Lecture Notes in Electrical Engineering 948

Rajeev Agrawal
Chandramani Kishore Singh
Ayush Goyal
Dinesh Kumar Singh *Editors*

Modern Electronics Devices and Communication Systems

Select Proceedings of MEDCOM 2021

Springer

Lecture Notes in Electrical Engineering

Volume 948

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases. ****

Rajeev Agrawal · Chandramani Kishore Singh ·
Ayush Goyal · Dinesh Kumar Singh
Editors

Modern Electronics Devices and Communication Systems

Select Proceedings of MEDCOM 2021

Springer

Editors
Rajeev Agrawal
G. L. Bajaj Institute of Technology
and Management
Greater Noida, India

Ayush Goyal
Department of Electrical Engineering
and Computer Science
Texas A&M University—Kingsville
Kingsville, USA

Chandramani Kishore Singh
Indian Institute of Science Bangalore
Bengaluru, India

Dinesh Kumar Singh
G. L. Bajaj Institute of Technology
and Management
Greater Noida, India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-19-6382-7 ISBN 978-981-19-6383-4 (eBook)
https://doi.org/10.1007/978-981-19-6383-4

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

MEDCOM 2021 Luminaries

Chief Patron

Dr. Ram Kishore Agarwal, Hon'ble Chairman, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India

Patron

Mr. Pankaj Agarwal, Hon'ble Vice-Chairman, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India

General Chair

Dr. Rajeev Agrawal, Director, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India

Technical Program Committee Chair

Dr. Subodh Wariya, Institute of Engineering and Technology, Lucknow, U.P., India

Conference Chair

Dr. Satyendra Sharma, HOD, ECE, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India

Organizing Secretary

Dr. Dinesh Kumar Singh, ECE, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India

Co-organizing Secretary

Dr. Amrita Rai, ECE, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India

Conveners

Dr. Piyush Yadav, ECE, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India
Dr. Shilpa Choudhary, ECE, G. L. Bajaj Institute of Technology and Management, Greater Noida, U.P., India

Preface

Innovation and research in a broad area of modern electronics and communication engineering needs to enhance its capabilities for recent challenges. Recent advances in different domains like sensor technologies, electronic devices, wireless communications, signal processing, computer vision medical image, data processing and globalization of digital society across various social areas have intensified the growth of sustainable electronics and communication systems. To effectively maintain a sustainable environment, it is constantly needed to meet the quality of service requirements of the emerging smart ICT applications. In this perspective, the book presents select proceedings of the International Conference on Modern Electronics Devices and Communication Systems (MedCom 2021) that brings together the state-of-the-art research works to propose novel architectures, models, and algorithms for solving the emerging challenges in almost all the areas of modern electronics devices and communication systems.

The book aims to bring teachers, researchers, scientists, engineers and scholars together, where they will exchange and share their experiences, new ideas and research results about all aspects of mentioned areas. This book will also present a strong fostering integration between modern electronics devices, communication and computational technologies. The global dependence in almost every sphere of human activity on computational analysis in modern era cannot be over emphasized. These trends are bound to increase with more and more new applications in the field of finance, marketing, health care, education, etc., which are coming under the gambit of computers.

Greater Noida, India — Rajeev Agrawal
Bengaluru, India — Chandramani Kishore Singh
Kingsville, USA — Ayush Goyal
Greater Noida, India — Dinesh Kumar Singh

Contents

A Review of State of Art Techniques for 3D Human Activity Recognition System 1
Bhavana Sharma and Jeebananda Panda

Comparative Analysis for Email Spam Detection Using Machine Learning Algorithms 11
Gayatri Gattani, Shamla Mantri, and Seema Nayak

Post-quantum Cryptography: A Solution to the Challenges of Classical Encryption Algorithms 23
Shagun Sharma, K. R. Ramkumar, Amanpreet Kaur, Taniya Hasija, Sudesh Mittal, and Bhupendra Singh

A Survey on Performance Analysis of Different Architectures of AES Algorithm on FPGA 39
Taniya Hasija, Amanpreet Kaur, K. R. Ramkumar, Shagun Sharma, Sudesh Mittal, and Bhupendra Singh

Designing Adiabatic Techniques for Logic Circuits 55
Himani Sharma, Nidhi Sharma, and Surya Deo Choudhary

Accuracy and Performance of Heart Failure Prediction System by Using Backpropagation Network (BPN) 67
Pawan Kumar Singh, Piyush Yadav, Pankaj Sharma, Virendra Pal Singh, and Durgesh Kumar

Analysis and Implementation of Microsoft Azure Machine Learning Studio Services with Respect to Machine Learning Algorithms 91
Soni Singh, K. R. Ramkumar, and Ashima Kukkar

Design and Analysis of Dual-Band Compact Antenna for 5G Communication Applications 107
Pankaj Jha, Seema Nayak, Samarth Golus, Prakriti Jaiswal, and Ayushi Kumari

Design and Analysis of Graphene-Based Metasurface Absorber for Temperature and Refractive Index Sensing in THz Spectrum 115
Ramkrishna and Rajveer S. Yaduvanshi

Optimizing and Validating Performance of 40 Gbps Optical System 131
Shradha Gupta, Shilpa Choudhary, Kanojia Sindhuben Babulal, and Sanjeev Sharma

Medical Image Analysis Using Soft Computing Feature Selection and Classification of Skin Cancer 141
Birendra Kumar Saraswat, Shipra Srivastava, Samender Singh, Arun Kumar Takuli, and Aditya Saxena

Comparative Study of Fuzzy, PID, and 2DOF-PID Controller for Liquid Level Control System 157
Gunjan Taneja, Anjali Jain, and Neelam Verma

IOT-Based Self-nourishing System for Plants 173
Asha Rani Mishra, Sanjeev Kumar Pippal, Bhavika Rajpal, Arshita Garg, and Sushant Singh

Social Distance Alerting Device for COVID-19 185
Lakshay Bansal, Anant Garg, Ishanvi Prasad, and Ravendra Singh

BWDS-Based Spatially Dabble Couple Horizontal-Vertical 4-Port PIPO DRA with Poly-Pattern Diversity 197
Ramkrishna and Rajveer S. Yaduvanshi

Design of U-Shaped MIMO Antenna for Wi-Fi Application 211
Om Prakash and Prabina Pattanayak

Comparative Study of Parametric Disturbances Effect on Cart-Inverted Pendulum System Stabilization 217
Neelam Verma and Sudarshan K. Valluru

Efficient Design of QCA ADD-SUB Subsystem for Future Microprocessors 231
Ashish Mohan, Shilpee Patil, and V. K. Pandeyr

Predication of Diabetes Through Machine Learning Technique 239
Arun Kumar Singh, Ashish Tripathi, Pushpa Choudhary, Prem Chand Vashist, and Arjun Singh

Study of Gate Workfunction Modulated FinFET with Effect of TID 253
Abhishek Ray, Alok Naugarhiya, and Guru Prasad Mishra

Prediction of Accuracy and Performance of Crop Production by Using Machine Learning Algorithm 261
Ramashankar Yadav, Pawan Kumar Singh, Ajay Sikandar, Manoj Kumar Tyagi, and Dinesh Kumar Yadav

Cognitive Load Classification in Alcoholic and Control Subjects Performing Delayed Matching Task Using Optimizable Classifier Networks 283
Sweeti

Energy Detection in Cognitive Radio Networks to Optimize the Spectrum Utilization Through Cooperative Spectrum Sensing Using Weights 291
Rahul Gupta and P. C. Gupta

Real-Time Tracking of Health Parameters Using IOT and Data Logger: Application to COVID-19 303
Insha Zehra, Rajat Mehrotra, M. A. Ansari, Rajeev Agrawal, and Aruna Pathak

Real-Time Facial Detection Using MATLAB (Utilizing VJ Method) 315
Preeti Sharma and Suryansh Goel

Hardening of EMP in Shielded Enclosure for Radiated Fields 323
Pathala Venkata Sai Charishma and P. V. Y. Jayasree

Review on Optimizing Text-Video Retrieval Using CLIP (Contrastive Language-Image Pre-training) 339
Shelly Sinha, Anupriya, and M. Rathna Chary

Argument Mining: A Categorical Review 353
Sakshi Arora, Ajay Rana, and Archana Singh

Optimizing Text: Representation of Similarity Using Machine Learning 369
Piyush Yadav, Siddharth Rajput, and Archana Singh

Experimental Analysis on Aging Evaluation of Transformer Oil via Data Collection 383
Harkamal Deep Singh and Jashandeep Singh

A Systematic Survey of Automatic Text Summarization Using Deep Learning Techniques 397
Madhuri Yadav and Rahul Katarya

An IoT-Based Smart Garbage Collection and Monitoring System 407
Taranpreet Singh Ruprah, Pramod Dharmadhikari, and Krunal Pawar

Estimation of Accuracy in Gender Identification Based on Voice Signals Using Different Classifiers 415
Abhishek Singhal and Devendra Kumar Sharma

Detection of Covid-19 Using CT-Scan Images and Deep Transfer Learning 425
Potluri Sairaj and Kritika Jain

Controlling Load Frequency in a Hybrid Power System Including Electrical Vehicles 435
Nagendra Kumar, Priyanka Datta, Kailash Sharma, Rajat Mehrotra, and Jay Singh

Lung Cancer Detection Using CT Scan Images 447
Kilari Veera Swamy, Jatavath Priyanka, Duggi Sri Venkata Likhitha, and Sreeja Veeragoni

A Wideband Microstrip Line-Based Balun Structure for High Power Amplifier Applications 457
Manishankar Prasad Gupta, Pradeep Gorre, Sandeep Kumar, and Hanjung Song

Principal Component Analysis of LISS—III Images Using QGIS 467
V. Vijayalakshmi, D. Mahesh Kumar, and S. C. Prasanna Kumar

A Review on Different Preprocessing and Feature Extraction Technique for SSVEP BCI Inference System 479
Mukesh Kumar Ojha and Dhiraj Gupta

Digital IC Tester Using Advanced Microcontroller Board Interfaced with GUI MATLAB 489
Shylaja V. Karatangi, Amrita Rai, Puneet Kumar Mishra, Krishanu Kundu, and Reshu Agarwal

A Simulation-Based Model for Detecting Fall of Elders Using Proteus 499
R. Harsha, Swikriti Gupta, and K. N. Veena

Simulation of a Microgrid with OpenDSS an Open-Source Software Package 513
Anjali Jain, Ashish Mani, and Anwar S. Siddiqui

Heart Condition Forecasting Using High-Accuracy ANN Algorithms 531
Pranshi Gupta, Piyush Yadav, Rajeev Agrawal, Varun Sharma, Ritik Dixit, Shubhashish Varshney, and Komal Kashish

Internet of Thing (IoT): Routing Protocol Classification and Possible Attack on IOT Environment 543
Bhupendra Patel, Khushi Patel, and Anand Patel

Operating Frequency Prediction of Annular Ring Microstrip Antenna Using Extreme Learning Machine 553
Kanhaiya Sharma, Ganga Prasad Pandey, Dinesh Kumar Singh, Krishna Chaya Addagarrala, and Shailaja Salagrama

Challenges and Future Perspectives of Low-Power VLSI Circuits: A Study 561
Paramjeet Chauhan and Saptarshi Gupta

About the Editors

Rajeev Agrawal is currently working as a professor and director at G. L. Bajaj Institute of Technology and Management, Greater Noida, India. Has an illustrated experience of more than 25 years in teaching and research, holds a B.E. degree in Electronics Engineering and M.Tech. degree in System Engineering. He received Ph.D. in the area of Wireless Communication Channels from the School of Computer and System Sciences, JNU, New Delhi. He was visiting professor at Kennesaw State University, Georgia, the USA under a joint research project in the area of Remote Patient Monitoring and Medical Imaging in the years 2010 and 2011. His research areas include planning and performance analysis of wireless networks and medical image analysis for automated diagnosis. He has more than 56 publications in international journals and proceedings. He has been awarded by various state and national agencies for his contribution to research and academics. He is also serving as a reviewer for several reputed international journals.

Chandramani Kishore Singh is currently an assistant professor with the department of electronic systems engineering, Indian Institute of Science, Bengaluru. He received M.E. and Ph.D. degrees in electrical communication engineering from the Indian Institute of Science, Bengaluru, India, in 2005 and 2012, respectively. He worked as a wireless communications engineer in ESQUBE Communications Solutions Pvt. Ltd., Bengaluru, from 2005 to 2006. He was a post-doctoral researcher with TREC, a joint research team between INRIA Rocquencourt and ENS Paris from 2012 to 2013, and with the Coordinated Science Laboratory, the University of Illinois at Urbana Champaign, USA, from 2013 to 2014. His research interests include communication networks, data centers, and smart grids. He has published more than 30 papers in respected international journals and conferences. Dr. Chandramani Kishore Singh has also received the Microsoft Research India Rising Star Award in 2011.

Ayush Goyal is an Assistant Professor in the Department of Electrical Engineering and Computer Science at Texas A&M University, Kingsville, Texas. He completed his B.S. in electrical engineering from Boise State University (BSU), Boise, USA,

and Ph.D. in Computer Science from Oxford University, Oxford, UK. He was a post-doctoral research fellow in the department of computer science, Tulane University, New Orleans, US. His areas of research interest are neurological disease prediction, brain MRI image analysis, and bioinformatics for connected healthcare. He has more than 40 publications in international journals and proceedings. He was awarded the Clarendon scholarship to pursue Ph.D. at Oxford University.

Dinesh Kumar Singh is an Associate Professor in the Department of Electronics and Communication Engineering, G. L. Bajaj Institute of Technology and Management. He completed his B.E. from Kumaon Engineering College, Almora, and M.E. from Rajeev Gandhi Proudyogiki Vishwavidyalaya, M.P. He got his Ph.D. degree from the Indian Institute of Technology Dhanbad. He has published several research papers in international journals and conferences. Dr. Singh's teaching and research areas are microwave engineering, antenna and wave propagation, and electromagnetic field theory. Dr. Singh is presently supervising 2 Ph.D. students.

A Review of State of Art Techniques for 3D Human Activity Recognition System

Bhavana Sharma and Jeebananda Panda

Abstract Recognizing human activities through video sequences and images is still a challenge due to background jumble, partial occlusion, changes in scale, viewpoint, lighting and appearance. A human activity classification technique has been comprehensively reviewed by the researchers. We have categorized human activity methodologies with object detection and feature extraction along with their subcategorization, advantages and restrictions. Moreover, we provide a comprehensive analysis of the existing, publicly available human activity datasets with applications and examine the prerequisites for an ideal human activity recognition dataset. At last, we present some open issues on human activity recognition and characteristics of future research directions.

