automatics method of extraction of etymology-related knowledge and for the detection of related words [11]. Obtain the accuracy rates of each languages using different

automatically result generating machine learning algorithms and then compare the results with the etymologies obtained automatically and measure for each language the accuracy of the etymology extraction. The best accuracy giving linguistic dataset will be chosen. Once the best linguistic dataset is obtained, one should concentrate on training the model to improve the system performance by upgrading and modifying the dataset using online dictionaries, and again manual results should be calculated and compared with the algorithmic results for achieving better accuracy [12].

5 Automatic Recognition of Related Words

The main objective is to define the association between the terms automatically and concentrate on cognates and borrowing by using an alignment approach and aligned subsequence [13]. Methodology used to identify cognates and discriminate between cognate and borrowing methods is as follows:

- *Align the related word pairs with a String alignment algorithm.*
- *Extraction of the related terms features.*
- *Machine learning distinguishing algorithms should be used to identify cognates and discriminate between cognate and borrowings.*

5.1 String Alignment

In order to fit pairs of terms, Needleman–Wunsch algorithm is used. This algorithm attempts to decide the best alignment over the whole input sequence range. The algorithm guarantees optimum alignment and is effective [14]. The key principle is that the optimum path to this stage would be every intermediate sequence that leads to optimum paths. Thus, by increasing expansion of intermediate subpaths, the optimum path can be calculated. One must regard terms as entry sequences for orthographic alignment and use a very simple substitution matrix that gives equivalent values, not taking diacritic factors into account.

5.2 Feature Extraction

The features are extracted around mismatches in alignments with the matched pairs of terms as an input. The insertion, deletion and substitution are three types of mismatches found while extracting the features [15].

Illustration of String Alignment and Feature Extraction

Consider the word *exhaustiv* which is derived from the roman language and word *esaustivo* which is derived from the Italian language, these two words are considered as cognate pairs. The String alignment is shown below:

**exhaustiv –
es – austivo**

- Mismatch 1: **x** & **s** occurred due to substitution.
- Mismatch 2: **h** & **-** occurred due to deletion.
- Mismatch 3: **-** & **o** occurred due to insertion.

5.3 Learning Algorithms

The spelling changes and the relationships between words are identified using Support Vector Machines and naive Bayes. With the Weka Laboratory, a collection of algorithms and tools bring program together. For SVM, the wrapper classes are used to supply to the LibSVM from Weka and the RBF kernel is used, which can deal with case where the relationship among the wrapper class attributes and kernel parameters are well maintained without any conflicts.

5.4 Evaluation Measures

The performance is assessed using the following evaluation measures:

- F-score
- Accuracy
- Recall
- Precision

5.5 Task Setup

- The data is divided into two subsets for each language pair, one for training and another one for testing with the ratio 3:1.
- The experiment is conducted for different values of n -gram of size n . The value ranges from one to three that belongs to n .
- Threefold cross validation is carried out over the training datasets.
- The grid search is performed over the training datasets to optimize the evaluation measures.

6 Linguistic Factors

The below are the high-predictive linguistic factors to improve system accuracy of the ML algorithms and to achieve better evaluation measures.

- **Part of speech**

Part of speech plays a vital role in analyzing whether incorporating information about the word element leads to changes in results. Verbs, nouns, adverbs and adjectives come to language-specific ends; therefore, one should assume that this factor may be helpful during orthographic learnings. For the learning algorithm, one should use the POS feature as a second categorical feature.

- **Hyphenization**

A hyphenated type of words is used for the algorithm alignment to see how it increases cohesion and therefore feature extraction by establishing distinctions between the syllables and derive the term hyphenization from the Romanian RoSyllabiDict for knowledge strings also some Perl modules 19 accessible for certain languages. The hyphen constraints will be considered as the extra terms.

- **Consonants**

System's effectiveness is analyzed by the matched consonant skeleton of the words (i.e., the word form of which vowels are discarded) is qualified and checked to provide less detail that is useful for word recognition. As vowels are dropped, the efficiency of the system decreases. In data collection, one should also train and check the decision tree classifier and its output.

- **Stems**

The first check is performed using stems rather than lemmas as data to understand that the associated appeals reflect the form of relation between the two terms. One should use the Snowball Stemmer and note that when stems are used instead of lemmas, their efficiency is increased.

- **Diacritics**

Diacritics elimination affects the system's efficiency, and many words were transformed by an increase of linguistic diacritics when a new language was entered. Therefore, diacritics are expected to play a role in the classification task. One should note that the precision of the test set is lower in nearly any case when diacritics are excluded. The study of the classification of the characteristics derived from the misalignment in this direction offers much clearer evidence that is more than a quarter of the top 500 features containing diacritics for both languages.

7 Reconstruction of Protowords

Two-step approach is followed as shown in Figure 2 to solve the problem of reconstruction of protowords. Firstly, a system is based on sequence labeling to recreate protowords. This approach is used separately in every modern language. Second

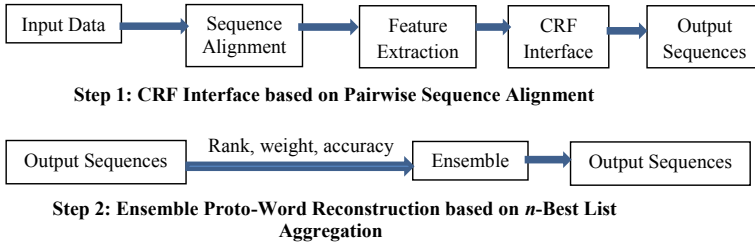


Fig 2 Protowords reconstruction methodology

is to pull together the optimum productions from different vocabulary, a variety of ensemble modes to incorporate knowledge from various structures and to calculate the system performance outcomes.

7.1 Fusion by Rank

The rank is calculated based on the individual production. The output word is u , and its rank weight w over the n lists is calculated as

$$w_r(u) = (1/k) * \sum_{i=1}^k w(u_i)$$

7.2 Fusion by Rank and Accuracy

Training accuracy is considered here for each language. The best obtained accuracy is considered and multiplied with the weight obtained by training accuracy for each language I calculated as

$$w_{ra}(u) = (1/k) * \sum_{i=1}^k w(u_i)\pi(i)$$

7.3 Fusion by Weight

The confidence score is considered here based on the individual productions. With the help of obtained confidence score, productions are reranked in sequence labeling system. L_i is calculated as

$$w_c(u) = (1/k) * \sum_{i=1}^k w(u_i)$$

7.4 Fusion by Weight and Accuracy

Training accuracy is considered here for each language. The best obtained accuracy from all the languages is considered and multiplied with the weight obtained by training accuracy in sequence labeling system calculated as

$$w_{ca}(u) = (1/k) * \sum_{i=1}^k w(u_i)\pi(i)$$

8 Modern Word Production

The hypothesis is that the term v in language L1 derives from the text u (by means of a borrowing). One should use stems rather than lemmas to produce modern words. The form of inflected or derivative words reduces both lemmatization and stemming in a more drastic way, to the common form of base word, but stemming. In other words, the words are reduced to lemmatization. In addition, stemming process eliminates suffixes and prefixes that are not lemmatized; this could make a difference in word production that leads to better results, instead of lemmas. The Snowball Stemmer is used, and 9 of our 20 donor languages have stemmers.

One possible reason that stemming does not improve when new words enter the language and so root cannot necessarily be more easily produced than the whole word (including the appeals) due to its morphological changes. Many morphophonological changes have been observed over the years, and there are changes even in declinations and conjugations.

8.1 Cognates production

This task is much like production. The interest in recognizing cognates using computation methods has been significant in the last couple of years, but very few studies deal with automatic cognate pairs development. Main goal is to decide whether the system works for cognates differently from borrowing.

9 Automatic-Related Word Production

The method of orthography-based production of related words is widely used and take into account the form of association between words and discern specifically between protowords. Main of protoword reconstruction without the use of any recipient's lexicon or data in the recipient's language, with less resources. A new approach to the production of related words is introduced by domain experts and a strategy based on random circumstances.

9.1 Conditional Random Fields

The system can learn orthographic patterns of spelling change between source and target language by harmonizing related words. Main approach is marking of sequences achieving transliterations that have proved helpful in our situation, the words are the numbers, and their characters are tokens. The aim is to get a sequence of characters composing its related word for each input word and use Custom Random Fields (CRFs) for this purpose. The CRF system uses the n -gram characters of the input words, extracted around the current token from a fixed window.

9.2 Pairwise Sequence Alignment

The Needleman–Wunsch global alignment algorithm are mainly used. Consider two words *frumos* from the roman language and another word *formosus* from the Latin language. The pair sequence alignment is as follows:

f o r – m o s u s
f – r u m o s – –

The associated label shall be the character that takes place in the same position in the target word for every character in the source word (after alignment). In the case of insertions, add the new character to the previous label because we could associate the inserted character as the label with no input character in the source language. Apply to each word and add two additional characters B and E, which represent the beginning and end of the word, for each input word. These specific characters are linked to the characters that are inserted in the target word at or at the end of the word. We remove the sticker to reduce the number of labels.

9.3 Evaluation Measures

The performance is assessed using the following evaluation measures:

Average Edit Distance

The editing gap between the words render and the standard, in order to determine how close, the outputs are to the appropriate form of the associated language. The unstandardized and the normalized editing distance are recorded for normalization in $[0, 1]$ interval.

Coverage

Coverage is also referred to as accuracy of top n items in the language dictionary. The functional value of evaluating the top n findings to provide potential output terms to a smaller list that can be evaluated by linguists. $n = 1$ is known for its accuracy of measurement.

Mean Reciprocal Rank

The mean mutual rank is a measure for measure for systems that produce an ordered output list for each input instance. In the case of an input word is calculated as:

$$\text{MRR} = (W_i) \frac{1}{m} \sum_{i=1}^m \frac{1}{\text{rank}_i}$$

where m indicates the total number of input instances, rank_i indicates the position of protoword w_i in output lists.

10 Conclusion

A method for automatically sequence alignment identification of related words is identified. For the first time the tool is employed to classify cognates for an automatically derived cognate dataset for four language pairs, for the detection of lexical shifts happening as words are joining new languages. Then the method of discrimination is based on their spelling between cognates and debt. The predictive analysis reveals that for cognates and borrowing, the orthographic indications are different than that underlying language factors. One should look for more fine-grained vocabulary and more languages to broaden the studies and a framework which is language-relating, and it might boost the efficiency of the program to integrate language skills. An automatic method to produce related words has been introduced and used a sequence labeling and sequence alignment approach which combined the results of the individual systems with assemblies. The benefit to the method is that less data is needed than previous approaches and that in historical linguists where resources are scarce, incomplete information is also accepted. First used the method to reconstruct

Latin protowords using multiple datasets in romantic languages to develop modern Romanian word forms as a receiver language. The languages are rather than language families in light of their etymons (ancestors) and cognates, and the recipient language could be a valid word and the created sequences also reflect the elderly shapes or the feminine shape of a word for the substantives.

References

1. Rama T (2016) Siamese convolutional networks for cognate identification. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers. Osaka, pp 1018–1027
2. Pagel M, Atkinson QD, Calude AS, Meade A (2013) Ultraconserved words point to deep language ancestry across Eurasia. *Proc Natl Acad Sci* 110(21):8471–8476
3. Gomes L, Lopes JGP (2011) Measuring spelling similarity for cognate identification. In: Proceedings of the 15th Portuguese conference on progress in artificial intelligence, EPIA 2011. Lisbon, pp 624–633
4. Gooskens C, Heeringa W, Beijering K (2008) Phonetic and lexical predictors of intelligibility. *Int J Humani Arts Comput* 2(1–2):63–81
5. Hall D, Klein D (2010) Finding cognate groups using phylogenies. In: Proceedings of ACL 2010. Uppsala, pp 1030–1039
6. List JM (2012) LexStat: automatic detection of cognates in multilingual wordlists. In: Proceedings of the EACL 2012 joint workshop of LINGVIS and UNCLH. Avignon, pp 117–125
7. Luong M-T, Brevdo E, Zhao R (2017) Neural machine translation (seq2seq) tutorial. <https://github.com/tensorflow/nmt>
8. List JM, Greenhill SJ, Gray RD (2017) The potential of automatic word comparison for historical linguistics. *PLOS ONE* 12(1):1–18
9. Luong T, Pham H, Christopher D (2015) Manning. 2015. Effective approaches to attention-based neural machine translation. In: Proceedings of EMNLP 2015. Lisbon, pp 1412–1421
10. McMahon A, Heggarty P, McMahon R, Slaska N (2005) Swadesh sublists and the benefits of borrowing: an Andean case study. *Trans the Philolo Soc* 103(2):147–170
11. Tsvetkov Y, Ammar W, Dyer C (2015) Constraint-based models of lexical borrowing. In: Proceedings of NAACLHLT 2015 Denver, CO, pp 598–608
12. Simard M, Foster GF, Isabelle P (1992) Using cognates to align sentences in bilingual corpora. In: Proceedings of the fourth international conference on theoretical and methodological issues in machine translation. Montreal, pp 67–81
13. Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press
14. Schuler GD (2002) Sequence alignment and database searching. In: Baxevanis AD, Ouellette BFF (ed) *Bioinformatics: a practical guide to the analysis of genes and proteins*, 43. Wiley, pp 187–214
15. St. Arnaud A, Beck D, Kondrak G (2017) Identifying cognate sets across dictionaries of related languages. In: Proceedings of the 2017 conference on empirical methods in natural language processing. Copenhagen, pp. 2519–2528

Proposing Host-Based Intruder Detector and Alert System (HIDAS) for Cloud Computing



Amit Kumar Chaturvedi, Punit Kumar, and Kalpana Sharma

Abstract The use of online applications is increasing regularly, and during this COVID-19 pandemic, its usefulness is envisaged to all of us. Data sharing either in the form of text, videos, or audio is the main requirement for online meeting or discussions, and during COVID-19 pandemic, most of the meetings like business, learning, presentation, or other types from private business sector to government are done through online applications. These applications are accessible through Internet from devices like mobile devices, laptops, palmtops, or desktop computers. Now, with the use of these online meeting and data sharing applications, the need of data security also increases. During this COVID-19, the activities of intruder in online applications have been noticed more than before. IDS solutions provide security against such problems. The IDS solutions are categorized into two categories according to their nature of work: (i) The Network-based Intruder Detection Systems (NIDS) and (ii) The Host-based Intruder Detection System (HIDS). In this paper, we are proposing a model Host-based Intruder Detection and Alert System (HIDAS) for securing the data and applications from such intruder attacks for public cloud computing H machines.

Keywords Cloud computing · Intruder detection system · IDS · Host Machine · Alerts · Prevention · Signature · Anomaly

1 Introduction

Most of the people and organizations started working through online applications for their business as well as personal requirements during the COVID-19 pandemic.

A. K. Chaturvedi (✉)
Department of MCA, Government Engineering College, Ajmer, India
e-mail: amit0581@gmail.com

P. Kumar · K. Sharma
Department of Computer Science, Bhagwant University, Ajmer, India
e-mail: p_punitkumar@rediffmail.com

K. Sharma
e-mail: kalpanasharma56@gmail.com

The online data traffic like text, audio, video, etc. are increasing regularly, and it is need of the hour also. Most of the applications are providing the facilities of online communication, meeting, teaching, etc. either through social media platforms, some specialized applications, for selling, purchasing house hold goods, or for other business purpose. All are using cloud computing services, and some have their own public cloud server. The security of these communication data items is the prime requirement of every cloud server.

According to the National Cyber Security Agency and The Computer Emergency Response Team of India (CERT-In), the number of cyber-attacks increases on personal computer networks and routers since professionals of different fields were asked from home in the wake of COVID-19 in the country, and in these attacks, phishing emails cases, ransomware attacks, etc. were registered, where attacker demanded for an amount against releasing their important data to different organizations and personals having good market reputation. Malicious programs or intruders take entry in such servers through many techniques and try to theft important data from these servers [2, 4, 9, 10].

Intruder Detection System (IDS) is such a system that will detect for any change in the configuration or server settings and will report to the system administrator about the changes [7, 8, 12]. The system admin will either allow for these changes or deny for that. Accordingly, the processes will be started or blocked on the server. So, the requirement for an Intruder Detection System and Alert system is increased during this COVID-19 pandemic year.

Intruder Detection System (IDS) solutions are categorized in two categories: (1) Host-based IDS and (2) network-based IDS [3, 5, 11]. In this paper, we will focus more and present the study of the existing Host-based IDS and proposing an innovative model named the Host-based Intruder Detection and Alert System (HIDAS).

The cloud computing is an open and fully distributed architecture, and hence, it will be an easy target for the intruders or attackers. Due to its architecture, there are two problems still persist with the cloud computing: (1) lack of control and (2) security [1, 6, 13, 14]. Because attacker changes its attacking strategies regularly, security policy should also to be changed and updated regularly accordingly. Besides that prompt notification of an attack is also important security measure, because it may alert the system administrator to take prompt and appropriate action in time, so that the damage will be as less as possible.

2 Intruder Detection System and Its Types

Intruder detection is not a new concept. Intruder detection is done in almost all the system, because these intruders always try to collect the information of the ongoing processes or old process details and passes such information to the attackers for using against the organization. So, beware from these intruders and adopt a security policy or identification process to catch them and secure the system from such intruders is

always a requirement for every system. As cloud computing has an open architecture and distributed system, so the chances for intruder attack during COVID-19 pandemic increases more because all professionals are working online on the important projects. But any how we have to stop the entry of such intruders and if they are entered, it is the requirement to know, do they have made any unauthorized changes in the system or in the configuration settings. This alert will help to take the requisite action for making the system secure and minimizing the damage.

In computer networks, we have two types of Intruder Detection Systems: (1) Network-based Intruder Detection System [NIDS] and (2) Host-based Intruder Detection System [HIDS]. In this paper, we are focusing on the Host-based Intruder Detection System. Host-based Intruder Detection Systems are mainly used in the public cloud computing, because it is distributed in nature. So, cloud application providers want to secure their cloud server from such intruder attacks.

There are some security measures like firewalls taken by many networks for stopping the attacks. These security measures are only effective for external attacks but are insufficient for internal attacks. The internal attacks are done by the valid users or by the complicated external attacks, e.g., DOS and DDoS. Such attacks cannot be tackled by firewalls like mechanism, but IDS systems will play an important role in these cases.

The Intruder Detection System [IDS] or Intruder Detection & Prevention System [IDPS] layer is used as additional prevention layer of security used by the cloud Host Server. It will detect the many known as well as some unknown attacks and gives alerts to the administrator.

The intruder identification process is a strategic issue for an IDS, and this process depends on two major things: (i) the service model and (ii) security mechanism. Here we are discussing the IDS implementation for the cloud computing, and the cloud computing has three service models: (i) SaaS, (ii) PaaS, and (iii) IaaS.

- Software as a Service (SaaS)—The SaaS users are almost dependent on the solutions provided by their service providers, and the process of Intruder detection or identification is on the part of service provider. Users may have only the option of getting some logs and deploying a custom monitoring and altering on that information.
- Platform as a Service (PaaS)—Like SaaS in PaaS, the process of IDS will also done by the service provider, because the Intrusion Detection System perform its work that is outside the scope of application area. Users may configure its application and platforms to log onto a central location, where IDS can setup monitoring, altering, and altering the administrator if something unauthentic changed. Here IDS will monitor the application users and platform also.
- Infrastructure as a Service (IaaS)—This service later of cloud computing is the most flexible for IDS deployment. There are more options for configuring IDS according to the requirement. At IaaS, the IDS can be performed at four primary positions:- (i) in the Virtual Machine itself, (ii) in the Hypervisor, (iii) in the virtual network, and (iv) in the traditional Network.

In this paper, we will discuss the Intrusion Detection System implementation at the hypervisor or Host Machine position. The hypervisor or Host Machine allows the IDS to monitor the hypervisor and anything happened between virtual machines and Host Machine. Cloud computing provides open-access distributed computing environment. The user will be at the Virtual Machine and the related data and application software reside at the Host Machine [15, 16]. So, all the processes execute between Host Machine and Virtual Machine.