Keywords Human activity recognition · Object detection · Feature extraction · Object classification · HAR datasets

1 Introduction

Population of elders is increasing with a rapid rate in most of the western countries and hence the challenges [1]. If we convert this in the form of percentage, by 2050, it will be 30% in Europe and China which is maximum globally and then 20.2% in United States of America (USA). This fact was established after a survey of WHO, that on an average, 28–35% of elderly people meet with an accident, because of falling, annually. According to WHO report, 37 million fall accidents are reported every year out of which 64.6 thousand people lose their life because of these accidents [2, 3]. In today's modern world of nuclear families, elderly people are living alone at their homes and they are more prone to meeting with such accidents while staying

B. Sharma (✉) · J. Panda
Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi, India
e-mail: bhavanasharma.iec@gmail.com

J. Panda
e-mail: jpanda@dce.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_1

at home. Hence, the study of fall detection, to improvise the science of detection, is more crucial and important [4–6]. In human activity recognition, to solve the existing and upcoming challenges toward activity recognition, researchers are using different techniques to beat the challenges of analysis. Based on the defined categories in different areas, methods and approaches may differ from one to another. Most commonly used cases of activity recognition is in medical and surveillance, where we talk about the devices and systems which are beneficial to humankind, in improving the life by mitigating the threats. By having a direct impact on saving lives, researchers are very keen to work in the field of video surveillance.

Human activity recognition through a comprehensive survey covers human activity recognition, i.e., 2D and 3D HAR based on RGB, depth and skeleton-based methodologies. The literature is updated with the application of recent advances field of human action recognition in Sect. 2. A structured arrangement of 2D and 3D object detection techniques has been discussed in Tables 1 and 2 highlighting different feature extraction techniques. Organization of our survey is as follows. Section 2 provides a panoramic summary of the related state-of-the-art survey works in the area of abnormal human action recognition followed by paper count analysis per year. It will help the reader to get an overview of key contributions of previous surveys done. Section 3 provides that human action recognition system is closely discussed with methodologies for detection, extractions and classification techniques evolved. Sections 4 and 5 outline recently introduced publicly available datasets used for activity recognition with challenges and applications. Finally, peculiar observations and possible directions are highlighted that need to further explore for research in the field of HAR.

2 Literature Survey

Shian-Ru Ke presented trends of HAR in video signals, and article explains the three different areas in activity recognition using core technology, human activity recognition systems and applications. This article throws light on application areas like surveillance, healthcare and entertainment industry where major focus is on surveillance in healthcare including its challenges [7]. Pau Climent-Pérez's article is based on HBA for ambient assisted living using AI. This study beautifully covers the estimation based on pose and gaze for movement identification. Later, it represents the latest work showcasing latest data tools and new datasets are described here [8]. Paul explained the techniques which are used in identification of human objects in surveillance video data with a benchmark datasets including directions for further research in living human identification and detection [9]. Fei Han represented an extensive survey of space time representations of human based on 3D skeletal data on categorization and analysis including modality, feature engineering, structure and transition including representation encoding [10]. Tej Singh explained key specifications of vision-based human activity recognition datasets which are discussed along with the algorithms according to the datasets best performance. Resolutions, actions/actors,

Table 1 Comparison of object detection methods

Techniques		Accuracy	Computational time	Advantages	Disadvantages
Background subtraction	Mixture of Gaussian model	Moderate	Moderate	Better response with Simple implementation and multi-modal scenarios	Not suitable for dynamic background and need to defined parameters
	Non-parametric background model	Moderate to high	Low to moderate	With significant post-processing, performs better in moving background	In occlusion, cannot performed
	Temporal differencing	High	Low to moderate	With sudden illumination changes, gives well performance in indoor environment	
	Warping background	High	Moderate to high	With high dynamic background, it is good in outdoor environment	Cannot work with occlusion
	Hierarchical background model	High	Low to moderate	Block-based and pixel-based approaches both are used and faster than pixel-based approach	Not good quality
Optical flow		Moderate	High	Good with dynamic camera and crowd detection	Highly computation intensive
Spatio-temporal filter		Moderate to high	Low to moderate	Perform good with low-resolution scenarios	More noise

frame rate, background and application domain are discussed in the paper [11]. Allah Bux described the image segmentation techniques and reviewed including challenges and future scope of research [12]. Athanasios Lentzas focused on the ABHAR for senior citizens. Analysis is done based on the taxonomy [13]. Michalis Vrigkas provided a comprehensive analysis of available datasets and examine the

Table 2 Comparison between different feature extraction methods

Techniques	Accuracy	Computational time	Advantages	Disadvantages
Shape-based method	Moderate	Low	With appropriate templates a simple pattern-matching approach is used	Not able to determine internal movements and in dynamic situations cannot performed
Motion-based method	Moderate	High	There is no need to predefined pattern templates	Cannot identify a non-moving human
Texture-based method	High	High	Good quality	More computation time

requirements for ideal datasets [14]. Tej Singh focused on the issue of benchmark datasets. Here article provide the comprehensive review to address this issue of benchmark datasets, action recognition-related RGB-D video datasets with 27 single-view datasets, 10 multi-view datasets, are provided [15], and various human activity recognition handcrafted and deep approaches are explained with 2D and 3D RGB and RGB_D dataset in this paper [16].

3 Methodology

See Fig. 1.

3.1 Input Image/Video

In the general process of recording the procedure of RGB cameras, a differentiation in the activity was created to analyze the sequence of actions, but at the same time, we have the challenges related to background clarity and lightning effect of the images which leads to the complexity while working on a design of the solution [17, 18]. Afterward, there was a regular improvement in the research methods to improve the factors like capturing of the depth of action in an optimal cost and real-time with the help of infrared radiation which provided a relief to the challenges related to lightning effects.

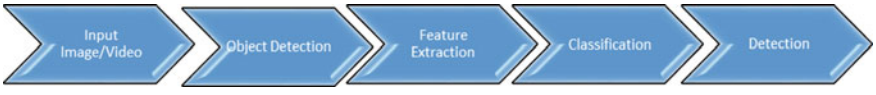


Fig. 1 Process of human activity detection

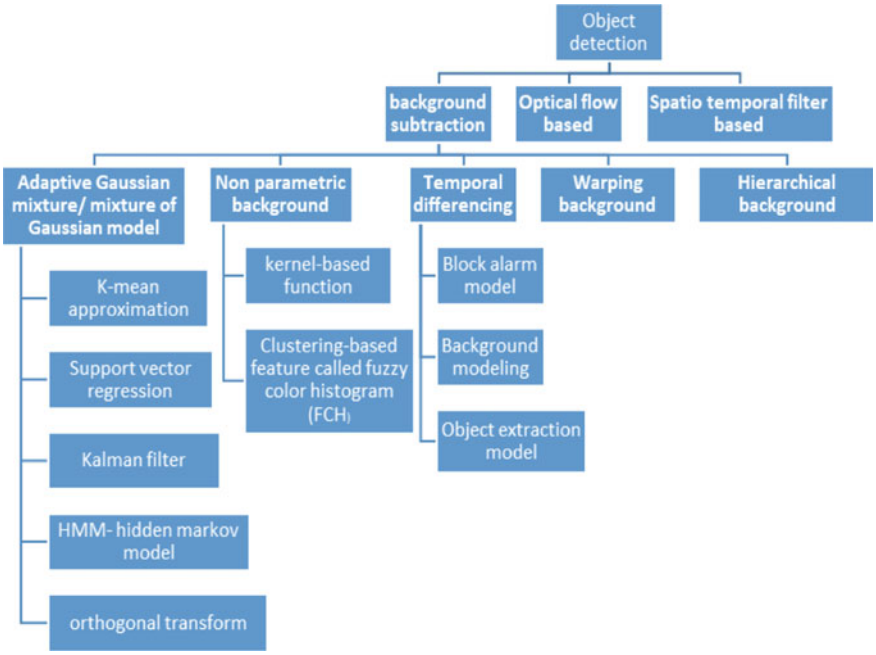


Fig. 2 Three types of human object detection

3.2 Object Detection

Background subtraction—In this techniques, a comparison of the moving object has been done based on the difference between current frames with the background frame. This comparison is done either pixel by pixel or block to block [19–21].

Optical flow based—Here in this technique, detection of moving object w.r.t. time based on the characteristic of flow vector has been used. There are challenges also related to lightening effect, motion sensitivity or noise which leads to high computational time [22].

Spatio-temporal filter based—This method is used to have reduced the computational requirement and noise by using data volume spanned by the moving object in a video signal [23]. This method is called 3D spatio-temporal because it works with spatial as well as time (Fig. 2).

3.3 Feature Extraction

This is a technique for the reduction of data dimension in which transforms lower dimension data into a modified featured space. Researcher selects a subset of features from the superset which will meet the forecasting requirements of target labels

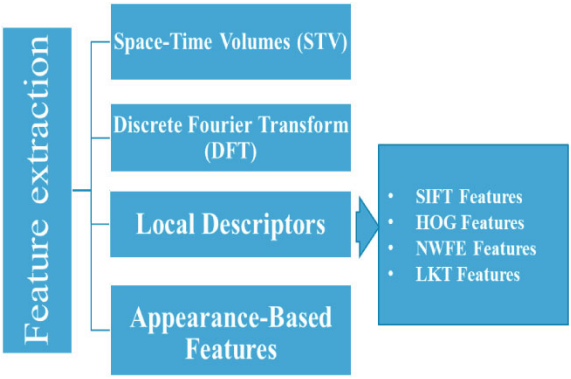


Fig. 3 Types of feature extraction

correctly diminishing the complexity of computation of different algorithm of learnings and predictions by subtracting the cost of remaining features left in the list [24]. Out of all methods, principal component analysis (PCA) gives the most reliable results in reduction of dimensions and extraction of attributes in case of linear structures [25, 26]. On the other hand, for nonlinear structure, linear discriminant analysis (LDA) technique is used to mitigate the challenges of PCA [27]. Linear discriminant analysis is used to separate the features with the aim of establishment of a linear transformation to attain the biggest class discrimination. The traditional LDA is used to find out a standard discriminant subspace (Fig. 3).

3.4 Object Classification

Shape-based method—The shape data of moving regions such as points, boxes and blobs is described foremost and after that deemed as a standard pattern recognition. While using the aforementioned approach, the large number of possible impressions of the body creates chaos between a moving human and other moving objects [28, 29]. An enormous challenge with this method is that it cannot apprehend the internal motion of the object in the contour area.

Motion-based method—In this method, we can overcome with the confusion between a moving human and other moving objects by using object motion characteristics with patterns analysis means to identify people in other moving objects, it uses the periodic property of captured images [30–32].

Texture-based method—Texture-based methods use intensity patterns for nearby pixels. This technique counts the gradient directions of local area of image and does calculations on a dense grid of evenly spaced cells. For better accuracy, it uses overlapping local contrast normalization.

4 Datasets

There are some important datasets.

KTH human motion dataset—This dataset contains six human actions performed by 25 subjects in four different situations. Running, jogging, walking, boxing, hand waving or clapping are performed in more than 2000 sequences. The backgrounds are homogeneous and uncluttered. Video files are classified by operation, to eliminate unnecessary operations easily [13].

Weizmann human action dataset—It uses static front-side cameras to record individual human movements from 10 subjects in different environments. Approximately 340 MB of video sequences are available. The actions performed include bending, walking, running, hand waving and different types of jumping.

HOHA—(Hollywood human actions)—HOHA dataset contains video sequences from 32 movies with annotations for eight action types: AnswerPhone, GetOutCar, HandShake, HugPerson, Kiss, SitDown, SitUp and Stand.

INRIA Xmas motion acquisition sequences—The video images of 390 $\times$ 291 pixel which is recorded from five different angles are included in these sequences. 11 actors perform 13 actions: check watch, cross arms, scratch head, sit down, get up, turn around, walk, wave, punch, kick, point, pick up, throw overhead and throw from bottom to top.

TUM kitchen dataset—This dataset aims ADLs in a kitchen scene with a low level of action. Multiple subjects perform tablet setting in different ways; transporting items one by one; and other behaviors are natural, grabbing multiple objects at once.

5 Challenges in HAR Dataset

In this section, we discuss the various current challenges in the dataset.

Background and environment conditions—In videos if there is moving object or background, it is very difficult to recognize human activity. There are so many types of background in a video signal like slow and fast, dynamic and static, airy and rainy, and crowded. Same recognition activity in environment conditions which contains various issues like rain, waves, trees and water is affected.

Similarity and Difference of actions—There are many actions that looks same in the videos like running, jogging, walking, etc. The same type of procedure affects classification accuracy. Similarity between classes of actions in datasets provides a fundamental challenge.

Occlusion—Occlusion occurs when an object is hiding the another object. The occlusion can be classified into two parts one is self-Occlusion and another one is partial occlusion. Occlusion is a greater challenge in computer vision applications such as human posture, object tracking and video monitoring.

View variations—In human identification system, any action recorded inside the video is the most crucial characteristic. Multiple views have larger facts

comparetively single view which leads to fair analysis of captured perspective in dataset.

6 Conclusion and Future Work

The literature survey encloses a wide area around HAR covering different methodologies and techniques of identification, detection and limitations along with its pros and cons. It also throws light on dataset benchmark and its quality which leads to the variation in results. Numerous HAR dataset challenges discussed.

In future, HAR systems need to address specific issues related to the quality of dataset and connect it to the real-life application development. In future, researcher will need to work on the challenges relating to noise, input quality data and various process-related challenges. Some meaningful datasets to represent abnormal actions in different scenarios are still a problem. Working on deep architectures from primary CNN to RCNN, RNN, auto-encoders can be extended to enhance the parameters of recognition systems.