Besides service models, the second part is the security mechanism adopted for handling the issues. Because the IDS is a security tool with different configuration settings and every configuration plays an important role in designing the security mechanism [17–20]. So, how much security has to be enforced on the Host Machine is based on the requirement.

3 Proposed HIDAS Model and Methodology

An intrusion is defined as an attempt to compromise the Confidentiality, Integrity, and Availability (CIA) of the system. This CIA defines the security mechanism of the system or network that has to bypass by the intruder. These intruders are triggered by the attacker in a system to access the data and processing information and breach the confidentiality of that system. If we are discussing this in the respect of cloud computing, the chances increase more. So, the requirement of an Intrusion Detection System on a cloud server or Host Machine is more necessary.

We are proposing an innovative Host-based Intrusion Detection and Alert System for implementing the better security mechanism with the provisions of existing Intrusion Detection System.

Let us first discuss the Intrusion Detection System types and its different approaches. The types of IDS we have already discussed above and some examples of NIDS and HIDS are mentioned in Fig. 1. The other important feature of the

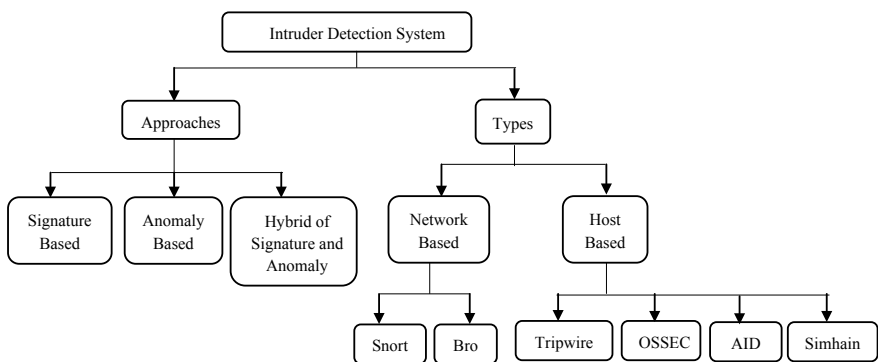


Fig. 1 IDS approaches and types

IDS is the approach used for designing the security mechanism of the IDS. So, there are three approaches used for designing the IDS: (i) signature based, (ii) anomaly based, and (iii) hybrid of signature and anomaly based.

So, keeping in view of the working of existing IDS systems, the innovative proposed HIDAS will work using the signature-based approach. The working of the HIDAS can be divided into four steps: (i) the identification of the intruder, (ii) regular monitoring the Host Machine and the activities between Host Machine and its multiple Virtual Machines, (iii) sending prompt alert message to the administrator with the policy id breached by the intruder, and (iv) taking immediate action to block the intruder and keeping track of damage done in the system.

When the HIDAS is installed on the Host Machine or cloud server, it will prepare a database and categorize the different files related to providing the services to its clients like client's impersonation data, server security and configuration related files, etc. It will not keep track of the data files of the clients because data files contents are regularly changed or updated by the user and it is under the jurisdiction of the user, but every file will be encrypted with individual client's information-based security key. This HIDAS will monitor the traffic comes to and from this Host Machine as well as all the activities, on the cloud Host Machine and between the Host Machine and multiple Virtual Machines under this Host Machine.

When the HIDAS installed on the Host Machine or cloud server, it will create a database of files and categorize these files as configuration files, policy files, and data files. The configuration files keeps the HIDAS internal security structure that how internal security is planned by the administrator at the Host Machine. These files should be kept with high security, because multiple tenants are connected with a Host Machine and all are sharing the buffer and hence is accessible to all; that buffer is the major risk area.

These files will be bounded with multiple rule names, where each rule will be used for configuring and enforcing security policy for the related files. As every rule requires a different level of security, three security levels are defined in HIDAS and are named as HIGH_SEC, MID_SEC, & LOW_SEC. So, the administrator will set these security levels and associate these with different rules and HIDAS system will be maintained the desired security. The configuration files are on the regular monitoring of HIDAS. In HIDAS, only admin is permitted to do addition, removal, or modification in the configuration files of the Host Machine (Fig. 2).

The second category of the Host Machine files is the policy files. These policy files are very important for designing the security in the system. The policy files will enforce the security on the rest of the files in the Host Machine. To perform this security mechanism, the policy files will then relate to the multiple rules and every rule will be assigned to multiple files. So, one rule will cover multiple files and this rule has some attributes like rule_name, List_of_files_under_rule, rule_holder_id, Security_level, rule_holder_email, etc. A rule may be under a different rule holder, and all the related problems under this rule will be handled by the rule holder.

The files then encrypted with a security key, and encrypted files are kept in the system that improves the security level of the system. The HIDAS now sets in the Host Machine and will regularly monitor the traffic comes to and from the Host

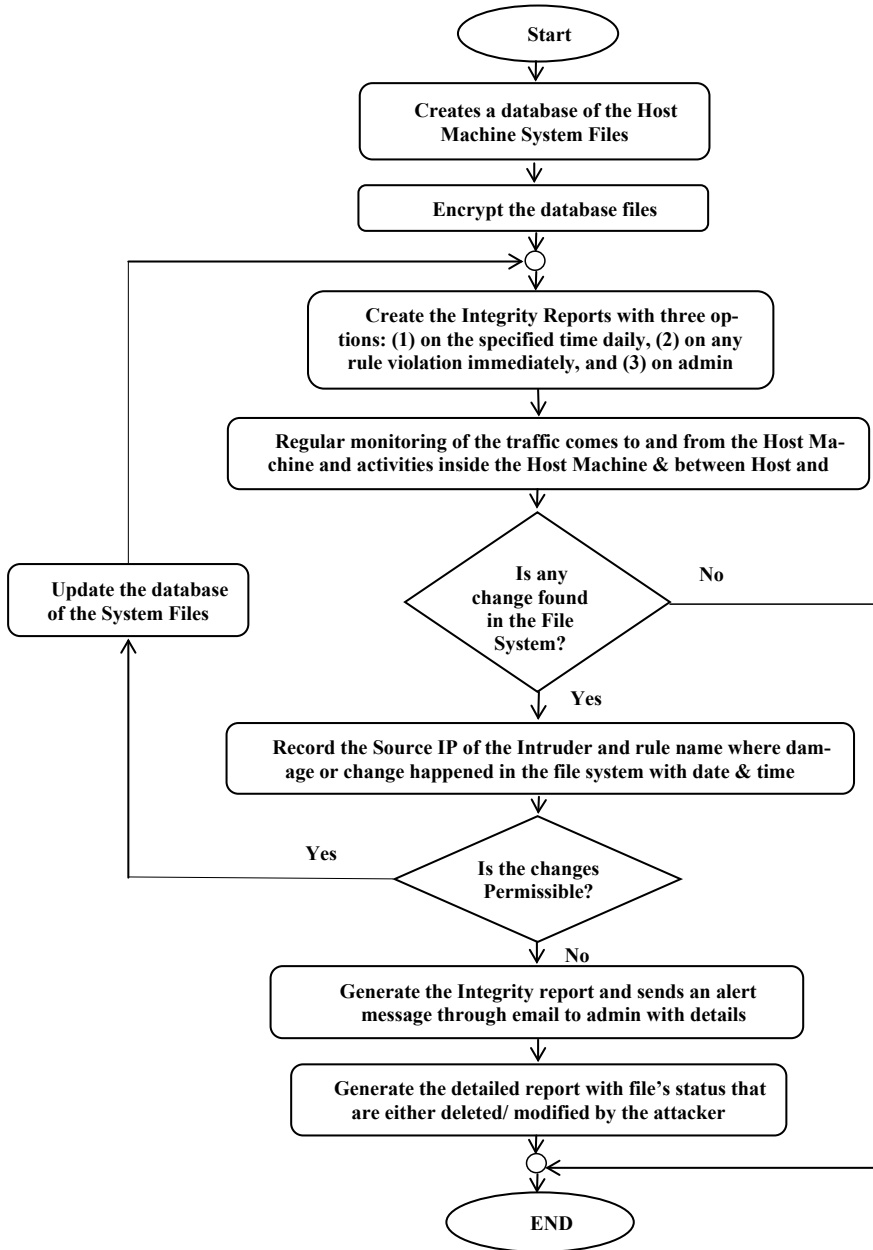


Fig. 2 Proposed HIDAS model for intrusion detection

Machine, different activities done inside the Host Machine, and the activities done between the Host Machine and different Virtual Machines.

The HIDAS system will then regularly monitor the traffic comes to and from the Host Machine, the activities done in the Host Machine related to the files bounded with the different rule names, and the activities done between Host Machine and the various Virtual Machines, i.e., clients and server. During the working if HIDAS found any changes in the file system, the system will record the intruder activities with its Source IP, the changes with date and time of the modifications and then put this information of the modification or changes done by the intruder to the administrator and ask for permission either allow or deny these changes. If the administrator allows these changes by checking that this is valid change or not, then the HIDAS will update the database of the system files with these changes. But if the administrator denies these changes, then generate the integrity check report and sends it to the administrator through email to the admin with these details. After then generate the detailed report with file's status that are either modified / deleted /or added by the intruder at a pre-specified folder the details or the action. This process of HIDAS will be continued throughout till it will be ended specifically on the Host Machine.

4 Active Alert and Prevention System

The HIDAS will have an active alert and prevention system, because prompt alert will help to reduce the level of damage and give sufficient time to manage the damage. HIDAS will regularly monitor the Host Machine, its internal activities and activities between Host Machine and its Virtual Machines, It will be designed in such a manner that it will identify the intruder immediately whenever it will try to do unauthentic activities with the configuration files, policy files, or unauthorized data files. But the internal security mechanism role is also very important to understand here. As even if a security system has provisions but if these options are not used in a strategic manner, they will be silent and not result oriented. So, the role of skilled administrator will be a part of establish an efficient security mechanism. There are the provisions of an active alert system in the HIDAS but how efficiently they be utilized will be on the part of the administrators skill and strategy.

A prevention system majorly depends on the environment, because each environment has different level of risks and different surfaces of attacks. Here, we are discussing the HIDAS in respect of cloud computing and different types of cloud have different risks and surfaces of attack. The different types of clouds are: (i) Public Cloud, (ii) Private Cloud, (iii) Community Cloud, and (iv) Hybrid Cloud. Here we are discussing the HIDAS in respect of Public Cloud, and the surface of attack in Public Cloud may be: (i) Cloud Provider, (ii) Co-tenants, and (c) User. We have discussed this above that in SaaS and PaaS service models, major security part is on the side of Cloud Provider. The major performance issues in public clouds are:

- **Workload:** It is always high in public clouds because tenants may be increased exponentially and workload varies for the users, but for the cloud provider, it will be always high all the time. The traffic is bursty and unpredictable in public clouds.
- **Security:** In case of public clouds, the security cannot be kept high because multiple tenants have their own issues. So, the security kept at low.
- **Latency time:** When the security level increases, the latency time will be decreased and when security level decreases the latency time will be increased.
- **Cost:** The security is a paid service and as much as security will be imposed on the system, the cost of security will be increased and reducing the security level will decrease the cost.

So, a prevention system should be configured keeping in view these all the points in mind for a public cloud system and the results be obtained accordingly. In HIDAS, we have considered these issues and will provide the provisions for them.

5 Conclusion

Intruders are always a problem for any system, in case of public cloud computing the issues matter more because as stated above that we cannot keep the security at high level in the present competitive environment, as the increase in security will proportionally increase in the cost of service. So, users or tenants and their Virtual Machines set with low security and on the cloud server the monitoring will be maintained efficiently with a planned security mechanism.

HIDAS is a model that provides the provisions for planning an efficient security policy with multiple rules and every rule is bounded with its attributes. A skilled and experienced administrator will plan an efficient security mechanism, and the public cloud provider will be able to serve its client in a secure and safe environment. Hence, HIDAS is a useful model for creating a less vulnerable cloud service environment for intruder attacks, and it is a good proposal for Host-based Intruder Detection System.

References

1. Roschke S, Cheng F, Meinel C (2009) An extensible and virtualization-compatible ids management architecture, 2:130–134
2. Modi C, Patel D, Borisaniya B, Patel H, Patel A, Rajarajan M (2013) A survey of intrusion detection techniques in cloud. *J Netw Comput Appl* 36(1):42–57
3. Mazzariello C, Bifulco R, Canonico R (2010) Integrating a network id into an open source cloud computing environment, pp 265–270
4. Jones AK, Sielken RS (2000) Computer system intrusion detection: a survey. *Computer Science Technical Report*, pp 1–25
5. Lo C-C, Huang C-C, Ku J (2010) A cooperative intrusion detection system framework for cloud computing networks. pp 280–284

6. Garfinkel T, Rosenblum M et al (2003) A virtual machine introspection-based architecture for intrusion detection. 3:191–206
7. Aborujilah A, Musa S (2016) Critical review of intrusion detection systems in cloud computing environment. In: 2016 international conference on information and communication technology (ICICTM).
8. Aryachandra AA, Arif YF, Anggis SN (2016) Intrusion detection system (IDS) server placement analysis in cloud computing. 2016. In: 4th international conference on information and communication technology (ICoICT)
9. Chiba Z, Abghour N, Moussaid K, El Omri A, Rida M (2016) A survey of intrusion detection systems for cloud computing environment. In: 2016 international conference on engineering & MIS (ICEMIS)
10. Yasir M, Shibli MA, Umme H, Rahat M (2013) Intrusion detection system in cloud computing: challenges and opportunities. In: 2nd national conference on information assurance (NCIA), pp 59–66
11. Modi CN, Patel DR, Patel A, Rajarajan M (2012) Integrating signature Apriori based network intrusion detection system (NIDS) in cloud computing. In: 2nd international conference on communication, computing and security, pp 905–912
12. Alharkan T, Martin P (2012) IDSaaS: intrusion detection system as a service in public clouds. In: Proceedings of the 12th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGrid), pp 686–687
13. Manthira SM, Rajeswari M (2014) Virtual host based intrusion detection system for cloud. *Int J Eng Technol (IJET)* 5:5023–5029
14. Khatri JK, Khilari G (2015) Advancement in virtualization based intrusion detection system in cloud environment. *Int J Sci Eng Technol Res (IJSETR)* 4:1510–1514
15. Gupta S, Kumar P (2015) Immediate system call sequence based approach for detecting malicious program executions in cloud environment. *Wireless Pers Commun* 81:405–425
16. Pandeewari N, Kumar G (2015) Anomaly detection system in cloud environment using fuzzy clustering based ANN, *Mobile networks and applications*, pp. 1–12
17. Muthukumar B, Rajendran PK (2015) Intelligent intrusion detection system for private cloud environment. *Commun Comput Inf Sci* 536:54–65
18. Modi CN, Patel D (2013) A novel hybrid-network intrusion detection system (H-NIDS) in cloud computing. In: IEEE symposium on computational intelligence in cyber security (CICS), pp 23–30
19. Ghosh P, Mandal AK, Kumar R (2015) An efficient network intrusion detection system. In: Chapter information systems design and intelligent applications, vol. 339 of the series *Advances in Intelligent Systems and Computing*, pp 91–99
20. Ambikavathi C, Srivatsa SK (2015) Improving virtual machine security through intelligent intrusion detection system. *Indian J Comput Sci Eng (IJCSE)* 6:39

Mutative BFO-Based Scheduling Algorithm for Cloud Environment



Saurabh Singhal and Ashish Sharma

Abstract Scheduling jobs in cloud computing is a challenging issue. Utilizing the resources efficiently and gaining maximum in cloud computing is one of the goals of the cloud service provider, whereas the end-user likes to minimize the cost and time while scheduling a job on the cloud environment. The rapid increase in the demand for high computational resources has led the path for the growth of data centers in cloud data centers. With an increase in data centers, energy consumption has become a concern to the environment. The cloud environment is also characterized by unpredictable demand from users leading to a challenge to map virtual resources with jobs while satisfying the user, service provider, and environmental constraints. In this paper, a job scheduling algorithm for mapping resources to the job is proposed by applying the mutation operator to Bacterial Foraging Optimization (BFO) algorithm named BFO-Mutative algorithm. The BFO algorithm is characterized by slow search space and local convergence. The proposed algorithm overcomes these issues by mutating each particle in every iteration that improves convergence rate and thus efficiency. The proposed algorithm maps job to resources considering deadline as a constraint and reduces execution cost and time. The proposed algorithm also optimizes energy consumption in cloud data centers. Through experiments implemented in CloudSim, the results verify that the proposed algorithm reduces makespan and energy efficiency while increasing throughput.

Keywords Job Scheduling · Bacterial Foraging Optimization · Mutation · Makespan · Energy efficiency

S. Singhal (✉) · A. Sharma
GLA University, Mathura, India
e-mail: saurabh.singhal@gla.ac.in

A. Sharma
e-mail: ashish.sharma@gla.ac.in

1 Introduction

Cloud computing is the large collection of heterogeneous virtualized resources that offer resources to its end users as services via the Internet. The services offered are flexible and elastic [6] and are mapped to the user as per their demand. This has ensured that the number of users and their demand for computational and storage resources is always increasing in the cloud environment. Users pay the service provider on pay-per-use basis, eliminating the requirement of purchasing any hardware or resource. The cloud services are categorized into Software as a Service, Platform as a Service, and Infrastructure as a Service. Providing users with the virtualized resource as per demand, the service providers have grouped a large number of computers at a place known as data centers. As demand is increasing, so is the number of data centers, which in turn has increased energy consumption. In 2006 as per reports [3], in the USA energy consumed by IT infrastructure was approximately 61 billion kWh cost to USD 4.5 bn. This is approximately double the energy consumed by the IT infrastructure in 2000. And if the same trend continues, the amount of energy consumed by data centers would become 4 times by 2020 [4]. However, the data centers are utilized to 20–30% [2] of their capacity wasting a large amount of energy.

Efficient utilization of resources and minimizing energy consumption are critical issues for the service providers. The objective of the scheduling algorithm is the mapping of jobs submitted by the end-user in such a way the job is completed in minimum cost and time in the constraint of a deadline. End-users and service providers are the two main entities in the cloud environment [8]. Cloud service providers have a large pool of virtualized computational resources allocated to users with a motive of maximum profit and resource utilization, whereas the end-users want to execute their job within the deadline at minimum cost and time. A trade-off exists between energy consumption and profit at the service provider end. We proposed a mutation-based BFO to minimize makespan time and energy while scheduling jobs in the cloud computing keeping maintain deadline of the submitted jobs. To reduce energy consumption, resource utilization has to be increased [15].

One of the challenging issues in a cloud environment is the scheduling of jobs. Traditional scheduling algorithms in the cloud environment consider the meeting of QoS requirements for users and resources rather than looking to maximize profits for the two entities. Job scheduling in cloud computing plays a vital role in meeting the QoS requirements of submitted jobs and increases the utilization of resources efficiently. Job scheduling also ensures that the profit objective of both entities is fulfilled. For this, the scheduler should consider strategies for utilizing resources at the service provider efficiently and economically.

Proposed scheduling algorithm BFO-Mutative aims to schedule the submitted jobs to cloud resources in such a way that jobs are completed in the minimum time while consuming minimum energy considering deadline as a constraint. This paper discusses a scheduling algorithm BFO-Mutative that considers the deadline of

submitted jobs as a constraint while scheduling and reduces the makespan time as well as the energy consumed by data centers.

The rest of the paper is organized as follows: Sect. 2 describes the state of the art for job scheduling. Section 3 has three parts; the proposed job scheduling model of cloud computing with its components is described in Sect. 3.1. Section 3.2 discusses the objective function; Sect. 3.3 discusses the proposed algorithm. Analysis and simulation results of the proposed algorithm are discussed in Sect. 4, and in Sect. 5 the conclusion of the paper is described.

2 Related Work

Scheduling is the process of allocation of workloads to cloud resources considering various QoS parameters such as execution time, deadline, cost, and other characteristics [14]. A workload may be divisible and divided into multiple sub-workload that execute concurrently on different resources. Workload may be dependent, where the output of one workload is required for other workloads or may be independent.