References

1. Merrouche F, Baha, N (2016) Depth camera based fall detection using human shape and movement. In: IEEE international Conference on signal and image processing
2. Ma X, Wang H, Xue B, Zhou M, Ji B, Li Y (2014) Depth-based human fall detection via shape features and improved extreme learning machine. IEEE J Biomed Health Inf
3. Bian Z-P, Chau L-P, Magnenat-Thalmann N (2014) Fall detection based on body part tracking using a depth camera. IEEE J Biomed Health Inf
4. Lentzas A, Vrakas D (2019) Non-intrusive human activity recognition and abnormal behavior detection on elderly people: a review. Springer Nature B.V
5. Pham C, Nguyen-Thai S, Tran-Quang H, Tran S, Vu H, Tran T-H, Le T-L (2020) SensCapsNet: deep neural network for non-obtrusive sensing based human activity recognition. IEEE Access
6. Popoola OP, Wang K (2012) Video-based abnormal human behavior recognition—a review. IEEE Trans Syst Man Cybern C: Appl Rev
7. Ke S-R, Thuc HLU, Lee Y-J, Hwang J-N, Yoo J-H, Choi K-H (2013) A review on video-based human activity. Recognition 2:88–131. https://doi.org/10.3390/computers2020088
8. Chaaraoui AA, Climent-Pérez P, Flórez-Revuelta F (2012) A review on vision techniques applied to human behaviour analysis for ambient-assisted living. Elsevier
9. Paul M, Haque SME, Chakraborty S (2013) Human detection in surveillance videos and it applications—a review. EURASIP J Adv Signal Process
10. Han F, Reily B, Hoff W, Zhang H (2016) Space-time representation of people based on 3D skeletal data: a review. Elsevier
11. Dhiman C, Vishwakarma DK (2019) A review of state-of-the-art techniques for abnormal human activity recognition. In: Engineering Applications of Artificial Intelligence Elsevier, pp 21–45
12. Dhiman C, Vishwakarma DK (2020) View-invariant deep architecture for human action recognition using two-stream motion and shape temporal dynamics. IEEE Trans Image Process
13. Singh T, Vishwakarma DK (2019) Human activity recognition in video benchmarks: a survey. Springer Nature Singapore

14. Jankowski S, Szymański Z, Mazurek P, Wagner J (2015) Neural network classifier for fall detection improved by Gram-Schmidt variable selection. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
15. Brun L, Percannella G, Saggese A, Vento M IAPR Fellow (2017) Action recognition by using kernels on aclets sequences. Elsevier
16. Jing C, Wei P, Sun H, Zheng N (2019) Spatiotemporal neural networks for action recognition based on joint Loss. Springer-Verlag, London Ltd., part of Springer Nature
17. Thien Huynh- Cam-Hao Hua, Nguyen Anh Tu , Taeho Hur , Jaehun Bang , Dohyeong Kim , Muhammad Bilal Amin , Byeong Ho Kang , Hyonwoo Seung , Soo-Yong Shin , Eun-Soo Kim , Sungyoung Lee (2018) "Hierarchical topic modeling with pose-transition feature for action recognition using 3D skeleton data", Elsevier
18. Sarakon S, Tamee K (2020) An individual model for human activity recognition using transfer deep learning. In: Joint international conference on digital arts
19. Cai X, Zhou W, Wu L, Luo J, Li H (2016) Effective active skeleton representation for low latency human action recognition. IEEE Trans Multimedia 18(2)
20. Suto J, Oniga S, Lung C, Orha I (2018) Comparison of offline and real-time human activity recognition results using machine learning techniques. Springer
21. Dhiman C, Vishwakarma DK (2019) A robust framework for abnormal human action recognition using R-transform and Zernike moments in depth videos. IEEE Sens J
22. Ladjailia A, Bouchrika I, Merouani HF, Harrati N, Mahfouf Z (2019) Human activity recognition via optical flow: decomposing activities into basic actions. Springer-Verlag London Ltd., part of Springer Nature
23. Ji X, Cheng J, Feng W, Tao D (2017) Skeleton embedded motion body partition for human action recognition using depth sequences. Elsevier
24. Vishwakarma DK, Rawat P, Kapoor R (2015) Human activity recognition using Gabor wavelet transform and Ridgelet transform. In: 3rd international conference on recent trends in computing—ICRTC
25. Lahiri D, Dhiman C, Vishwakarma DK (2017) Abnormal human action recognition using average energy images. In: Conference on information and communication technology
26. Abdull Sukor AS, Zakaria A, Abdul Rahim N (2018) Activity recognition using accelerometer sensor and machine learning classifiers. In: 2018 IEEE 14th international colloquium on signal processing & its applications (CSPA 2018), Penang, Malaysia, 9–10 March [2018]
27. Tao D, Jin L, Yuan Y, Xue Y (2016) Ensemble manifold rank preserving for acceleration-based human activity recognition. IEEE Trans Neur Netw Learn Syst
28. Akagündüz E, Aslan M, Şengür A, Wang H, İnce MC (2015) Silhouette orientation volumes for efficient fall detection in depth videos. IEEE J Biomed Health Inf
29. Mazurek P, Morawski RZ (2015) Application of Naïve Bayes classifier in fall detection systems based on infrared depth sensors. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
30. Wagner J, Morawski RZ (2015) Applicability of mel-cepstrum in a fall detection system based on infrared depth sensors. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
31. Jankowski S, Szymański Z, Dziomin U, Mazurek P, Wagner J (2015) Deep learning classifier for fall detection based on IR distance sensor data. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
32. Zhang H, Parker LE (2011) 4-dimensional local spatio-temporal features for human activity recognition. In: IEEE international conference on intelligent robots and systems, San Francisco

Comparative Analysis for Email Spam Detection Using Machine Learning Algorithms

Gayatri Gattani, Shamla Mantri, and Seema Nayak

Abstract The prominence of the use of communication over the Internet is increasing progressively. Being economical, faster, and easy user interface, the number of email users is increasing tremendously. These led to the gradually increasing activity of spam. Spam emails are unrequested and unimportant emails in bulk. Due to this, there arise major Internet and email security issues that also include a problem of electronic storing space and waste of time. Thus, the identification of spam emails is very necessary. In this paper, four supervised machine learning algorithms, which are Naïve Bayes, support vector machine (SVM), logistic regression, and random forest classifier, are proposed for spam and ham emails classification. Experiments using these four algorithms are performed on prepared feature sets on two different datasets to select the best model with the highest accuracy and less overfitting or underfitting for spam detection. To automate the workflow of building the model and its evaluation, a machine learning pipeline is used in this project. Experimental results show that the overall accuracy of the random forest classifier model is the highest and also has less complexity.

Keywords Emails · Spam email · Spam email identification · Machine learning · Naïve Bayes · Support vector machine · Logistic regression · Random forest classifier

G. Gattani (✉) · S. Mantri
School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT WPU, Pune, India
e-mail: gayatrigattani2001@gmail.com

S. Mantri
e-mail: shamla.mantri@mitcoe.edu.in

S. Nayak
Department of Electronics and Communication Engineering, IIMT College of Engineering, Greater Noida, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_2

1 Introduction

In recent decades, the use of technology and the Internet has reached to peak. Being fast, cheap, and accessible, the extension of the use of email has increased tremendously. This resulted in a dramatic increase in spam emails [1]. These emails are junk emails that are almost identical and sent to multiple recipients randomly [2]. The changing way of communication by Internet on a very large scale has led to the expansion of new communication services, such as email [3]. According to a recent study, over 4 billion of the population use email. Due to its simplicity and accessibility, the mark of people using email is increasing day by day. It is extremely fast and cost-effective. With the escalation in the broadening of emails, there is also a rise in spam emails, and the unnecessary and undesirable bulk mails sent to several users haphazardly. Spam mails not only cause the problem of electronic storing space but also are the carrier of malware and hoard the network bandwidth, space, and computational power [4]. A study estimates that approximate measure of spam emails is 85%.

While the number of spam emails increases, the certainty of a user not reading a non-spam email increases. Due to the loss of network bandwidth and time consumed by users to demarcate between normal and spam [5], various spam filtering techniques have been introduced. These techniques can be categorized based on the use and non-use of machine learning algorithms. The use of ML algorithms provides an automated approach where the model trains itself based on features extracted from the dataset. As easy to implement and short training time, Naïve Bayes is a popular spam filter [6]. The main objective is to collate the accuracy of four major classification systems that include SVM, random forest classifier, Naïve Bayes, and logistic regression and select the best model for spam detection.

2 Literature Survey

Spam: Unnecessary emails sent by unknown people randomly in bulk are spam mails. These spam mails are vulnerable to major user security and also cause the problem for electronic storing space. The following are the major spam categories (Table 1).

Table 1 Frequency of major scam categories and danger level caused by them

Categories	Frequency of receiving	Danger
Ads	High	Moderate
Chain letter	Low	High
Email spoofing	Low	High
Hoaxes	Moderate	Moderate
Money scams	Moderate	High

Table 2 Some previously used techniques in spam filtering and their accuracy [9–12]

Year	Author	Classification technique	Dataset	Highest accuracy
2008	Bo Yu, Zong-ben Xu	Naive Bayes, NN, SVM, RVM	SpamAssassin & Babble Text	SVM: 95.2% and 96.0%
2011	W. A. Awad & S. M. ELseuofi	Naïve Bayes, KNN, ANN, SVM, artificial immune	SpamAssassin spam corpus	NB: 99.46%
2013	Sumant Sharma & Amit Arora	ML techniques provided by WEKA tool	SPAMBASE	94.28%
2014	Andronicus A. Akinyelu & Aderemi O. Adewumi	Classification of phishing email Using random forest ML technique	2000 phishing and ham emails set	99.7%
2017	A. S. Yuksel, S. F.Cankya, & I. S. Uncu	Cloud-based approach combining predictive analysis and ML techniques (SVM and decision tree)	SpamAssassin	SVM: 97.6%
2018	Deepika Mallampati	Naïve Bayes, J48, MLP	–	MLP: 99.3%
2021	Manoj Sethi, Sumesha Chandra, Vinayak Chaudhary & Yash	NB-multinomial, logistic regression, SVM, NN	SpamAssassin	NN: 99.02%

Spam classification: Email systems without spam classification techniques are highly open to risks. The dangers open to email systems without spam filtering are spyware, phishing, ransomware [7]. Thus, the classification of such messages can be seen as another defense mechanism against such dangers. In the previous years, various techniques of spam identification have been developed. Domain name server blacklist (DNSBL) and white list, high-volume spammers (HVSs) and low-volume spammers (LVSs) classification, machine learning-based Web spam classification, support vector machine classifier model, TruSMS systems, cloud-based approach [3], and ML algorithms like Naïve Bayes, random forest classifier, neural networks [8] are some of the classification techniques developed by researchers earlier (Table 2).

2.1 Existing Approaches

Global email users are increasing day by day. In 2024, it is set to grow up to 4.48 billion [13]. As the use of email increases, spam increases too. This causes a decrement in productivity since manually spam filtering is time-consuming, and also the electronic

storing space is reduced. Spam also increases the cyber threat to users through various phishing and malware attacks. Not only this, it has been discovered that on yearly basis, spam is accountable for over 77% of whole global email traffic [13].

In today, two common approaches, namely knowledge engineering and machine learning, are used for spam filtering. A collection of rules in knowledge engineering are used to identify mails as ham or spam. This method can lead to large time wastage and also does not guarantee the results as there is a continued need for an update in the specified set of rules. Thus, it is mainly used by naïve users [14].

Machine learning is completely based on the datasets. It just needs the training datasets, and the algorithm used itself learns the classification rules from the set of training samples from datasets. Thus, machine learning is proved to be more effective than knowledge engineering [14]. Examples of machine algorithms used for the classification of spam include Naïve Bayes, support vector machine, artificial immune systems, neural networks, logistic regression, deep learning, and many more.

The best possible outcome for any algorithm can be checked using various evaluation techniques in machine learning. This evaluation technique also helps in recognizing the overfitting and underfitting of the model. Cross-validation score, F1-score, confusion matrix, precision, recall, accuracy, regression metrics, and mean squared error can be used for evaluating the model. The three major metrics to weigh up a classification model are accuracy, precision, and recall [15].

Three major methods that are reliable for present spam detection systems are linguistic-based (used in places like a search engine), behavior-based (user-dependent since the need of change in rules from time to time), and graph-based (detect abnormal forms in data showing the behavior of spammers) [16].

3 Proposed Method

3.1 Proposed Algorithm and Workflow

Two datasets are used for this experiment to select the best algorithm with the highest accuracy. Dataset 1 is taken from Kaggle SMS Spam Collection [17]. This dataset contains 5574 messages tagged ham or spam. Dataset 2 is taken from the collection of emails from_Apache SpamAssassin's public datasets_ available on Kaggle as spam or not spam dataset [18]. There are 2500 non-spam and 500 spam emails in this dataset. The experiment is performed using four simple machine learning classification algorithms that are logistic regression, support vector machine (SVM), random forest classifier, and logistic regression on a prepared feature set of two datasets.

Through evaluation using confusion matrix, evaluation metrics, k cross-validation score, and accuracy, the perfect model with the highest accuracy and reduced underfitting or overfitting is selected. Selection of parameter k in k cross-validation score and splitting ratio of datasets play an efficient contribution in assessing the accuracy

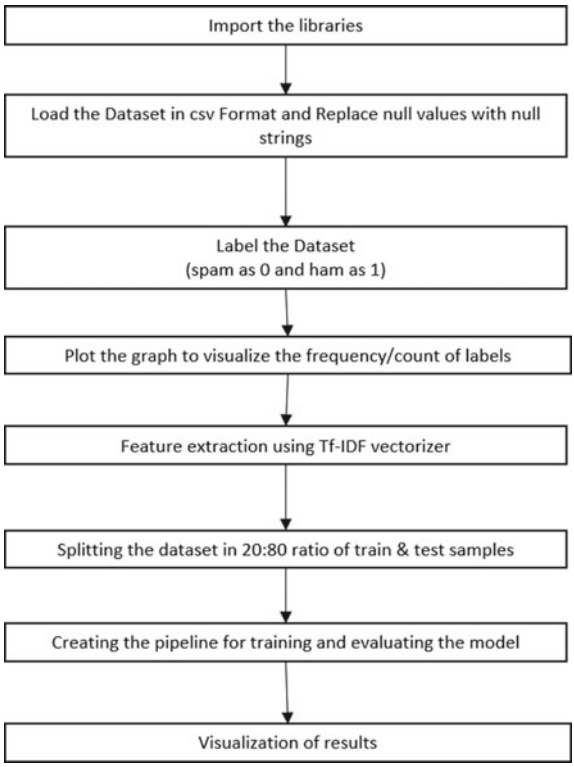


Fig. 1 Flowchart of model

of the model. The accuracy and overfitting/underfitting results are visualized using a heat map (Fig. 1).

Data Preprocessing

Both the datasets are taken from Kaggle [17, 18]. Dataset 1 comprises 5574 messages tagged according to being ham or spam. Here, we need to label the spam messages as 0 and ham messages as 1 for further simplicity. Dataset 2 comprises a collection of 3000 emails taken from "_Apache SpamAssassin's public datasets_". There are 2500 non-spam emails and 500 spam emails in this dataset. Here, the dataset initially contains the labeled data, that is, spam mails as 1 and ham emails as 0. All the null values in both datasets are converted to null strings for the normalization of plain text (Fig. 2).