In [12], the authors proposed a Continuous Double Auction based on the bargaining model which executes the scientific application in a distributed environment for a heterogeneous environment. In [18], the authors presented a dynamic resource scheduling algorithm based on the market for assigning tasks to service and optimizing tasks to the virtual machine in local data centers to centers to minimize the overall running cost while maintaining QoS parameters. The algorithm optimizes the makespan and reduces the running cost of jobs for datacenters.

In [11], the authors proposed a resource scheduling algorithm based on a bag of tasks in which the task was selected based on its arrival. The method used FCFS for scheduling resources for the task. This mechanism was able to minimize the cost, completion time, and improved CPU performance. In [1], the scheduling algorithm proposed by the authors finds all virtual machines that match the criteria for executing a given task. Then, the task was assigned to the virtual machine that has load below the threshold. In [5], the authors proposed a scheduling technique based on community awareness for improving the waiting time of the job. The scheduling is done without having any prior information on the real-time processing of different nodes. This is done in a decentralized manner.

In [7], the authors had proposed a resource scheduling technique based on feedback for reducing contention, using the notion of job preemption. A task scheduling policy based on DAG is used for resource allocation to jobs and generation of the priority list of jobs. Based on the feedback obtained, the resources are scheduled. A dynamic resource scheduling policy based on ACO had been discussed by authors in [17]. The policy while balancing the load reduces the execution time of tasks.

The authors in [10] proposed a resource scheduling technique based on ACO for distributing load to improve resource utilization and reduce cost. Instead of updating their result, the algorithm updates a single result continuously based on resource utilization. Bee life algorithm-based policy for NP-complete applications

is discussed in [13] that reduces their makespan. The makespan was reduced by optimizing resources to the distributed workloads. In [16], the authors presented a resource allocation policy for HPC applications that were based on PSO that reduces price, job rejection ratio, and makespan, and at the same time tries to maximize jobs meeting deadlines.

A scheduling algorithm based on PSO was described by authors in [9]. In the paper, the author had presented a technique that executes scientific workflow within the defined deadline. The approach is used for the identification of minimum cost configuration requirements for executing a particular workflow application without degrading the performance.

3 Proposed Work

This section has been divided into three subsections; firstly we describe the proposed job scheduling model of cloud computing with its components. The next subsection discusses the objective function used in the proposed algorithm, and finally, the proposed algorithm is discussed.

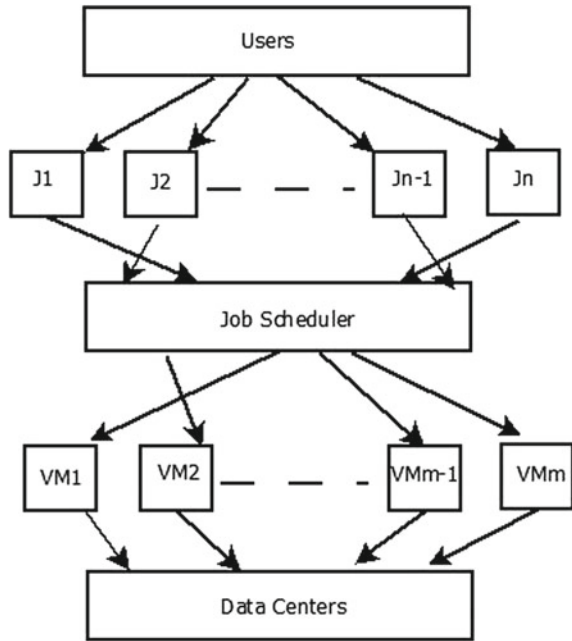
3.1 Job Scheduling Model

The Job Scheduling Model is illustrated in Fig. 1. The users spread over different locations submit their jobs to the cloud environment. In the cloud, the jobs are handled by the broker who collects the jobs coming from various users. The jobs are placed in the waiting queue, where the scheduler picks each one of them and schedules them as per the job requirement and resource availability. So, a mapping of jobs to the resources is done by the task scheduler. This mapping is the schedule for a job for its execution. As per the schedule prepared, the jobs are dispatched to the virtual machine for execution. The number of virtual machines varies as per the requirement and availability in the data centers.

3.2 Problem Formation

The executing of submitted jobs in the least amount of time is the objective of the cloud environment. However, this is a complex problem as a trade-off always exists between the users and service provider profit. For balancing this, a trade-off solution is required. In this paper, we have formulated this into the objective function and is represented in its mathematical form. To optimize QoS parameters like makespan time and execution cost, objective function has been used.

Fig. 1 Job scheduling model



Makespan time is the time duration of executing the total submitted tasks to the environment. In cloud computing, it defines the maximum time for executing cloudlets running on different data centers. Execution cost is the cost that the user has to pay as the job was executing on the resource. The service provider looks to minimize makespan time, whereas the user is concerned in minimizing the execution cost to it. Therefore, in this paper, we define the objective function based on makespan time with the deadline of jobs as a constraint. The objective function used in the proposed algorithm is

$$MS(X) = \min(\sum_{i=1}^n \sum_{j=1}^m C_{ij} * M_{ij})$$

MS = Makespan time.

C = the cost of executing job_{*i*} on virtual machine_{*j*}.

M = the mapping of job_{*i*} on virtual machine_{*j*}.

n = total number of submitted job.

m = total number of virtual machine available.

3.3 Problem Formation

In this paper, a mutative-Bacterial Foraging Algorithm is discussed. The Bacterial Foraging Algorithm imitates the behavior of bacteria. With the help of chemicals present in the environment, the bacteria perceive information about the direction of food. Using the mechanisms of locomotion, bacteria can move from one location to another. In the environment, bacterial cells are treated as agents. Depending on the cell interaction, the bacterial cells may swarm sources of food or can ignore other cells. We have applied a Gaussian-based mutation operator to the existing BFO algorithm. The proposed mutative-Bacterial Foraging Algorithm (MBFO), searches for the underutilized virtual machine or resources in data centers and then schedules a task to them. The flowchart of the proposed algorithm is illustrated in Fig. 2.

The algorithm uses three control variables l , k , and j over three parameters N_{ed} , N_{re} , and N_c respectively. The value of these control variables is initialized from zero and is in the range of parametric variables. N_{ed} maintains the number of steps in removal and dispersal, the number of reproduction steps is given N_{re} while N_c gives

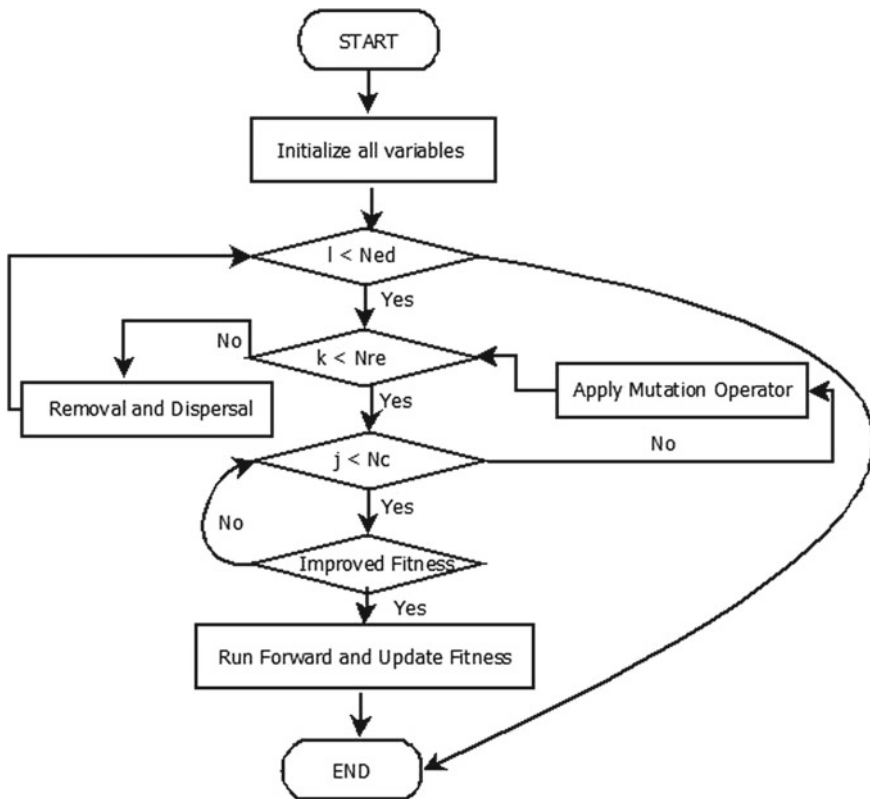


Fig. 2 Flowchart of the proposed algorithm

Table 1 Experimental setup

Entity	Parameters	Values
User	Number of cloudlets	50–300
Cloudlets	Length	500–10,000
	Number of hosts	2
	RAM	4 GB
	Storage	20 GB
	Numbers of VMs	8
	Operating system	Windows
VM	Numbers of CPUs	1
Data centers	Number of data centers	5–20

the number of chemotaxis steps. The algorithm searches for data centers and looks to improve the fitness function in the search space.

4 Result

This section discusses the results that were obtained by simulating the environment of the proposed algorithm BFO-Mutative on CloudSim. CloudSim 3.0.3 simulator is used of simulating the proposed algorithm which is configured with the Eclipse tool. The proposed algorithm minimizes the makespan time while improving fitness function.

4.1 Problem Formation

Makespan time is the time duration of executing the total submitted tasks to the environment. In cloud computing, it defines the maximum time for executing cloudlets running on different data centers. Makespan time has been considered as the parameter for measuring the performance of the proposed algorithm. The proposed algorithm is compared with other nature-inspired algorithms like BFO, ACO, and PSO. The parameters that have been considered for the experimental setup are shown in Table 1.

4.2 Problem Formation

The results obtained by running proposed algorithm are shown in Figs. 3, 4, 5, and 6 for the value of data centers 5, 10, 15, and 20, respectively. It has been observed the

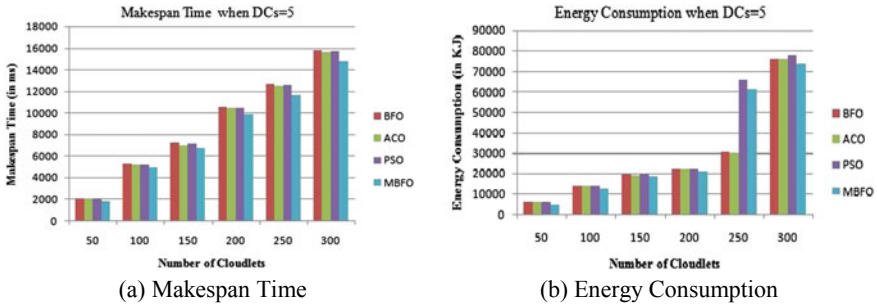


Fig. 3 Performance comparison varying value of jobs and DCs = 05

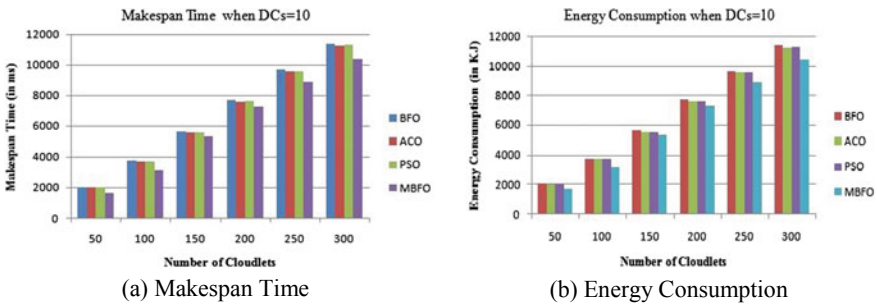


Fig. 4 Performance comparison varying value of jobs and DCs = 10

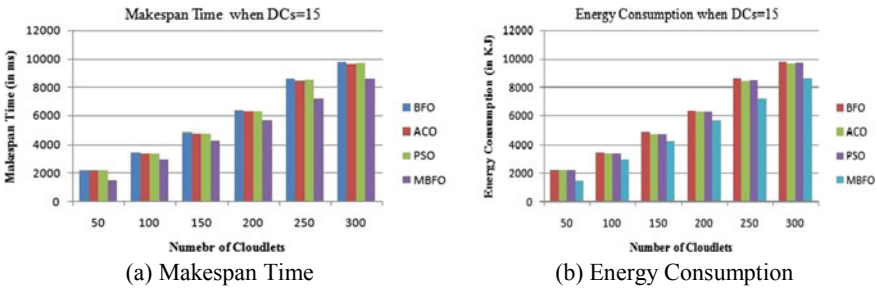


Fig. 5 Performance comparison varying value of jobs and DCs = 15

proposed algorithm gives a better value of makespan time energy consumption than the algorithm present in literature. For different sets and values of cloudlets and data centers, the result improves in the range of 5–15%.

Makespan Time

Different scenarios have been created for measuring the performance of the proposed algorithm. The number of data centers is in of range from 5 to 20 and jobs are in

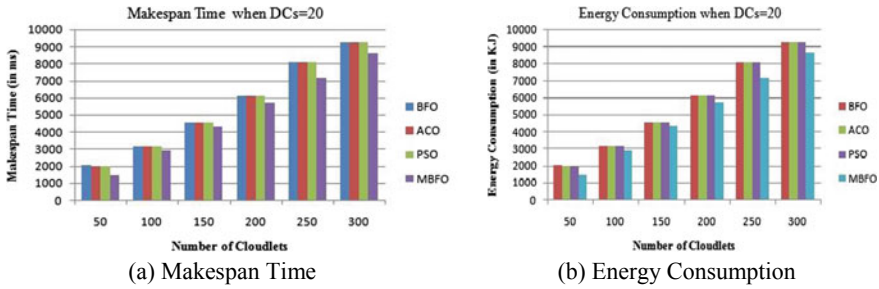


Fig. 6 Performance comparison varying value of jobs and DCs = 20

the range from 50 to 300. Various parametric values of BFO (such as bacteria and iteration) are also considered. As the number of jobs increases in data centers, the makespan also increases. The proposed algorithm is able to reduce the value of makespan time 5–10%.

Energy Consumption

The data centers in cloud computing consume a large amount of electrical energy and generate large greenhouse effect gases. This not only increases the operational cost of the data centers but also affects the environment. To minimize the consumption of energy is one of the challenges in the field of cloud computing. To minimize the consumption of energy, resource utilization must be increased so that the idle time of data centers is low. A trade-off always exists because of this between cost and energy consumption. As the number of jobs on data centers increases, resource utilization also increases, thus reducing energy consumption. The proposed algorithm is able to reduce energy consumption by 10–15%.

We have calculated the makespan time of each virtual machine by varying the data centers and keeping the number of fixed with varying job length from 500 to 10,000 million instructions.

Further, energy consumption has been calculated on varying the jobs from 50 to 300 of random lengths between 500 and 10,000 million and ranging the number of data centers from 5 to 20 with a different configuration. As the number of data centers increases, the number of machines available to execute jobs is more thus reducing the makespan time. Also, the time the machine remains busy with executing tasks is also reduced bringing the energy consumption down. This is evident from Figs. 3, 4, 5, and 6.

5 Conclusion

Scheduling in cloud computing is an important problem as the cloud service provider has to manage a large number of tasks and users in the environment. In this paper, we have discussed the problem of job scheduling in the cloud and have tried to propose

a solution. The proposed scheduling algorithm BFO-Mutative reduces the overall makespan time and improves fitness function. This algorithm manages resources effectively distributed over a large number of data centers. The proposed algorithm shows the improved result in both cases of varying data centers and varying tasks. In the future, we can look for different other parameters that help in measuring the performance of load balancing algorithms. Also, the mutation operator can be changed or could be used in combination.

References

1. Babu AA, Rajam VMA (2017) Resource scheduling algorithms in cloud environment-a survey. In: 2017 second international conference on recent trends and challenges in computational models (ICRTCCM). IEEE, pp 25–30
2. Barroso LA, Höölzle U (2007) The case for energy-proportional computing. *Computer* 40(12):33–37
3. Brown RE, Brown R, Masanet E, Nordman B, Tschudi B, Shehabi A, Stanley J, Koomey J, Sartor D, Chan P et al (2007) Report to congress on server and data center energy efficiency: public law 109–431. Tech. rep., Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States)
4. Forrest W (2008) How to cut data centre carbon emissions? Website (9.78)10–58
5. Huang Y, Bessis N, Norrington P, Kuonen P, Hirsbrunner B (2013) Exploring decentralized dynamic scheduling for grids and clouds using the community-aware scheduling algorithm. *Future Gener Comput Syst* 29(1):402–415
6. Kumar M, Dubey K, Sharma S (2018) Elastic and flexible deadline constraint load balancing algorithm for cloud computing. *Procedia Comput Sci* 125:717–724
7. Li J, Qiu M, Niu J, Gao W, Zong Z, Qin X (2010) Feedback dynamic algorithms for preemptable job scheduling in cloud systems. In: 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology. vol. 1. IEEE, pp 561–564
8. Madni SHH, Abd Lati MS, Coulibaly Y et al (2017) Recent advancements in resource allocation techniques for cloud computing environment: a systematic review. *Cluster Comput* 20(3):2489–2533
9. Netjinda N, Sirinaovakul B, Achalakul T (2014) Cost optimal scheduling in iaas for dependent workload with particle swarm optimization. *J Supercomput* 68(3):1579–1603
10. Nishant K, Sharma P, Krishna V, Gupta C, Singh KP, Rastogi R et al (2012) Load balancing of nodes in cloud using ant colony optimization. In: 2012 UKSim 14th international conference on computer modelling and simulation. IEEE, pp 3–8
11. Opreescu AM, Kielmann T (2010) Bag-of-tasks scheduling under budget constraints. In: 2010 IEEE second international conference on cloud computing technology and science. Mutative BFO-based scheduling algorithm for cloud environment, 11. IEEE, pp 351–359
12. Prodan R, Wiecek M, Fard HM (2011) Double auction-based scheduling of scientific applications in distributed grid and cloud environments. *J Grid Comput* 9(4):531–548
13. Raju R, Babukarthik R, Chandramohan D, Dhavachelvan P, Vengattaraman T (2013) Minimizing the makespan using hybrid algorithm for cloud computing. In: 2013 3rd IEEE international advance computing conference (IACC).IEEE, pp 957–962
14. Rimal BP, Jukan A, Katsaros D, Goeleven Y (2011) Architectural requirements for cloud computing systems: an enterprise cloud approach. *J Grid Comput* 9(1):3–26
15. Singhal S, Sharma A (2020) Load balancing algorithm in cloud computing using mutation based pso algorithm. In: International conference on advances in computing and data sciences. Springer, pp 224–233

16. Somasundaram TS, Govindarajan K (2014) Cloudrb: a framework for scheduling and managing high-performance computing (hpc) applications in science cloud. *Fut Gener Comput Syst* 34:47–65
17. Song X, Gao L, Wang J (2011) Job scheduling based on ant colony optimization in cloud computing. In: 2011 International conference on computer science and service system (CSSS). IEEE, pp 3309–3312
18. Wu Z, Liu X, Ni Z, Yuan D, Yang Y (2013) A market-oriented hierarchical scheduling strategy in cloud workflow systems. *J Supercomput* 63(1):256–293

Key Management Using Particle Swarm Optimization in MANET



Inderpreet Kaur, Parth Pulastiya, and Vivek Anil Pandey

Abstract With the accretion in the demand for robustness in the cryptographic algorithms, we are compelled to find a practical answer to this desire of reliable and dynamic mode of communication. Though many of the cryptographic devices possess mobility and hence are engaged within a mobile ad hoc network (MANET), yet the security goals still remain the same, viz. availability, integrity, confidentiality, non-repudiation, and authenticity. This work is an attempt to devise a secured and authenticated procedure for key management employing asymmetric cryptography. The following procedure would be employed in MANETs. This project is an amalgamation of the two most researched topics in the current cyber world. The approach presented here is an attempt where we exploit the randomness associated with the crossover and mutation process of the genetic algorithm; moreover, to make it unbreachable, a scheduled random factor is included which is derived using the particle swarm optimization. At last, the scheme is implemented in Python 3 and implemented for encrypting a plaintext document into ciphertext and decrypting the ciphertext into the respective plaintext. Furthermore, the algorithm devised here is common and can be further implemented for any secured transmission of data.