Feature Extraction

The feature set will be prepared using term frequency-inverse document frequency (TF-IDF) vectorizer by transforming the feature text into feature vectors and converting it to lowercase. Parameter min_df is set to 1 that means to ignore the terms that appear in less than one document. The terms that appear irregular, min_df is used to remove them [19]. The next parameter, stopwords, is set to English to return

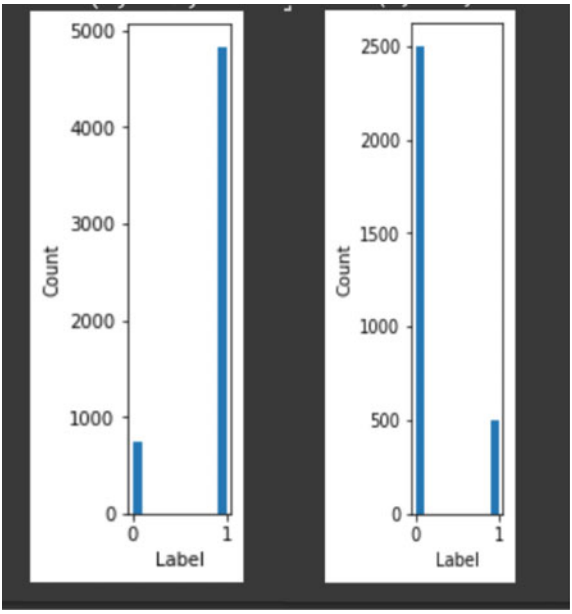


Fig. 2 Visualization of labeled dataset 1 (left) and dataset 2 (right)

the relevant stop list. The parameter lowercase is set to true to convert all characters to lowercase.

Pipeline

To automate the workflow of producing a machine learning model and evaluation of spam detection using different algorithms, a pipeline is created. The different algorithms used in this experiment are as follows:

Logistic Regression

A supervised machine learning algorithm is used for solving classification problems. It is a simple yet very effective algorithm for binary classification. The basis of this algorithm is the logistic function (sigmoid function), which takes any real-valued number and maps it in the value between 0 and 1 [20].

$$\text{Logistic Function: } y = 1/(1 + e^{-x})$$
$$\text{i.e., } 1 + e^{-x} = 1/y$$
$$e^{-x} = (1-y)/y$$
$$e^{x} = y/(1-y)$$
$$x = \log(y/(1-y))$$

Naïve Bayes

A simple probabilistic classifier uses the Bayes theorem that calculates a set of probabilities by counting the frequency and combination of values in the dataset [21].

$$P(A|B) = P(B|A)\, P(A)/P(B)$$

Using Bayesian probability terminology, the above equation can be written as [22]

$$\text{Posterior} = \text{Prior} * \text{Likelihood/Evidence}$$

Random Forest Classifier

It uses ensemble learning and regression technique to solve data classification problems [23]. It is a supervised machine learning algorithm that gets a prediction from each decision tree created.

Support Vector Machine

SVM is a supervised machine learning algorithm that classifies the data points by finding an optimal hyperplane. There are support vectors that help to maximize the classifier margin.

3.2 Performance Evaluation Criteria for Algorithm

This section is to measure and analyze the accuracy of different algorithms used in the model to estimate the results that fit best between the model and testing dataset. In this experiment, we have computed confusion matrix, evaluation metrics, and k cross-validation to assess our model and different algorithms.

Confusion Matrix

The table having four outcomes computed by the binary classifier is called confusion matrix. Measures, such as error rate, accuracy, specificity, sensitivity, and precision, are derived from the confusion matrix [24]. The four outcomes mentioned above are true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Accuracy, recall, precision, and F-score are calculated using these four outcomes. Here, in this experiment, we have considered accuracy, recall, precision, F-score, and error rate to evaluate the models.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{Total no. of dataset samples})$$

Sensitivity ***is*** also known as recall or true positive rate. It is used to measure the ability of a test to be positive when the condition is present [25].

$$\begin{aligned}\text{Sensitivity or Recall} &= \text{TP}/(\text{TP} + \text{FN}) \\ &= \text{TP}/(\text{Total positive})\end{aligned}$$

Precision is also known as positive predictive value [25]. The value ranges from 0 to 1.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

F-score is calculated with precision and recall, as follows:

$$\text{F-score} = (2 * \text{precision} * \text{recall})/(\text{precision} + \text{recall})$$

K Cross-validation Score

Cross-validation is a data resampling method to assess the generalization ability of predictive models, and cross-validation is a resampling (in such a way that no two samples overlap) method to assess the abstraction ability of models to predict the outcomes and stave off the overfitting [25]. The parameter k is the number of sets in which the sample is to be split, such that no set contains element in common. In this experiment, value of k is 4. cv_score_mean and cv_score_std are calculated to verify the accuracy results and find deviation in cv_score, respectively.

4 Results

Accuracy and precision are the important parameters in the above experiment to evaluate the different algorithms used. Other functions such as F1-score, error rate, and recall are also calculated to compare the performance of four algorithms on the above-mentioned two datasets. As shown in Fig. 3, the accuracy for random forest classifier is highest followed by the logistic regression classification algorithm. F1-score and precision of random forest classifier outperform all the other algorithms in the experiment. Figures 5 and 6 depict the evaluation results of all four algorithms on both datasets, respectively (Fig. 4).

The error rate in identifying whether the mail is spam or ham is lowest for random forest classifiers and highest for Naïve Bayes. Cv_score_std for Naïve Bayes in dataset 1 (0.0027) and logistic regression in dataset 2 (0.0066) is the lowest out of the four algorithms. Lower the cv_score_std, lower is the overfitting. However, cv_score_mean that verifies the accuracy results is highest for random forest, 0.9757 and 0.9673 in dataset 1 and 2, respectively.

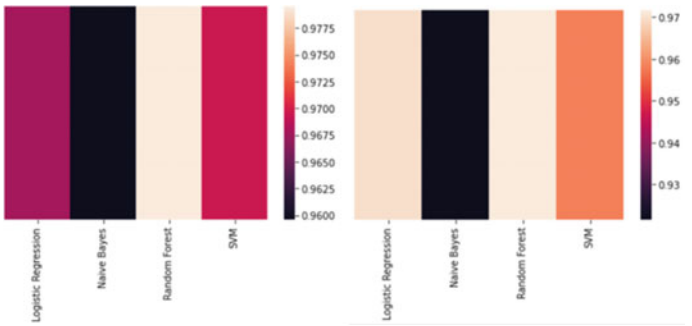


Fig. 3 Accuracy of dataset 1 (left) and dataset 2 (right)

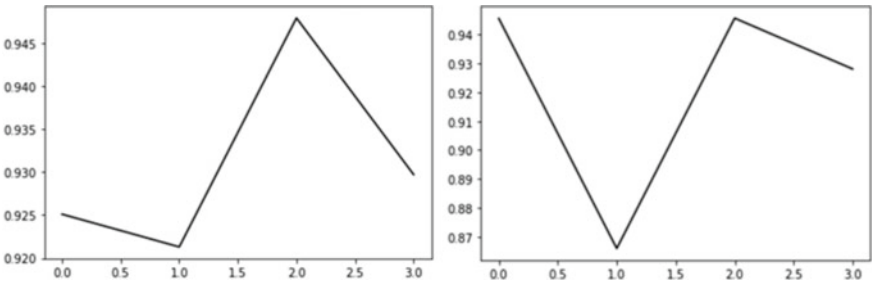


Fig. 4 F1-score of dataset 1 (left) and dataset 2 (right)

	Model	Accuracy	f1-score	precision	recall	error_rate
0	Logistic Regression	0.9677	0.9251	0.9819	0.8839	0.0322
1	Naive Bayes	0.9596	0.9213	0.8963	0.9522	0.0403
2	Random Forest	0.9767	0.9479	0.9835	0.9188	0.0233
3	SVM	0.9695	0.9297	0.9829	0.8903	0.0304

==========

	Model	cv_mean	cv_std
0	Logistic Regression	0.9669	0.0034
1	Naive Bayes	0.9549	0.0027
2	Random Forest	0.9757	0.0041
3	SVM	0.9691	0.0036

Fig. 5 Evaluation of dataset 1

Among all the models used, random forest has the highest accuracy, that is, 0.9767 and 0.9717 in both datasets, respectively.

	Model	Accuracy	f1-score	precision	recall	error_rate
0	Logistic Regression	0.9700	0.9456	0.9456	0.9456	0.0300
1	Naive Bayes	0.9216	0.8659	0.8476	0.8883	0.0783
2	Random Forest	0.9717	0.9457	0.9738	0.9222	0.0283
3	SVM	0.9583	0.9281	0.9090	0.9507	0.0416

==========

	Model	cv_mean	cv_std
0	Logistic Regression	0.9616	0.0066
1	Naive Bayes	0.8793	0.0300
2	Random Forest	0.9673	0.0075
3	SVM	0.9093	0.0304

Fig. 6 Evaluation of dataset 2

5 Conclusion and Future Work

In the comparative analysis of machine learning algorithms to classify emails as spam or ham using two different datasets, the random forest classifier is the best binary classifier out of all the four supervised algorithms. Feature extraction is done using the TF-IDF vectorizer, and the application of pipeline automates the workflow of training and evaluating the model using four different classification algorithms and different evaluation methods. Here, two different datasets are used to analyze the results on different data to select the model with high accuracy and less error rate.

The future work includes assessing the model with various effective algorithms to automate the task of filtering spam and non-spam emails using different features. This research proposes to test the model using different feature sets on different types of datasets to analyze and increase the efficiency of the prototype to identify the email as spam or non-spam.

References

1. Yu B, Xu Z-b (2008) A comparative study for content-based dynamic spam classification using four machine learning algorithms. Knowledge-based systems, vol 21. Elsevier, pp 355–362
2. Abdulhamid SM, Shuaib M, Osho O, Ismaila I, Alhassan JK (2018) Comparative analysis of classification algorithms for email spam detection. MECS, I J Comput Netw Inf Secur, pp 60–67
3. Yuksel AS, Cankya SF, Uncu IS (2016) Design of a machine learning based predictive analytics system for spam problem. In: ICCESEN, pp 500–504
4. Huang L, Jia J, Ingram E, Peng W (2018) Enhancing the Naïve Bayes spam filter through intelligent text modification detection. IEEE, pp 849–854
5. Guzella TS, Caminhas WM (2009) A review of machine learning approaches to Spam Filtering. Elsevier, pp 10206–10222

6. Rusland NF, Wahid N, Kasim S, Hafit H (2017) Analysis of Naïve Bayes algorithm for email spam filtering across multiple datasets. In: IRIS
7. Lloyed-Davis F (2017) The dangers of Spam email. Acronyms, November
8. Sethi M, Chandra S, Chaudhary V, Yash (2021) Email Spam detection using machine learning and neural networks. IRJET 8(4):349–355
9. Awad WA, ELseuofi SM (2011) Machine learning methods for spam e-mail classification. IJCSIT 3:173–184
10. Sharma S, Arora A (2018) Adaptive approach for spam detection. IISTE 7:14–21
11. Mallampati D (2018) An efficient spam filtering using supervised machine learning techniques. IJSRCSE, pp 14–21
12. Akinyelu AA, Adewumi AO (2014) Classification of phishing email using random forest machine learning technique. Hindawi, Article ID 425731
13. Spam Mail Detection using Machine Learning. PerfecteLearning, April 2021
14. Dada EG, Bassi JS, Chiroma H, Abdulhamid SM, Adetunmbi AO, Ajibuwa OE (2019) Machine learning for email spam filtering: review, approaches and open research problems. Heliyon 5(6)
15. Evaluating a Machine Learning Model. https://www.jeremyjordan.me/evaluating-a-machine-learning-model/
16. Mohammed M, Mostafa S, Ibrahim Obaid O, Zeebaree S, Abd Ghani MK, Mustapha A, Jubair M, Hassan M, Ibrahim D, Al-Dhief F (2019) An anti-spam detection model for emails of multi-natural language
17. SMS spam collection dataset. https://www.kaggle.com/uciml/sms-spam-collection-dataset
18. Apache, Spam or not spam dataset. A collection of emails from spamassassin.apache.org
19. TF-IDF vectorizer. Scikit Learn Library (2021)
20. Yeldirim S (2020) Logistic regression as a classification algorithm. Towards Data Science
21. Yang S (2019) Introduction to Naïve Bayes classifier. Towards Data Science
22. Venkatraman A (2019) Naïve Bayes for machine learning—from zero to hero. FloydHub
23. Dada E, Joseph S (2018) Random forests machine learning technique for email spam filtering, vol 9. University of Maiduguri, pp 29–36
24. Classeval (2017) Basic evaluation measures from the confusion matrix. Classevalblogs
25. Berrar D (2018) Cross-validation, vol 1. Elsevier, pp 542–545

Post-quantum Cryptography: A Solution to the Challenges of Classical Encryption Algorithms

Shagun Sharma, K. R. Ramkumar, Amanpreet Kaur, Taniya Hasija, Sudesh Mittal, and Bhupendra Singh

Abstract With the rise in quantum computing, there is an increase in the security concerns of existing cryptographic algorithms such as Rivest Shamir Adleman (RSA) and advanced encryption standard (AES) along with elliptic curve cryptography (ECC). The problem with RSA and ECC is that they depend on the challenges of integer factorization and discrete logarithms, respectively, whereas AES can be the target of brute force attack. In the recent years, there is a huge advancement in quantum computing due to which the existing encryption standards found vulnerable as they are weak to handle quantum attacks. Hence, to solve the problems with quantum computing, a new set of algorithms came into existence called as post-quantum cryptography (PQC) which is safe and secure from quantum attacks. In this review, firstly, the quantum computers and number of classical algorithms are introduced, afterward the classification of PQC algorithms has been shown. The aim of this review is to present the technical overview about PQC algorithms and their working procedures to protect classical cryptographic-based encrypted information from quantum attacks during online transactions and communications.

S. Sharma (✉) · K. R. Ramkumar · A. Kaur · T. Hasija · S. Mittal
Chitkara University Institute of Engineering & Technology, Chitkara University, Rajpura, Punjab, India
e-mail: shagun.sharma@chitkara.edu.in

K. R. Ramkumar
e-mail: k.ramkumar@chitkara.edu.in

A. Kaur
e-mail: amanpreet.kaur@chitkara.edu.in

T. Hasija
e-mail: taniya@chitkara.edu.in

S. Mittal
e-mail: sudesh.mittal@chitkara.edu.in

B. Singh
Centre for Artificial Intelligence & Robotics, DRDO, Bengaluru, India
e-mail: bhupendra@cair.drdo.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_3

Keywords Cryptography · Post-quantum cryptography · Quantum computer · Signature · Lattice · McEliece · Rainbow

1 Introduction

In the present time, cryptography plays a major role for ensuring the privacy and security of online communications and transactions [1]. It is the art of encrypting and decrypting the data. Cryptographic-based algorithms are used by industries along with government firms to establish an authentic and confidential communication having the presence of integrity in it. The public key cryptography is a major concept of classical encryption algorithms, which allows participants to securely communicate over an insecure network [1, 2]. It solves the key distribution challenges by reliably verifying the participant identities using digital signatures which is widely used for protecting online applications, software, e-identity documents, e-passports, and online contracts [3]. The public key cryptography uses the concept of two different keys: Public key is used to perform encryption which is known to everyone, and second is the private key used for decryption which is known to the single party called receiver [1] as shown in Fig. 1. Most of the public key algorithms such as digital signatures and Rivest Shamir Adleman (RSA) are based on the eminent mathematical challenges [4]. The RSA algorithm is based on prime factorization of numbers, whereas unidirectional Euler's function is the center point of digital signatures. Additionally, elliptic curve cryptography (ECC) is a public key algorithm which is used in discrete logarithms to compute the point groups on an elliptic curve.