Keywords MANET · Genetic Algorithm · PSO · Crossover · Mutation · Asymmetric Key · Encryption · Decryption

I. Kaur · P. Pulastiya · V. A. Pandey (✉)

Galgotias College of Engineering & Technology (GCET), Uttar Pradesh, Greater Noida, India

e-mail: Viv3k.pand3y@gmail.com

I. Kaur

e-mail: Inderpreet.kaur@galgotiacollege.edu

P. Pulastiya

e-mail: Parthpulastiya@gmail.com

1 Introduction

Genetic algorithm (GA) is the heuristic procedure that is inspired by the Darwinian theory of Natural Selection and Evolution. They are an implementation of Darwin's principle of survival of the fittest and normal genetics.

Cryptography is the conventionally exploited mechanism for validation of the authentic user in addition for secure information. Devoid of the information of the appropriate decrypting keys, the transformation is infeasible considering both in time and cost. [1–4]. In this work, we employ the genetic algorithm along with the particle swarm optimization for generating key–value pair for asymmetric key cryptography that is solely a unique procedure and is not a publicly available technique like other traditional cryptographic methods, i.e., RSA and DES algorithms. Hence, if the trespasser could get access to the key, it would be computationally infeasible to break the proposed algorithm which makes the encryption process so intense that it preserves the confidentiality and authenticity of the data.

1.1 Cryptography

Cryptography is the progression of obtaining the key assigned enciphering a plaintext into ciphertext, and hence, this key is exploited in finding the plaintext of ciphertext. Cryptography will concede us various keywords produced through the cipher method. This is done by transforming the readable data (plaintext) into a meaningless form (ciphertext). Almost every cryptographic systems are broadly divided into two self-governing dimensions:-

Varieties of Operations—Every encrypted algorithm is generally dependent on two basic standards, substitution and transposition. It means that all there should be no loss of data at any point, and all the proceedings should be reversible.

Key Utilized—The length of the key that is used to encrypt the information determines the security's strength [5–7].

In this work, the GA algorithm picks a character from the plaintext as a block of four characters for each iteration of the algorithm. Each parent is a string of two characters. Here, we have chosen two vectors of 14 bits each as a parent for the purpose of reproduction. In view of the fact that the complication is of encryption, there is no exceptional preference provided to any specific selection scheme. The entire vectors are elected in sequence in accordance with their order of appearance in a text file. The asymmetric key generated from the above discussion consists of three basic components:

- 1 Four arbitrarily produced crossover points in the limit 0–13; all different and then sorted.
- 2 Three arbitrarily produced mutation points in the limit 0–13.
- 3 Permutation factor generated with PSO which will be in the range of 1–7.

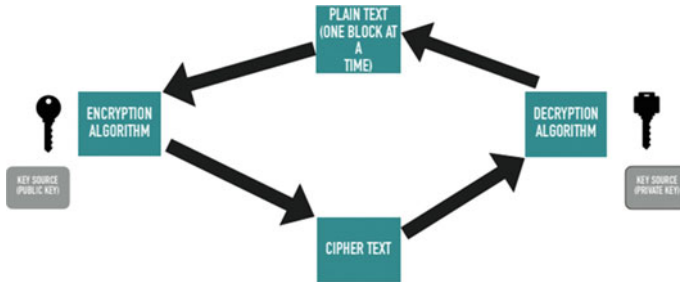


Fig. 1 Block diagram of asymmetric key encryption

The fervor of the key is determined by the number of crossover and mutation points, and the entire process is depicted in the Fig. 1:-

Genetic Algorithm

In computer science, GA is an evolutionary approach of meta heuristics that is inspired by the theory of natural selection and is a Super set of the evolutionary algorithms. They are used to providing high-quality solutions for finicky problems by relying on biologically inspired operants. A characteristic GA requires two conditions.

Genetic demonstration of the solution provides fitness function for the purpose of evaluating the solution. GA consists of three fundamental operations.

1. Selection,
2. Crossover and
3. Mutation

The first action is to select the individuals the reproduction of the individual. In our algorithm, we have selected two vectors of 14-bit each be encrypted; hence, no special preference is given and are picked sequentially in the order of their appearance in the plain text data [8–10].

Crossover is the foremost phase of the genetic algorithm. After the selection of the parents for mating, a point is chosen to perform the crossover from with the genes of the parents. [11]. After performing the crossover operation, to prevent the algorithm from being fascinated into a local minimum, the genetic algorithm comprises the mutation process. In our problem, we have employed the flipping method for crossover and compliment method for mutation; thus, the corresponding child chromosome is produced.

1.2 Particle Swarm Optimization

Particle(s) swarm optimization (PSO) is a ciphering procedure which is composed of two components of science, i.e., computer science and social science. The foundations of PSO are based on the basis [12].

Social Concepts: It can be referred to as “human intelligence results from social interactions.”

Swarm Intelligence: It may be illustrated as the collective behavior of decentralized, self-organized schemes. These systems can be natural or artificial.

In PSO, the term “particles” is referred to the members of the population group that needs an optimized result for locating the food for themselves. [13].

2 Literature Survey

In literature to date available, many types of genetic algorithm have been proposed. Nowadays, the security of the data lime becomes the main issue while communicating through mobile devices. In such cases, these devices must be secured. Upon examining the various utilities of PSO ill respective fields of computing, we conclude that PSO is swift and straightforward to implement in cryptography as well. But there are very few papers that are using the randomness of the crossover and mutation for the encryption process. And it is well said that the best predictor of the future fulfillment is its past.

Thus, one can refer PSO a comprehensive prescription in cryptography particle(s) swarm optimization which has been scrutinized for later intensification in such a way that the modish analysis might be concentrated to yield a more reliable explication by augmenting the efficacy and lessening the constraints. Strong key generation is very important for secure data transfer [14].

3 Proposed Algorithm

The algorithm for the encryption process by means of the engagement of GA and PSO is given below:

1. Produce four crossover points and three mutation points in the limit 0–13.
2. Arrange the crossover points in ascending order.
3. Produce the permutation factor by using particle swarm optimization. Apply PSO in the previously generated random numbers for generating the subsequent numbers of the public key.
4. These subsequent numbers are called a random factor and are concatenated after every point using the formula stated below: random factor point 4-permutation factor/mod 14

5. Produce the public key depending on crossover points, mutation points, permutation factor and random factors.
6. Produce the private key in accordance with the public key.
7. Write the key pairs to their respective files.
8. Read four blocks of a single character. NOTE: if there are not sufficient characters then put O into the block.
9. Go to Step 4.
10. Add on the obligatory number of spaces, in order to generate a 7-bit block for each character using their binary codes.
11. Apply translation on the blocks generated. The translation is the conversion of the characters into ASCII and the further into 7-bit binary-coded representation.
12. Execute the crossover and mutation procedures upon the parent blocks that are present in the mating pool as described in Step II.
13. Inscribe the encrypted blocks to the file.
14. If the file encloses more data subsequently go to step 8.
15. END

The functioning of the algorithm is demonstrated four characters represented as 14-bit characters from a text file to be encrypted. Consider the two blocks which contain characters represented as binary coded numbers as:

FIRST BLOCK: $-b_0, b_1, b_2, b_3 \dots b_{13}$.

SECOND BLOCK: $-c_0, c_1, c_2, c_3 \dots c_{13}$.

in which each b_i and c_i indicates a binary character which is the part of character. For example, let us say the text is CSEB.

$C = 01000011 \quad S = 01010011 \quad E = 01000101 \quad B = 01000010$.

In this, b_i constitutes the C and S , whereas c_i constitutes the letters E and B . For this reason, the length of the key is 14-bits here, and it is explicitly dependent on the amount of crossover and mutation points. Each alternating character starting from the first character is the part of the private key which would be helpful for decrypting the ciphertext and storing it into their respective files.

4 Experiment and Results

The cryptographic algorithm formulated is implemented in Python 3 which is exploited for encrypting and decrypting the text file simultaneously. The work is then extended for the purpose of encrypting the content and decrypt the ciphertext

again to plaintext. The text document is accessed through the Python application. The various files used in the software package are described in the table below.

4.1 Enciphering

The process of converting a message or a piece of text into a coded form. Python Tkinter Interface for Asymmetric encryption is revealed below (Fig. 2):

In the above figure, we can clearly see the content of the plaintext file along with the public key generated using our proposed algorithm which encompasses the symbiosis of genetic algorithm along with the particle swarm optimization.

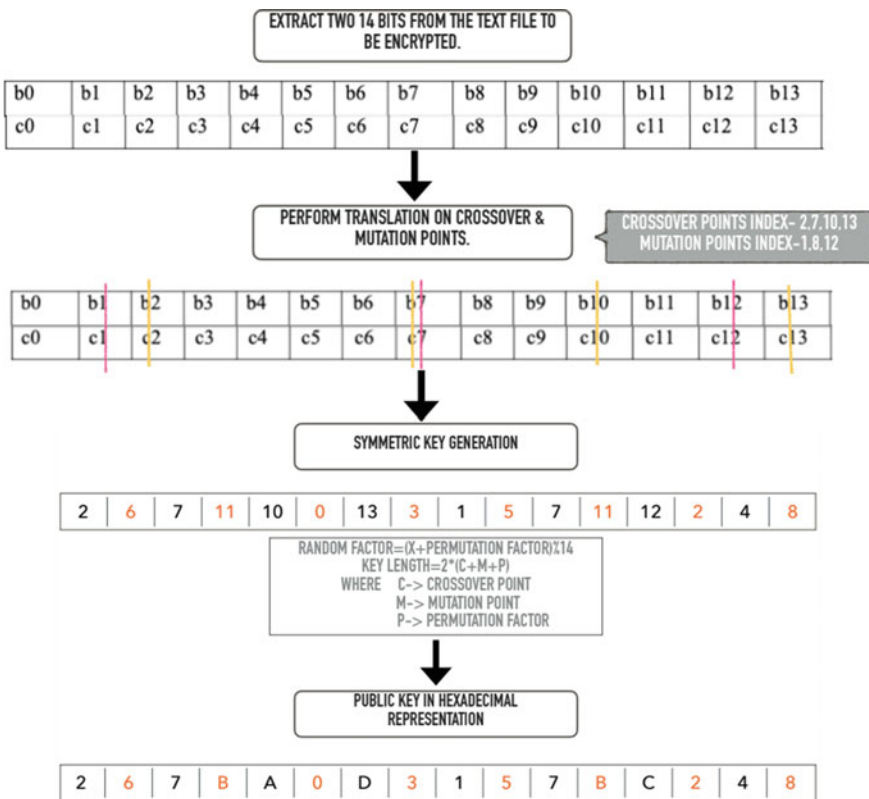


Fig. 2 Public key encryption using GA and PSO

5 Deciphering

It is process of reverse conversion of a coded text, or an encrypted message into the plaintext for making it sensible so that the receiver could understand the message. The decrypted text as well as the private key is shown below.