However, with the increase in the advance communications and computing systems, it is very easy to break asymmetric algorithms within a fraction of seconds [5] such as RSA is having a key size between 2048 and 4096 bits, however, it is well understood that the RSA implemented using 768 and 1024 bits key is already broken in the year of 2010. With the advancements of quantum computers-based upon quantum physics, it became very easy to solve the complex mathematical problems within a limited period of time [1, 4] as it could solve integer factorization and the complicated challenges related to discrete mathematics. Along with these advantages, it became a difficulty for public key cryptographic algorithms to remain working due to a factor that they could be broken easily. As an example, in the year of 1994, an algorithm called Shor quantum came into existence which exploited public key cryptographic algorithms and another algorithm called Grover quantum (1996) threatened symmetric key cryptographic algorithms.

Fig. 1 Public key cryptography

In basic, discrete algorithm and elliptic curve discrete algorithm along with integer factorization are mathematics based challenges which are very easy to break using Shor's quantum algorithm [6]. Hence, it has been found a big concern because these mathematical problems are used as the common equations for protecting confidential information over an insecure network called Internet. Later on for solving the impact of quantum computers on cryptographic algorithms, National Institute of Standards and Technology (NIST) started researching [4] which resulted a domain called PQC.

2 Background

In this chapter, a brief introduction about classical cryptography, quantum computers, and PQC has been provided. Along with which various classes of PQC are given to protect classical cryptographic algorithms from quantum attacks.

2.1 Classical Cryptography

Cryptography is invented by a famous cryptographer named as J. Coron as a technique which could help users to communicate and connect with other users in a secure way by guarantying privacy, security along with authenticity [2]. It protects private information from attackers and solves the challenges of information leakage [7]. It is a mathematical function which is used to transform the information by using a variable called key [8]. There are two types of cryptographic algorithms; symmetric key algorithm having same key for encryption and decryption, and another is asymmetric key having two different keys called public and private keys for encryption and decryption, respectively.

2.2 Quantum Computers

Quantum computers are the computers which are based on the working of quantum physics [1, 4]. The performance of such computers is very high as compared to classical computers [9]. Quantum computers can implement Shor's quantum algorithm to solve integer factorization and discrete logarithms which are used by ECC and RSA in the polynomial time [5, 10, 11]. These computers are found very powerful when dealing with complex problems of mathematics, which are very difficult to solve using classical computers as a result, they are found to be vulnerable toward the present cryptosystems. The quantum computers have been already placed in the numerous areas of the world. Hence, it is required to have some quantum resistant algorithms to protect the information from quantum attacks.

2.3 *Post-quantum Cryptography (PQC)*

PQC is a cryptographic system which came into existence with the evolution in quantum computers [4]. It is an algorithm which resists quantum attacks on cryptosystems by using matrix mathematics and ciphers [4, 7]. Matrix mathematics is found to be safe from quantum attacks as compared to classical methods such as integer factorization and discrete logarithms [7]. The classical computers cannot break classical encryptions algorithms, but quantum computers are very strong at breaking the base of cryptographic algorithms [12]. Hence, it is required to replace classical encryption algorithms with PQC algorithms. The PQC is classified into four types as shown in Fig. 2.

Lattice-based cryptographic algorithm is a public key encryption and performs best among all the algorithms due to high flexibility of lattices which can be built on the basis of shortest vector problem (SVP) [13–15]. It requires Euclidean length of a lattice vector present in the structure. In such algorithms, lattices are the objects which play a major role in cryptography. Lattices are first analyzed by famous mathematicians named as Carl Friedrich Gauss and Joseph Louis Lagrange [5]. These algorithms are highly resistant to quantum attacks and sub-exponentials plus are most studied which can be used for the implementation of digital signature, homomorphic encryption along with key exchange mechanisms [13]. Due to their quantum resistant behavior, they are having high efficiency and security than classical encryption algorithms [16].

Code-based cryptographic algorithm is a public key encryption algorithm [5] developed by Robert McEllice. This algorithm is replaced various traditional cryptographic algorithms due to its advantages such as high speed and security. It uses codes which are having error corrections by offering a traditional algorithm for asymmetric key encryption [15]. The legitimate receivers receive the information along with error correcting code and make the corrections using private key wherever required without letting other parties know about the secret of deciphering [4].

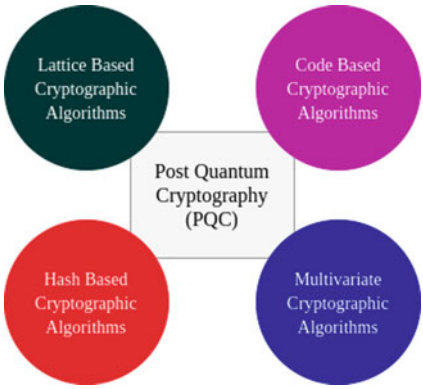


Fig. 2 Classification of post-quantum cryptography

Table 1 Classification of post-quantum cryptography family

Lattice-based cryptosystem	Multivariate cryptosystem	Hash-based cryptosystem	Code-based cryptosystem
BLISS	Rainbow	Lamport Signature	McEliece
CRYSTALS-KYBER	MQDSS	Merkle Signature	BIKE
Frodo KEM	LUOV	SPHINCS	LEDAcrypt
NTRU	GeMSS	Picnic	NTS-KEM
ROUND5			ROLLO
SABER			Rank QuasiCyclic
Three bears			
Ring TWE			
Ring LWE			

The security of code-based cryptographic algorithm relies on coding theory, i.e., the hardness of problems [17].

Hash-based cryptographic algorithm is first invented by Ralph Merkle in the year of 1979 [5]. Such algorithms are dependent on hash functions along with security as the major concern [5, 15]. It focuses on designing of digital signatures and mainly work to secure hash functions [17]. This algorithm was designed to convert a single time signatures into multiple signatures [4]. It is having a tree like structure where root of the tree denotes public key, leaf node denotes authentication key, and layer denotes the data block having a connection to every corresponding hash node.

Multivariate cryptographic algorithm is a public key cryptography consisting of numerous polynomial equations having number of variables in a finite field [4]. These algorithms are used for solving digital signature problems along with asymmetric key challenges [12, 17]. The set of polynomials in such algorithms are called as multivariate polynomials. By using the concept of polynomial, it is very easy and fast to verify the digital signatures having short signature length which is useful for authentication of legitimate party [5]. Algorithms under each PQC family are shown in Table 1.

3 Literature Review

Numerous classical cryptographic algorithms have been defined along with the introduction of quantum computers [3]. With the increase in security concerns of classical encryption techniques, authors have shown the need of PQC algorithms along with the challenges associated with them. Various types of PQC and classical cryptographic algorithms have been analyzed and found that with the emergence in quantum computers, it is most required to have quantum resistant algorithms [4]. It has been mentioned that by using PQC algorithms, the classical encryption algorithms can be replaced with the high security and privacy. Numerous PQC algorithms have been studied and listed them under their types [13]. The results of survey have mentioned number of properties of PQC algorithms and the areas in which they could be used.

A PQC algorithm-based deep survey has been done along with the development of quantum resistant algorithms [18]. The uses of PQC in real-time encryption and Web applications have been mentioned along with their impacts with respect to time was compared with classical cryptographic algorithms. Number of classical cryptographic algorithms have been introduced along with their uses in the real-world applications [2]. The research on quantum computers has been done, and their impacts on classical cryptography were noted in the form of security challenges. By looking at the classical cryptographic security issues, a PQC-based solution has been introduced. A discussion on classical cryptographic algorithms has been done, and how quantum computers can affect them was also mentioned [1]. Various security concerns of classical cryptography were focused to make a secure solution using PQC algorithms. Various drawbacks of quantum computing on classical encryption algorithms along with the implementation of PQC in the real-world applications have been shown [14]. Various mathematical facts and hardware architectures were introduced to utilize in the classical encryption as well as PQC algorithms. A quantum resistant signature system has been developed by keeping confidentiality and integrity of cloud manufacturing in mind [19]. The developed system was based on public key cryptography having a property of authentication security. NIST finalist (PQC-based algorithms) has been compared to see their performance on Raspberry Pi 4 Model B and Linux [7]. The comparison has resulted that most of the algorithms work better on Raspberry Pi 4B by consuming less time as compared on Linux. The CPU speed was kept as the major factor which has shown approximate 1197 MHz and 1.5 GHz speed for Linux and Raspberry Pi 4B, respectively. A brief survey has been done to get the information of using PQC algorithms on isogeny of elliptic curve cryptography (ECC) for their practical implementation on mobile applications [6]. The comparison of PQC and classical encryption algorithms has been made to see the stability, advantages, disadvantages along with their speed on ECC-based systems. Authors have introduced the challenges of adopting PQC algorithms for encrypting and decrypting the data [8]. The necessary sources along with planning requirements to migrate from classical encryption standards to PQC algorithms have been mentioned. The PQC security has been analyzed in messaging specification signal [9]. A deep survey on signal specification has been made to get the knowledge, whether it could be used as a secure PQC algorithm or not. As a result, signal specification has been found as an insecure PQC algorithm, along with it the reasons of not calling it as a quantum secure algorithm were also mentioned. Numerous encryption-based techniques such as quantum networks, quantum key distribution, and quantum random number generator have been introduced [10]. Furthermore, the security concerns with classical encryption algorithms were also mentioned. Various quantum resistant algorithms and signature schemes for asymmetric key encryption have been presented [20]. The comparison among numerous quantum resistant algorithms was done which resulted lattice-based and code-based PQC systems as the best algorithms to use for secure and confidential communications. Use of PQC algorithms to secure Internet of things (IoT) networks has been introduced [11]. Furthermore, the challenges associated with the implementation of PQC algorithms on IoT networks were also mentioned. Various PQC-based algorithms have been introduced

along with the comparison among NIST PQC algorithms [15]. Authors have implemented a design called high-level synthesis (HLS) which was used to map C language specifications in ASIC and FPGA implementations. Numerous PQC algorithms have been introduced for the use in IoT systems [21]. The study has shown that PQC algorithms can protect IoT systems from quantum attacks. In this analysis, the challenges and future trends associated with PQC algorithms were also mentioned. Authors have introduced various PQC algorithms along with a deep survey on lattice-based quantum cryptographic algorithms [22]. The rudimentary knowledge about implementing lattice-based cryptosystems using hardware and software has been shown. Security comparison among PQC algorithms [23] has been performed by looking at the challenges such as integer factorization and discrete logarithms. The analysis was implemented to look at the complexity issues in various PQC algorithms. The results of the study have shown isogeny algorithms as the best technique having the best results for security concerns. Various quantum attacks like integer factorization and discrete logarithms have been introduced that could be implemented on classical encryption algorithms such as RSA and ECC [5]. The algorithms associated with each type of PQC have been mentioned and introduced. Privacy enhancement technologies have been introduced which are suitable for privacy concerns in IoT and Internet of vehicles (IoV) networks [24]. The challenges with privacy enhancement technologies have been shown, and future direction of using PQC in implementing secure IoT and IoV networks was also mentioned. Various strategies have been recommended to implement the solutions for cyber and quantum attacks [25]. These strategies include lattice-based, code-based, multivariate, and hash-based PQC algorithms. It has been analyzed that how quantum computers can very badly effect on classical encryption algorithms [26]. To solve this issue, a survey has been done and identified that lattice-based cryptographic algorithms provide and ensure that they belong to the best and most quantum secured class because of their quantum resistant nature. Basic security concerns in IoT systems have been identified [27]. The classical algorithms used to implement in the IoT systems to ensure security were mentioned along with the challenges of using them. The need of PQC algorithms in IoT devices was mentioned for guaranteeing high security and safety from quantum attacks. A comparison among classical encryption algorithm by using SWOT framework has been analyzed and identified the need of using PQC algorithms to ensure network security for avoiding quantum attacks on security systems [28]. With the increase in quantum computers, it became very important to find the solution to secure online information against quantum attacks [29]. With this idea, authors have identified and reviewed various PQC algorithms to secure cloud data from classical as well quantum attacks to provide information security.

4 Related Work

Various PQC algorithms are defined below along with their key generation process.

4.1 Rainbow

It is a multivariate PQC algorithm [13] proposed by D. Schmidt and J. Ding which came into existence in the year of 2005. It is one of the finalist of NIST project [2] which has very fast signing and verification process due to short signature size. However, it requires a very large public key which made it unsuitable to use as normal digital signature. It is having 207 bit of classical and 169 bit of quantum security [1]. In the classic rainbow algorithm, the sizes of public and private key are 710,640 bytes and 511,448 bytes, respectively.

The steps to retain the authenticity using rainbow algorithms are mentioned below-Let

i. Finite field: k.
ii. First invertible affine linear map: L_m in k.
iii. Second invertible affine linear map: L_p in k.
iv. Central map: F in k.
v. Message size: x.
vi. Signature size: y.
vii. Message: z vector in k in the form of $z(z_0, z_1,\ldots, z_{x-1})$.
viii. Signature: w vector in k in the form of $w(w_0, w_1,\ldots, w_{y-1})$.
ix. Private key: L_m, F, L_p.
x. Public key: L_m o F o L_p.

Signing process

To sign a rainbow message—$z(z_0, z_1,\ldots,z_{x-1}) \in k$. it is required to solve the below given equation.

$$FoL_p\left(w_0, w_1, \ldots, w_{y-1}\right) = L_m^{-1}(z_0, z_1, \ldots, z_{x-1})$$

Now, generation of signature is required, hence use hash function (SHA-256) to compute hash value of message.

$$z' = \text{hash}(z)$$

Compute invertible affine linear map L_m

$$\overline{z} = L_m^{-1}(z_0^{'}, z_1^{'}, \ldots, z_{x-1}^{'})$$

Compute inverse of central linear map on $\overline{z}$

$$\overline{w} = F^{-1}(\overline{z_0}, \overline{z_1}, \ldots, \overline{z_{x-1}})$$

Now, compute invertible affine linear map L_p on $\overline{w}$

$$w = L_p^{-1}\left(\overline{w_0}, \overline{w_1}, \ldots, \overline{w_{y-1}}\right)$$

Finally, the value of w is signature.

Signature verification process

To change the default, adjust the template as follows.

Compute

$$z''\left(z_0'', z_1'', ..., z_{x-1}''\right) = \overline{F}\left(w_0, w_1, \ldots, w_{y-1}\right)$$

If

$$z'' = z'$$

then, the signature is accepted else rejected.

4.2 McEliece

It is a code-based cryptosystem developed by RobertMcEliece in the year of 1978 [2, 5, 12]. It encrypts the data by using matrix [2]. These matrices are having some errors added in them which makes it very difficult for third party to decrypt the data. This algorithm is having an issue of heavy key length which makes it very difficult to use in embedded devices [14, 20]. The process of decryption and encryption is very fast as compared to RSA algorithms [20] (Table 2).