Subsequent to decryption with the assistance of the related private key, the similar plaintext is restored back. It is to be noted that the hexadecimal key–value pair utilized in the cryptographic procedure is exposed in the table.

Thus, the results are promising and depicts that the algorithm performs encryption and decryption with complete robustness and authentication.

6 Conclusion and Future Scope

The proposed scheme by means of making use of the arbitrariness implicated in the crossover and mutation process for the purpose of producing an asymmetric key pair for encryption and decryption of messages. The points of the crossover and mutation along with the permutation and random factor are employed in the generation of a public and private keys. Here, we have utilized four crossover points, three mutation points, a permutation factor and subsequent random factor concatenated. The length of the key is 16 bits. The scheme is additionally made stronger by means of making it complicated to break through permuting the asymmetric key by means of a permutation factor derived by using PSO. The arbitrariness in cooperation with permutation makes the scheme robust and complicated to break. To end with, the scheme is implemented in Python and executed for the encryption and decryption of a text file and a text document.

Our future work includes working out a modus operandi to compute the strengths of the algorithm by means of the displacement of every character in the original string, and also for further security, we can employ biometric systems for the application software.

Acknowledgements We signify a deep sense of gratitude to our project guide, Dr Inderpreet Kaur, for her assistance for our perusal on the assigned topic. We are highly obliged to every single person who laid their efforts in the reviewing process and accommodated us in reaching our set time limit.

References

1. Lafe O (1997) Data compression and encryption using cellular automata transforms. *Eng Appl Artif Intell* 10(6):391–581
2. Sun Y, Liu KJR (2007) Hierarchical group access control for secure multicast communications. *IEEE/ACM Trans Netw* 802(6):1514–1526

3. Lu W, Traore I (2004) Detecting new forms of network intrusion using genetic programming. *Intelligence* 20(3):475–494
4. Kwok HS, Tang WKS (2007) A fast image encryption system based on chaotic maps with finite precision representation. *Chaos Solitons Fractals* 32(4):1518–1529
5. Lin H-Y, Chiang T-C (2011) Efficient key agreements in dynamic multicast height balanced tree for secure multicast communications in Ad Hoc networks. *EURASIP J Wireless Commun and Netw* 1
6. Mortazavi SA, Pour AN, Kato T (2011) All efficient distributed group key management using hierarchical approach with diffie-Hellman and symmetric algorithm: DIISA. In: 2011 international symposium on computer networks and distributed systems (CNDS)
7. Wong CK, Gouda M, Lam SS (1998) Secure group communications using key graphs. *ACM SIGCOMM Comput Comm Rev* 28(4):68–79
8. Zeng G, Wang B, Ding T, Xiao L, Mutka M (2010) Efficient multicast algorithms multichannel wireless mesh networks. *IEEE Trans Parallel Distrib Syst* 21(1):86–99
9. Sherman AT, McGrew DA (2003) Key establishment in large dynamic groups using one-way function trees. *IEEE Trans Softw Eng* 29(5):444–458
10. Lobo LMRJ, Chavan SB (2012) Use of genetic algorithm in network security. *Int J Comput Appl* 53(8):1–7
11. Lin JC, Lai F, Lee C Efficient group key management protocol with one-way key derivation. In: *Proceeding IEEE Conference on Computer Networks*, pp. 336–343
12. Challal Y, Seba H (2005) Group key management protocols: a novel taxonomy *J Inf Technol* 2(1):105–118
13. Goldberg DE (1989) *Genetic algorithms in search, optimisation, and machine learning*. Pearson Education
14. Pandey VA, Pulastiya P, Inderpreet K, Rastogi S (2020) A survey on key management using particle swarm optimisation in MANET

Author Index

A

Adarsha, H. P., 463
Agarwal, Aakash, 285
Aggarwal, Reshu, 71
Aggarwal, Vikrant, 235
Agrawal, Aman, 337
Agrawal, Pragya, 293
Agrawal, Reeya, 81
Ahire, Kiran, 313
Ahlawat, Vivek, 495
Ahuja, Smriti, 21
Arekar, Maulika, 359
Arora, Sanya, 41, 51
Asha, 59
Askani, Priyanka, 293
Aswin Kumer, S. V., 411

B

Bagora, Kshitiz, 21
Bagul, Manali, 313
Bangad, Rahul, 371
Bansal, Atul, 455
Bansal, Kriti, 337
Bansal, Mohak, 21
Bansal, Nancy, 337
Bera, Tushar Kanti, 271
Bhargava, Anuja, 455
Bhongade, Aashish, 371

C

Chahar, Meetendra Singh, 517
Chandrika, C. P., 463
Chaturvedi, Amit Kumar, 517, 529, 579
Chavan, Appasaheb, 463

Chavhan, Gajanan H., 423, 507
Chavhan, Pranali G., 423
Chhattani, Nishant, 349

D

Dahiya, Omdev, 549
Damodara, Komal, 11
Dandawate, Yogesh H., 507
Deshpande, Mangesh S., 507
Dhanawate, Swapnil, 313

G

Gaidhane, Vilas H., 153
Gangwar, Himanshu, 285
Garg, Anurag, 3
Garg, Pardeep, 171
Gaurav, Prashant, 111
Gayathri, P., 323
Ghose, Udayan, 385
Ghule, Neeraj, 371
Goswami, Aditya, 101
Goyal, Vishal, 81
Gupta, Eshita, 3
Gupta, Karan, 153
Gupta, Pari, 445
Gupta, Shruti, 349
Gupta, Sindhu Hak, 31

J

Jain, Amit, 235
Jayachitra, T., 165

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 192, <https://doi.org/10.1007/978-981-33-6546-9>

K

Kallimani, Jagadish S., 463, 557, 567
 Kanakaraja, P., 247, 259, 411
 Kapil, Spardha, 171
 Kaur, Inderpreet, 601
 Kaur, Sanmukh, 11, 41, 51, 201
 Kaushik, Tarun, 11
 Khandelwal, Utkal, 539
 Kotamraju, Sarat K., 247, 259, 411
 Kumar, Ajitesh, 145
 Kumar, Ashwani, 201
 Kumar, Divesh, 495
 Kumari, Mona, 145
 Kumari, Shalini, 445
 Kumari, Varsha, 385
 Kumar, Jitendra, 189
 Kumar, Manish, 93
 Kumar, Pranjul, 201
 Kumar, Punit, 579
 Kumar, Rakesh, 235
 Kumer, S. V. Aswin, 247, 259

M

Mahalle, Parikshit N., 423
 Maheshwari, Varun, 71
 Mamatha, H. R., 293
 Mangaonkar, Nikhita, 359
 Mathur, H. D., 211, 223
 Mishra, Akhilesh Kumar, 211, 223
 Mishra, Puneet, 211, 223

N

Nadipalli, L. S. P. Sairam, 247, 259
 Nagabhushan, Aparna, 463
 Nair, Nivedita, 11
 Narayan, Yogendra, 495
 Naugarhiya, Alok, 123, 135
 Navputra, Arun, 71
 Nayak, Ranjitha, 293
 Niharika, Neha, 123
 Nilofer, N. S., 431

P

Pabla, Harmeet Singh, 235
 Pandey, Nikita, 171
 Pandey, Vivek Anil, 601
 Pandit, Pooja, 71
 Pandoi, Deepika, 399
 Panicker, Suja Sreejith, 313, 323
 Pathak, Pooja, 285
 Patil, Balaji, 371

Prashant, 111
 Priyadarshani, Kumari Nibha, 135
 Priyadarshini, Rashmi, 165
 Pulastiya, Parth, 601
 Pundhir, Sandhya, 385
 Puri, Ridhima, 31
 Purwar, Vaibhav, 181

R

Ragavendran, U., 21
 Rajawat, Asmita, 31
 Ramesh, N. V. K., 411
 Ramkumar, R., 431
 Ranjan, Ashish, 371
 Rathor, Sandeep, 445
 Rawal, Arpana, 349

S

Sachdeva, Jaideep, 349
 Sahoo, Sourav, 303
 Sairam Nadipalli, L. S. P., 411
 Sajini, G., 557, 567
 Samal, Manisha, 41, 51
 Sarwate, Aditya, 303
 Sehgal, Sahil, 303
 Sengar, Praveen, 529
 Shafi, Pathan Mohd, 423
 Sharma, Ashish, 589
 Sharma, Himanshu, 483
 Sharma, Kalpana, 517, 529, 579
 Sharma, Nimisha, 31
 Sharma, Rohini, 101
 Sharma, Shubham, 71
 Shivani, 473
 Shukla, Atul Kumar, 71
 Singhal, Saurabh, 589
 Singh, Mahendra, 71
 Singh, Navaneet Kumar, 123
 Singh, Navdeep, 473
 Singh, Rishu, 445
 Singh, Sangeeta, 111, 123, 135
 Singh, Shradhya, 123
 Singh, Trilok Pratap, 539
 Solanki, Kamna, 549
 Sri Kavya, K. Ch., 247, 259

T

Thakkar, Manisha, 303
 Thakur, Bharat Mohan, 337
 Tomar, V. K., 101
 Tripathi, Sweta, 181

V

Varshney, Kratika, [181](#)
Venkatesh, Vivek, [359](#)
Vishnu Kartik, M., [41](#), [51](#)

W

Wani, Anurag, [303](#)