Key generation

For generating public and private keys, Bob chooses three matrices as follows

Matrix 1: $H \in F_2^{n*k}$.

It is a generator matrix for a linear code which can be decodable up to "e" errors.

Matrix 2: $I \in F_2^{k*k}$ is a random invertible matrix.

Matrix 3: $J \in F_2^{n*n}$ is a random permutation matrix.

Private key: H, I and J matrix.

Public key: $\left(\hat{H} = H * I * J\right)$.

Encryption

Alice has, message: "m".

Public key: $\hat{H}$.

Table 2 Comparison among post-quantum cryptography-based algorithms

Algorithms	Signature (ms)	Signature verification (ms)	Public key size	Private key size	Signature size	Encryption time	Decryption time	Year [23]	Advantages [14]	Disadvantages [14]
Rainbow (multivariate PQC)	257.1 ms [5]	288.0 ms [5]	7,10,640 bytes[1]	5,11,448 bytes [1]	512 bits [2]	NA	NA	2005	Fast	Less secure
McEliece (code-based PQC)	NA	NA	500 kb [5]	1000 kb [5]	NA	NA	NA	1978	Mature and secure	Excess memory necessities
Ring LWE (lattice-based PQC)	NA	NA	NA	NA	NA	68 ms [5]	18.8 ms [5]	NA	Robust, flexible with excellent security	Not fully tested
Merkle signature (hash-based PQC)	NA	NA	0.08 kb [5]	~ 250 kb [5]	NA	NA	NA	1990	Extremely fast	Deals with signature only

e: She selects a zero vector of length n and add t random 1's in it.

$$Y = m\hat{H} + e$$

Decryption

Bob has Y and private key to decrypt the message.

$$Y = m\hat{H} + e$$

where

$$\hat{H} = H * I * J$$

$$Y = m(H * I * J) + e$$

Apply J^{-1} on both R.H.S and L.H.S

$$Y^{'} = m(H * I) + e'$$

Bob applies a decoding method to remove e' and gets:

$$Y' = m(H * I)$$

Apply H^{-1} on both R.H.S and L.H.S

$$Y'' = m * I$$

Apply I^{-1} on both R.H.S and L.H.S

$$Y''' = m$$

4.3 Ring Learning with Errors (LWEs)

It comes under lattice-based cryptosystem [18] which is used for building public key cryptography [4]. This algorithm focuses on polynomial ring concept over finite field [5]. It chooses the coefficients from finite field and constructs the arithmetic of polynomials.

Key generation

Message bit: M_1

Prime number: Q_1
Select a random polynomial (P_1)
Select a random secret polynomial (S_1)
Select a random error polynomial (E_1)

Create a new polynomial:

$$P_1 * S_1 + E_1 = B_1$$

Public key: P_1 and B_1.
Private key: S_1

Encryption

$$u_i = \Sigma(P_1 \text{ Samples})(\text{mod}\, Q_1)$$

$$v_i = \Sigma(B_1 \text{ Samples})(Q_1/2 * M_1)$$

Now, encrypted values are (u_i, v_i).

Decryption

$$D_i = v_i - S_1 * u_i (\text{mod}\, Q_1)$$

If,

$$D_i < \frac{Q_i}{2}, \text{ the message } M_1 \text{ is } 0$$

Else

$$D_i > \frac{Q_i}{2}, \text{ the message } M_1 \text{ is } 1$$

4.4 Merkle Signature

It is a hash-based cryptosystem invented by Ralph Merkle in the year of 1990 [5]. It changes the signatures from one time to multiple [4]. It is having a tree like structure, in which root shows public key and leaf shows one-time authentication key. This type of algorithms is based on the security of hash functions [5]. Furthermore, it has a disadvantage, i.e., it produces number of signature at once.

Key generation

For generating a public key, create "n" private and public key pairs (x_i, y_i) of one-time signing scheme, i.e., lamport scheme.

To show that (x_i, y_i) are the actual key pairs to verifier, signer needs to include corresponding nodes of the tree to authenticate that

$$h = a_{0,i}$$

where $a_{i,j}$ is the node of the tree having height i and j as left–right position.

Now, detect the route having $n + 1$ nodes from $a_{o,i}$ to root node called A.

The route is $A_0, A_1, \ldots, A_n$ with

$$A_0 = a_{0,i}(\text{Leaf node}) = H(y_i) \text{ and}$$

$$A_n = a_{n,i}(\text{root node})$$

For computing the route, value of each child and neighbor node should be known to the verifier. Hence, identify child of A_{i+1} and sibling node of A_i called L_i.

$$A_{i+1} = H(A_i||L_i)$$

After computing all the nodes from L_0 to L_{n-1}.

$$\text{Signature} = \left(S'||y_i||L_1||L_2||L_3||\ldots||L_{n-1}\right)$$

For all $1 \leq \leq i \leq 2^n$, compute a hash function of y_i to generate a hash tree.

$$h = H(y_i)$$

Public key: y_i.

Private key: (x_i, y_i).

Signature generation

Message: M

To sign M, signer uses the pair (x_i, y_i) which is not previously used for signing any other message resulting in signed message along S' with respective public key (y_i).

Signature verification

Receiver has public key, message, and signature.

At first, one-time signature of M is verified.

If signature is valid, then calculates:

$$A_0 = h(y_i)$$

Now for j from 1 to $n-1$, the nodes of A_j route are computed as

$$A_j = h(A_{j-1}||L_{j-1})$$

If A_n is equal to the public key, the signature is valid else not.

5 Conclusion

The modern advancements in the quantum computing algorithms have made classical cryptographic algorithms restricted to quantum attacks. However, there are only limited number of quantum computers available, but with the instant increase in technology, the flourishment of quantum computers will also increase which is a biggest challenge for classical cryptographic algorithms to manage because such attacks can compromise the main properties of cryptography such as integrity, confidentiality, authenticity, and authorization. Till date, the classical encryption algorithm used in online transactions, i.e., RSA is safe and only broken theoretically by using Shor's algorithm. In future, if it will be practically implemented on quantum computers, the complete set of financial data will be hacked by breaking the base of cryptographic algorithms. Hence, to deal with aforementioned challenge, an emerging field called PQC came into existence which is quantum resistant. From the literature survey, it has been analyzed that with an emergence in quantum computers, the need of PQC algorithms has been also increased. The classical encryption algorithms are safe, but in future, they need to be replaced with PQC algorithms due to their quantum insecure behavior. In this survey, the classification of PQC has been done along with the various algorithms under them. Further, the advantage, disadvantages, key length, and signature size are also mentioned. Lastly, this paper can help researchers and academia to get the knowledge about PQC-based emerging algorithms.

For the future direction, it is required to develop and use the PQC algorithms to avoid classical as well as quantum attacks in the process of the online transactions and communications to secure data and provide high integrity, confidentiality, authenticity, and authorization.

References

1. Yunakovsky SE, Kot M, Pozhar N, Nabokov D, Kudinov M, Guglya A, Kiktenko EO, Kolycheva E, Borisov A, Fedorov AK (2021) Towards security recommendations for public-key infrastructures for production environments in the post-quantum era. EPJ Quantum Technol, 1–19
2. Gill B, Shailer S (2021) Analysis of security solutions for a post-quantum world
3. Campagna M, LaMacchia B, Ott D (2021) Post quantum cryptography: readiness challenges and the approaching storm

4. Wang S (2021) Research on post-quantum cryptosystem based on IoT devices. J Phys: Conf Ser 1883:1–5.
5. Roy KS, Kalita HK (2019) A survey on post-quantum cryptography for constrained devices. Int J Appl Eng Res, 2608–2615
6. Bobrysheva J, Zapechnikov S (2020) The relevance of using post-quantum cryptography on the isogenies of elliptic curves for mobile application protection. In: Advanced technologies in robotics and intelligent systems. Springer, Cham, pp 99–103
7. Önder, Elçin (2021) Measuring the performance of post-quantum cryptography on embedded systems. Diss. Worcester Polytechnic Institute, pp 1–83
8. Barker W, Polk W, Souppaya M (2020) Getting ready for post-quantum cryptography: explore challenges associated with adoption and use of post-quantum cryptographic algorithms. The publications of NIST Cyber Security White Paper, CSRC. NIST. GOV 26
9. Bobrysheva J, Zapechnikov S (2020) Post-quantum security of messaging protocols: analysis of double ratcheting algorithm. In: IEEE conference of Russian young researchers in electrical and electronic engineering (EIConRus), pp 777–780
10. Arnold DW, Perez-Delgado C (2020) Quantum cryptography: security for the post-quantum world
11. Lohachab A, Lohachab A, Jangra A (2020) A comprehensive survey of prominent cryptographic aspects for securing communication in post-quantum IoT networks. Internet of Things, 1–23
12. Wang, Wen (2021) Hardware architectures for post-quantum cryptography. Diss. Yale University, pp 1–24
13. Onuora AC, Madubuike CE, Otiko AO, Nworie JN (2020) Post-quantum cryptographic algorithm: a systematic review of round-2 candidates
14. Asif, Rameez (2021) Post-quantum cryptosystems for internet-of-things: a survey on lattice-based algorithms. IoT, 71–91
15. Basu K, Soni D, Nabeel M, Karri R (2019) NIST post-quantum cryptography—a hardware evaluation study. IACR Cryptol, pp 1–16
16. Pujeri U, Aithal PS, Pujeri R (2021) Survey of lattice to design post quantum cryptographic algorithm using lattice. Int J Eng Trends Technol, 92–96
17. Balamurugan C, Singh K, Ganesan G, Rajarajan M (2021) Code-based post-quantum cryptography, pp 1–24
18. Tutoveanu, Anton (2021) Active implementation of end-to-end post-quantum encryption. IACR Cryptol, pp 1–12
19. Yi, Haibo (2021) A post-quantum secure communication system for cloud manufacturing safety. J Intell Manuf, 679–688
20. Tera, Helen (2019) Introduction to post-quantum cryptography in scope of NIST's post-quantum competition
21. Fernández-Caramés, Tiago M (2019) From pre-quantum to post-quantum IoT security: a survey on quantum-resistant cryptosystems for the Internet of Things. IEEE Internet Things J, 6457–6480
22. Nejatollahi H, Dutt N, Ray S, Regazzoni F, Banerjee I, Cammarota R (2019) Post-quantum lattice-based cryptography implementations: a survey. ACM Comput Surv (CSUR), 1–41
23. Borges, F, Reis PR, Pereira D (2020) A comparison of security and its performance for key agreements in post-quantum cryptography. IEEE Access 8:142413–142422
24. Malina L, Dzurenda P, Ricci S, Hajny J, Srivastava H, Matulevičius R, Affia A-AO, Laurent M, Sultan NH, Tang Q (2021) Post-quantum era privacy protection for intelligent infrastructures. IEEE Access 9:36038–36077
25. Badhwar, Raj (2021) The need for post-quantum cryptography. The CISO's next frontier. Springer, Cham, pp 15–30
26. Dadheech, Aarti (2021) Post-quantum lattice-based cryptography: a quantum-resistant cryptosystem. Limitations and future applications of quantum cryptography. IGI Global, pp 102–123

27. Seyhan K, Nguyen, Nguyen N, Akleylek S, Cengiz K (2021) Lattice-based cryptosystems for the security of resource-constrained IoT devices in post-quantum world: a survey. Cluster Comput, 1–20
28. Vaishnavi A, Pillai S (2021) Cybersecurity in the quantum era—a study of perceived risks in conventional cryptography and discussion on post quantum methods. J Phys Conf Ser, 1–11
29. Ilias SM, CeronmaniSharmila V (2021) Recent developments and methods of cloud data security in post-quantum perspective. In: 2021 international conference on artificial intelligence and smart systems (ICAIS). IEEE, pp 1293–1300

A Survey on Performance Analysis of Different Architectures of AES Algorithm on FPGA

Taniya Hasija, Amanpreet Kaur, K. R. Ramkumar, Shagun Sharma, Sudesh Mittal, and Bhupendra Singh

Abstract Encryption is the primary way for ensuring communication security. The symmetric key method, often known as Advanced Encryption Standard (AES), is a well-known technique in the field of security. AES can be implemented in either software or hardware. In the current study, Field Programmable Gate Arrays (FPGAs) are utilized to implement AES. Number of studies have been done on experiments of AES using FPGAs. Till date, no study has been done on the architectures that are being utilized to implement AES on FPGA. This paper provides an in-depth examination of several hardware implementation of AES on FPGA in terms of through put and performance. This survey article enables the engineers to choose the best architecture of FPGAs to implement AES algorithm in terms of design as per the requirement. The surveyed architectures are sequential, pipelined, iterative, and parallel. Parallel architectures with pipelining between rounds have shown an excellent throughput. Certain improved S-box and key expansion approaches have also been employed by the researchers to reduce the hardware areas.

Keywords AES algorithm · AES architectures · Pipeline architecture · Parallel architecture

T. Hasija (✉) · A. Kaur · K. R. Ramkumar · S. Sharma · S. Mittal
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
e-mail: taniya@chitkara.edu.in

A. Kaur
e-mail: amanpreet.kaur@chitkara.edu.in

K. R. Ramkumar
e-mail: k.ramkumar@chitkara.edu.in

S. Sharma
e-mail: shagun.sharma@chitkara.edu.in

S. Mittal
e-mail: sudesh.mittal@chitkara.edu.in

B. Singh
Centre for Artificial Intelligence & Robotics, DRDO, Bangalore, India
e-mail: bhupendra@cair.drdo.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_4

1 Introduction

Internet messaging and data transfer are the most preferred modes of communication in today's society. Security is the primary concern of every business or government organization. If a company provides data security to its clients, it will be regarded trustworthy; otherwise, the organization will be considered unreliable. Cryptography is a type of cybersecurity that encrypts and decrypts data in order to ensure its secrecy, integrity, and validity [1]. Encryption is the act of converting data into cipher text (incomprehensible text), and decryption is the process of converting that cipher text back to intelligible language [2]. Cryptography employs two different types of security algorithms: symmetric key cryptography and asymmetric key cryptography. The only distinction between the two is how the key is utilized. In symmetric key cryptography, the same key is used for encryption and description, but in asymmetric key cryptography, two separate keys are used for encryption and description [3]. Some symmetric algorithms are DES, 3DES, AES, and Blowfish [4, 5]. Asymmetric key cryptography, on the other hand, includes Diffie Hallman, RSA, ECC, and digital signatures [5].

The National Institute of Standards and Technology (NIST) released a new algorithm to replace the data encryption standards (DES) algorithm, and the name given to the new algorithm is AES. Algorithm implementation can be divided into two categories: software implementation and hardware implementation. The drawback of DES is that it is only implemented on hardware, whereas AES is implemented in both software and hardware. In the year 2000, Rijmen and Daemen developed the Rijndael algorithm, which was later adopted as the AES method by NIST [6]. AES offers a number of benefits over DES, including fast throughput, ease of upgrading, software hardware implementation, portability, and flexibility [7]. The AES algorithm became the most popular cryptographic algorithm in 2006, and it attracted the attention of the entire globe. The size of the key is one of the reasons for switching from DES to AES. The key size of DES is 56 bits, which is breakable by today's hacking techniques. To address the risk of key attacks, AES uses a 128-bit key, a 192-bit key, or a 256-bit key, depending on the number of rounds [3, 8].

Due to the non-readability of hardware functionality and the fact that it cannot be manipulated by an intruder, hardware implementations of algorithms are considered more secure than software. Because standard hardware equipment lacks flexibility, the AES code is implemented on advanced hardware devices known as FPGA [9]. FPGAs are dynamic circuits with functionalities that are not fixed and may be modified over time or in response to a need [10]. In FPGAs, SRAM-controlled transistors are utilized, and a large number of logic cells are integrated into the wire network [10]. Iterative, loop unrolling, pipelining, sub-pipelining, and parallel architectures have been implemented on FPGAs. Some researchers have compared the hardware implementation of DES and AES algorithms [11–13]. Number of architectures are implemented on AES using FPGAs, but no study has been conducted on the throughput analyzes for AES. In this work, we will conduct a thorough analysis of the performance of various architectures implemented on FPGAs. The designs are compared

on the basis of throughput, hardware area, cost, and complexity. The literature study of AES algorithm implementations is depicted in Sect. 2. Section 3 goes through the theoretical background of the AES algorithm. In Sect. 4, the various architectures used in AES are discussed. The comparison table and performance analysis of implemented AES designs are included in Sect. 5. The conclusion and references are included in the next section.

2 Literature

The well-known AES algorithm first came into light in 2001 after the acceptance of Rijndael algorithm. Deaman and Rijman [14] introduced a new algorithm for the cryptography techniques. In their proposed algorithm, they represented mathematical preliminaries and then the round specifications and functionalities and also the key scheduling. Later, NIST accepted Rijndael algorithm as standard and named it as AES. Elbirt et al. [9] presented their work in IEEE transactions, in which the AES algorithm was finalized. They built RC6, Rijndael, snake, and Twofish algorithm designs on FPGA in their work, and they tested these algorithms in four modes: speed optimized non-feedback mode, area optimized non-feedback mode, speed optimized feedback mode, and area optimized feedback mode. The result is that the Rijndael algorithm performs well in hardware implementation and is the titled as AES algorithm. Wollinger and Paar [15] have published an article about the security of FPGAs. They examined the various types of attacks in their article and presented a research on how FPGAs are effective at coping with them. Zhang and Wang [16] worked on the AES algorithm and used FPGA to construct the AES outer-rounder pipelined design. They utilized block RAM to store the S-box values in their research. The provided system has a throughput of 34.7 Gbps at 271 MHz. Yoo et al. [17] developed a vertex II pro FPGA version of the AES algorithm in 2005. They used a parallel architecture for the AES algorithm, which resulted in a system throughput of 29.77 Gbps. Fan and Hwang [8] presented a study on completely sequential and pipelined design for the AES algorithm on FPGAs in 2007. To achieve high throughput, a content addressed memory-based approach is utilized. They employed a high-speed mix column block and a low-cost add round key architecture in their suggested work. For architectural designs, Xilinx ISETM 7 and XSTTM IDEs are utilized. The sequential design has a throughput of 0.876 Gbps, whereas the pipelined architecture has a throughput of 32 Gbps. For the recursive algorithm architecture on FPGA, Sklyarov [18] represents code for the recursive algorithm. He suggested a hierarchical recursive finite state machine paradigm. For implementation, the Xillin Spartan-II series is employed. Swankoski et al. [19] proposed two FPGA architectures for the AES algorithm. The first is a completely parallel loop unrolled design, while the second is a 32-bit datapath round-based architecture. The performance of the unrolled architecture is superior in terms of speed to area since there is no controller overhead, but it uses 1 2 million gate devices and has a throughput of 25 Gbps. For the first architecture, Xillin Spartan-III is utilized. The AES algorithm is investigated by Deshpande and Bhosle

[20] for parallel, pipelined, and sequential systems. For the implementation of these three designs, they employed 11 keys. In terms of throughput and area, a finite state machine is suggested and proven to be efficient. Kumar et al. [3] developed AES on the vivado ISE design suit in 2020, and the results were shown in Artix 7 FPGAs. During implementation, 3987 slice tables, 4115 lookup tables, 269 I/O, and 1 global buffer are used.

There have been a lot of papers published on AES implementation on FPGA, with various architectures being utilized for AES implementation. In this survey, we evaluated and analyzed the performance of different architectures so that one may have a good understanding of the performance of each architecture and select the best architecture for his work domain.

3 AES Algorithm

AES algorithm encrypts 128-bit plain text into cipher and cipher text into plain text using identical key [11]. The length of the key is of 128-bits for 10 rounds, 192-bits for 12 rounds, and 256-bits for 14 rounds. 128-bit plain text which is input to the AES algorithm is firstly converted into data block of 4×4 matrix having 16 elements, and each element is of one byte [21]. The algorithm is subdivided into three parts [22]. The key is XORed to the input matrix in the first stage, while the second stage includes nine cipher rounds, each with four functions: subBytes(), shiftRows(), mixColumn(), and addRoundKey() and in third stage which is also called the 10th round implements three functions which are subBytes(), shiftRows(), and addRoundKeys() [23]. The structure of AES algorithms is shown in Fig. 1. The same structure as in encryption is utilized for decryption, but the inverse of all functions is employed, as illustrated in Fig. 1.

The following are the four functions that we are using during the span of 10 rounds:

3.1 subBytes Transformation

Using the substitution box [24], each byte of the input matrix is replaced with a new byte in the subByte function. The substitution box is created using multiplicative inverse over GF(2^8), as well as a multiplicative inverse affine transformation. The inverse subByte function is used in the decryption step to obtain the original value [21].

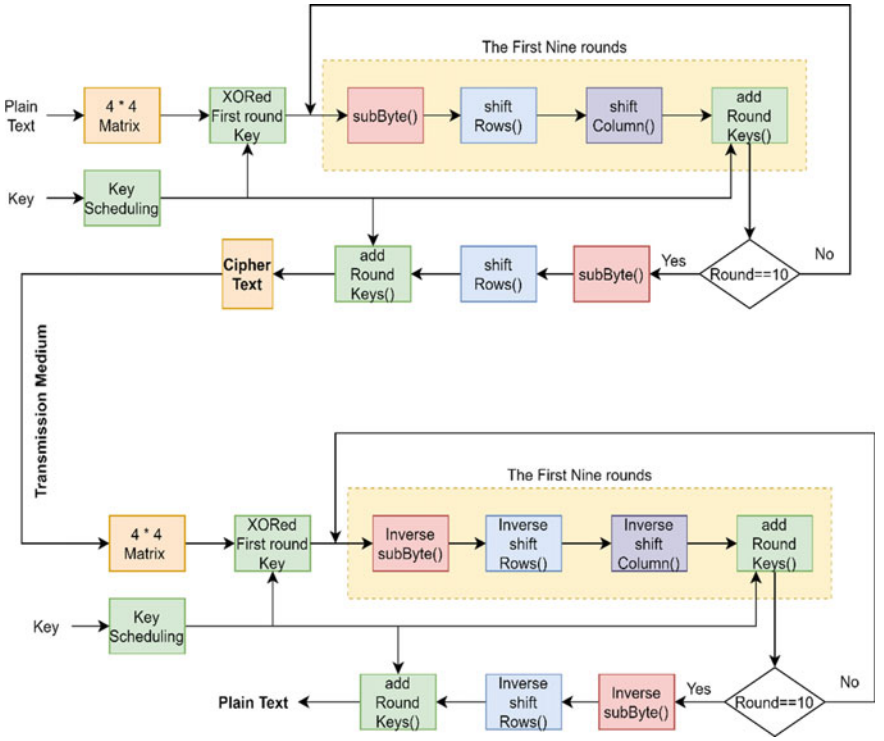


Fig. 1 Basic structure of AES algorithm

3.2 shiftRow Transformation

Four rows of the matrix are moved to the left in this transformation. The first row remains intact, but the second is moved by one cycle, the third by two, and the fourth by three. The identical method is used in the inverse shifting row, but the shifting is done to the right side.

3.3 mixColumn Transformation

Each column is treated as a polynomial in this transformation, and it is multiplied by a fixed polynomial before being moduled with x^4 + 1 [17]. Inverse mixColumn follows the inverse of the multiplication to get plain text.

3.4 addRoundKey Transformation

The key is created in the key scheduling block, where a 128-bit key is permuted using shifting row or column operations before being used to the encryption rounds. The mixColumn output is then XORed with this key, and a new XORed output is delivered to the following round [14].

4 Different Architectures for AES

The fundamental objective of any hardware architecture is to achieve high throughput while using the least amount of space possible. A variety of hardware architectures have been designed that meet these objectives. From which some popular architectures are explained below:

4.1 Iterative Architecture

Iterative architecture is the most efficient technique since it requires the least amount of hardware. The architecture of iterative AES is shown in Fig. 2. This design has a low register to register latency, but encryption requires a significant number of clock cycles. The only expense associated with this approach is the hardware needed for round key and S-box multiplexing [25]. The cost of operating in RAM mode can be lowered by employing a 4-bit lookup table within the configurable logic blocks. Lookup tables are used to store round keys; one lookup table may hold up to 16 round keys, eliminating the need for registers, which are expensive in terms of hardware [9].

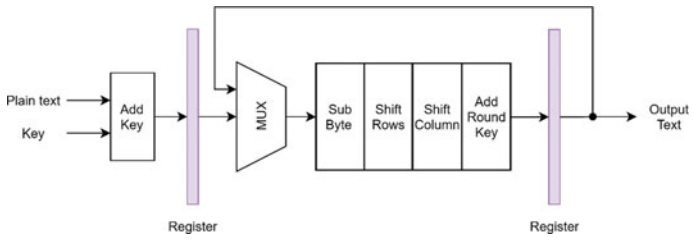


Fig. 2 Iterative architecture

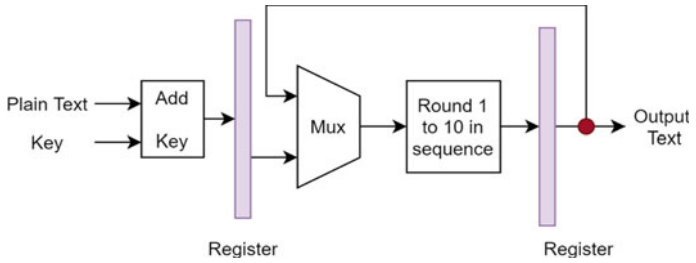


Fig. 3 Sequential architecture [8]

4.2 Sequential or Loop Unrolling Architecture

Multiple rounds are unrolled, up to the total number of rounds required for the cipher text. All loops are unrolled in this design, and the entire algorithm is implemented as a single combinatorial logic block. This design has a significant hardware demand. There are two new registers, as well as input and output interfaces. To encrypt 128-bit plain text, 11 clock cycles are needed [25]. The architecture of sequential architecture is depicted in Fig. 3.

4.3 Pipelined Architecture

Because a large number of blocks are processed at the same time, the pipeline layout allows for high throughput. When round 2 is run for the cipher text production, 1st one will not be on ideal state, the next plain text is supplied as input to the round 1st, while 2nd is processing the previous input. This may be accomplished by inserting a register between the intermediate rounds. The critical route latency is reduced by adding registers [16]. The pipeline architectures are further subdivided into different categories. In outer round only pipelined architecture, pipelining is applied only in first round and in last round. While in external pipeline architecture, all 10 rounds are pipelined, and throughput is high. In sub-pipelining, the subByte step of all rounds is also subdivided into four steps, and all steps are pipelined, so that delay can be overcome. In sub-pipelined, the hardware area is quite high, but the throughput is also very good. Figure 4 depicts the pipelined design.

4.4 Parallel Architecture

For achieving maximum utilization of FPGAs logic cells, gates, and I/O resources, parallel architecture is implemented, which in return has high throughput and

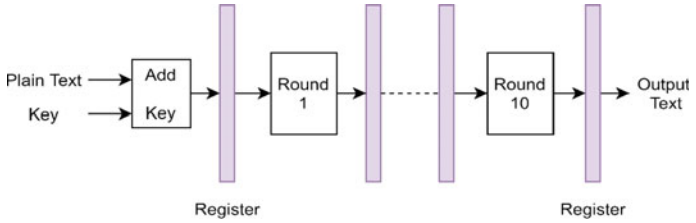


Fig. 4 Pipelined architecture [16]

maximum utilization of resources. The parallel designing provides more degree of flexibility.

Two or more than two pipeline architectures are implemented in parallel, utilization of resources is done efficiently, and performance is also increased [7]. Parallel architecture allows multiple parallel encryptions of plain text. Independent key scheduling is used in parallel architectures as if shared key is used by two parallel blocks, these blocks are not at the same point in the encryption [17]. The parallel processing of AES is depicted in Fig. 5. Multi-core systems are used for the implementation of parallel architecture where two or more cores are in a single processor. In simple words, core is a processing or execution unit. In multi-core systems, each core has its own execution pipeline and also has resources to run the block independently [26].

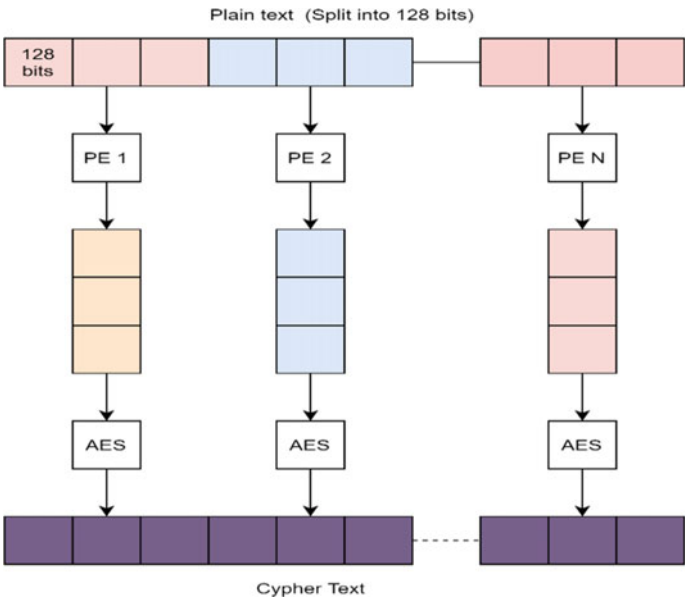


Fig. 5 Parallel architecture

5 Performance Analysis of AES Architectures

Because of its widespread use, the AES algorithm is utilized in nearly every aspect of security. A number of AES architectures have been implemented to achieve reduced costs and higher output by knowing the device or algorithm's demand for throughput. In literature section, a number of researcher have contributed significantly to the field of AES and have built hardware designs of algorithms by describing the throughput of the architecture in terms of time complexity and hardware area complexity. There has been no comprehensive research of the various hardware design algorithms for AES till date. We compared implemented designs and performed a performance study based on the throughput of their implementation in this part. A performance analysis table for the AES algorithm on FPGAs utilizing various designs is shown in Table 1. The comparison is made up of seven criteria that describe implementation specifics. The defined characteristics are the author's implemented architecture, the device used to implement that architecture, the mode used (mode is a method of determining whether output encrypted data is given to the next input plain text or not), the slices used, the frequency and then the number of cycles used to run a block, and the main parameter throughput of the implemented system.

AES FPGAs have been used by reference no. 9 to construct three designs, as indicated in the Table 1. The work [9] has been approved by NIST as a standard. AES throughput has been improved in the sources [8, 13, 19, 25, 27] by using a pipelined design that includes inner pipelining, outer pipelining as well as deep pipeline, sub-pipeline, and hybrid pipelining architectures, compared to source [9]. Sequential architecture's throughput rate is not very great [8]. The iterative framework has been developed by authors [24] and [32] and has a throughput of 14.383 Gbps on the average. The authors [7, 14, 16, 20, 26] worked on parallel architectures, implementing parallelism via loop unrolling, pipelining, and lookup tables. With minimal time complexity and large throughputs, parallel architecture provides its own set of advantages.

6 Conclusion

This article offers a detailed analysis of the AES algorithm and the architectures being used to implement it. Different architectures are developed to meet the needs of various requirements. Some architectures are focused on low hardware, while others are focused on high throughput, while yet others are focused on maximizing resource utilization. In this study, we conducted a survey of existing work that compares the performance of several designs for the AES algorithm on FPGAs. It has been concluded that iterative design is well suited in terms of lowest hardware area, but that parallel architecture with pipelining between rounds leads to efficient resource utilization and high throughput. A completely unrolled three-stage inner and outer double pipelined design and parallel lookup table architecture provide the greatest throughput of 60.29 Gbps in the surveyed articles.

Table 1 Different architectures implemented on AES and different perimeters used during implementations

S. No.	Author	Architecture	Device used	Mode	Slices	Frequency	Cycles (MHz)	Throughput
1	Elbirt et al. [9]	Loop unrolling with two rounds having speed optimized feedback mode and non-feedback mode	VHLD code on FPGA, Xilinx Virtex XCV1000BG560-4	ECB (Electronic codebook mode)/cipher text feedback mode (CFB)	5302	14.1 MHz	6 per block	300 Mbps
2	Elbirt et al. [9]	Pipeline architecture with two stages having speed optimized non-feedback mode	Xilinx Virtex XCV1000BG560-4	ECB	10,533	20 MHz	2.2 per block	1165 Mbps

(continued)

Table 1 (continued)

S. No.	Author	Architecture	Device used	Mode	Slices	Frequency	Cycles (MHz)	Throughput
3	Elbirt et al. [9]	Sub-pipeline architecture, a partially pipelined architecture with five stages and one Sub-pipeline stage per pipeline stage having speed optimized non-feedback mode	Xilinx Virtex XCV1000BG560-4	ECB	10,992	31.8 MHz	2.1 per block	1937.9 Mbps
4	Chodowiec et al. [27]	Pipeline architecture	Virtex XCV-1000BG560-6	ECB	12,600	100–130 MHz	22 complete one cipher text generation	12.2 Gbps
5	Swankoski et al. [7]	Parallel architecture with 10 blocks	Vertex-II pro FPGAs (XC2VP50)	ECB	20,249	150.61 MHz	Not mentioned	19.28 Gbps
6	Yoo et al. [17]	Parallel pipelined architecture	Virtex-II Pro FPGA (chip contains same 10 units, each unit perform one round of algorithm)	ECB	6541	222.2 MHz	7.96 ns for one complete 128-bit cipher text generation	29.77 Gbps

(continued)

Table 1 (continued)

S. No.	Author	Architecture	Device used	Mode	Slices	Frequency	Cycles (MHz)	Throughput
7	Good and Benaissa et al. [16]	Fully parallel loop unrolled architecture	Xilinx Spartan-III (XC3S2000)	ECB	25,107	196.1 MHz	70 cycles for all rounds, to pass AES cipher	25 Gbps
8	Fan and Hwang [8]	Sequential architecture	Xilinx ISE™ 7 (XC2V3000)	Not mentioned	7617	75.3 MHz	11 clock cycles required for encrypting 128-bit	0.876 Gbps
9	Fan and Hwang [8]	Fully pipelined architecture	XST™ (XC4VLX200)	Not mentioned	86,806	250 MHz	1 for complete cipher of the 128-bit text	32 Gbps
10	Wang et al. [22]	Multi-core architecture	Vertex-II XC2V6000	ECB/ CBC (cipher block chaining)	27,561	102 MHz	3 for one cipher output	1.29 to 3.7 Gbps
11	Zhang and Wang [16]	Outer round only pipelined architecture	Virtex-II Pro FPGA XC2VP70	ECB	2389 CLB with 200 Block RAMs	271.15 MHz	11 clock cycles required for encrypting 128-bit	34.7 Gbps
12	Borker et al. [24]	Iterative architecture	Xilinx's Vertex XCV600 BG 560–6	CFB	1853 slices, 512 flip flops, 3645 lookup tables	140.39 MHz	51 for 128-bit data encryption	352 Mbps

(continued)

Table 1 (continued)

S. No.	Author	Architecture	Device used	Mode	Slices	Frequency	Cycles (MHz)	Throughput
13	Rahimunnisa et al. [28]	Parallel operation in the folded architecture	Virtex-6 XC6VLX75T	Not mentioned	2056 registers, 3788 LUTs, 48 block RAMs	505 MHz	1.978 ns	37.1 Gbps
14	Deshpande and Bhosle [20]	Parallel as encryption and decryption is done parallel, pipelined as all rounds are accepting input and giving output to next round	Implemented in Zynq (xc7z020-2clg484) device and tested on Zedboard	Not mentioned	Slice registers-2008 Lookup tables (LUT)- 1364	239.648 MHz	1 for complete cipher of 16 byte data	5.75 Gbps
15	Yadav et al. [29]	Outer pipeline architecture and inner line pipelining on different stages	Xilinx 9.2 on Spartan and Virtex series (xc4vlx20)	ECB	2273 slices and LUTs	388.667 MHz	11 Clock cycles	49.49 Gbps
16	Zhang et al. [30]	Fully unrolled three-stage inner and outer double pipelined architecture and parallel lookup tables	Vertex-6 (XC6VLX240T)	Not mentioned	2252 slices, 244 block RAMs	470.99 MHz	31 clock cycles for on cipher output	60.29 Gbps

(continued)

Table 1 (continued)

S. No.	Author	Architecture	Device used	Mode	Slices	Frequency	Cycles (MHz)	Throughput
17	Zodpe and Sapkal [21]	New hybrid non-pipelined (PN sequence generator is used for S-box values)	Spartan6 XC6SLX150-3FGG900 FPGA	Not mentioned	5566	237.45 MHz	10 clock cycle to generate output	30.39 Gbps (security is enhanced)
18	Chen et al. [31]	Deep pipeline architecture and full expansion technology	Implemented with Vivado HLS (v2016.4) built on Xilinx xc7k325tffg676-2 1 chip	ECB	Block RAM- 288, flip flop- 13,647, lookup tables-1,26,040	Not mentioned	62 clock cycles for cipher of given plain text	31.30 Gbps
19	Murugan et al. [32]	Iterative architecture with sub-pipelined S-box technique	Spartan-3, virtex-4	Cipher feedback mode	Slices-1132, flip flops- 680, 4 in–out UTs-2156	112.37 MHz for vertex-4. 67.75 MHz for Spartan-3	Not mentioned	14,383 Mbps for vertex-4, 8672 Mbps for Spartan-4

References

1. Diffie W, Hellman ME (2019) New directions in cryptography. Secur Commun Asymmetric Cryptosyst, 143–180
2. Jindal P, Kaushik A, Kumar K (2020) Design and implementation of advanced encryption standard algorithm on 7th series field programmable gate array. In: 2020 7th international conference on smart structures and systems (ICSSS), pp 1–3
3. Kumar K, Ramkumar KR, Kaur A (2020) A design implementation and comparative analysis of advanced encryption standard (AES) algorithm on FPGA. In: 2020 8th international conference on reliability, infocom technologies and optimization, pp 182–185
4. Thakur J, Kumar N (2011) DES, AES and Blowfish: symmetric key cryptography algorithms simulation based performance analysis. Int J Emerging Technol Adv Eng 1(2):6–12
5. Chandra S, Paira S, Alam SS, Sanyal G (2014) A comparative survey of symmetric and asymmetric key cryptography. In: 2014 international conference on electronics, communication and computational engineering (ICECCE), pp 83–93
6. Rijmen V, Daemen J (2001) Advanced encryption standard. In: Proceedings of federal information processing standards publications. National Institute of Standards and Technology, pp 19–22
7. Swankoski EJ, Brooks RR, Narayanan V, Kandemir M, Irwin MJ (2004) A parallel architecture for secure FPGA symmetric encryption. In: Proceedings of the international parallel and distributed processing symposium (IPDPS 2004) (Abstracts CD-ROM), vol 18, pp 1803–1810
8. Fan CP, Hwang JK (2018) FPGA implementations of high throughput sequential and fully pipelined AES algorithm. Int J Electr Eng 15:447–455
9. Elbirt AJ, Yip W, Chetwynd B, Paar C (2001) An FPGA-based performance evaluation of the AES block cipher candidate algorithm finalists. IEEE Trans Very Large Scale Integr Syst 9:545–557
10. Prasanna VK, Dandalis A (2004) FPGA-based cryptography for internet security. Perform Eval, 1–6
11. Devi A, Sharma A, Rangra A (2015) A review on DES, AES and blowfish for image encryption & decryption. Int J Comput Sci Inf Technol 4(6):12646–12651
12. Yazdeen AA, Zeebaree SR, Sadeeq MM, Kak SF, Ahmed OM, Zebari RR (2021) FPGA implementations for data encryption and decryption via concurrent and parallel computation: a review. Qubahan Acad J 1(2):8–16
13. Padmavathi B, Kumari SR (2013) A survey on performance analysis of DES, AES and RSA algorithm along with LSB substitution. IJSR, 2319–7064
14. Daemen J, Rijmen V (1999) AES proposal: Rijndael, NIST AES website csrc.nist.gov/encryption/aes
15. Wollinger T, Paar C (2003) How secure are FPGAs in cryptographic applications? Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 2778, PP 91–100
16. Zhang Y, Wang X (2010) Pipelined implementation of AES encryption based on FPGA. In: Proceedings of the 2010 IEEE international conference on information theory and information security, pp 170–173
17. Yoo SM, Kotturi D, Pan DW, Blizzard J (2005) An AES crypto chip using a high-speed parallel pipelined architecture. Microprocess Microsyst 29(7):317–326
18. Sklyarov V (2004) FPGA-based implementation of recursive algorithms. Microprocess Microsyst 28(5–6, SPEC. ISS.):197–211
19. Good T, Benaissa M (2005) AES on FPGA from the fastest to the smallest. Lect Notes Comput Sci 3659:427–440
20. Deshpande PU, Bhosale SA (2016) AES encryption engines of many core processor arrays on FPGA by using parallel, pipeline and sequential technique. In: International conference on energy systems and applications (ICESA 2015), no Icesa, pp 75–80
21. Zodpe H, Sapkal A (2020) An efficient AES implementation using FPGA with enhanced security features. J King Saud Univ-Eng Sci 32(2):115–122

22. Wang MY, Su CP, Horng CL, Wu CW, Huang CT (2010) Single- and multi-core configurable AES architectures for flexible security. IEEE Trans Very Large Scale Integr Syst 18(4):541–552
23. Mali M, Novak F, Biasizzo A (2005) Hardware implementation of AES algorithm. J Electr Eng 56(9–10):265–269
24. Borkar AM, Kshirsagar RV, Vyawahare MV (2011) FPGA implementation of AES algorithm. In: ICECT 2011—2011 3rd international conference on electronics computer technology, vol 3, pp 401–405
25. Standard AE, Tv HD (2007) architectural designs for the advanced encryption standard. Cryptogr Algorithms Reconfigurable Hardw, 245–289
26. Nagendra M, Chandra Sekhar M (2014) Performance improvement of advanced encryption algorithm using parallel computation. Int J Softw Eng Its Appl 8(2):287–296
27. Chodowiec P, Khuon P, Gaj K (2011) Fast implementations of secret-key block ciphers using mixed inner- and outer-round pipelining. In: Proceedings of the 2001 ACM/SIGDA ninth international symposium on Field programmable gate arrays, pp 94–102
28. Rahimunnisa K, Karthigaikumar P, Rasheed S, Jayakumar J, SureshKumar S (2014) FPGA implementation of AES algorithm for high throughput using folded parallel architecture. Secur Commun Netw, pp 2225–2236
29. Yadav D, Rajawat A (2016) Area and throughput analysis of different AES Architectures for FPGA implementations. In: 2016 IEEE international symposium on nanoelectronic and information systems (iNIS), pp 67–71
30. Zhang X, Li M, Hu J (2018) Optimization and implementation of AES algorithm based on FPGA. In: 2018 IEEE 4th international conference on computer and communications (ICCC 2018), pp 2704–2709
31. Chen S, Hu W, Li Z (2019) High performance data encryption with AES implementation on FPGA. In: Proceedings of IEEE 5th international conference on big data security on cloud (BigDataSecurity), IEEE international conference on high performance and smart computing,(HPSC) and IEEE international conference on intelligent data and security (IDS), pp 149–153
32. Arul Murugan C, Karthigaikumar P, Sathya Priya S (2020) FPGA implementation of hardware architecture with AES encryptor using sub-pipelined S-box techniques for compact applications. Automatika 61(4):682–693

Designing Adiabatic Techniques for Logic Circuits

Himani Sharma, Nidhi Sharma, and Surya Deo Choudhary

Abstract VLSI designers are increasingly concerned with power dissipation and chip size, as we all know. Desperate measures have been taken to reduce energy consumption, such as the use of low-tech adiabatic technology, which significantly relies on parameter variations DFAL and ECRL, and two logic families are examined. In this paper gates such as inverter and NAND gates with parameters like as load capacitance and supply voltage fluctuation are used to demonstrate the basic concepts and intricacies. Inverter and 2 × 1 MUX simulations are then performed using microwind. We can now compare these circuit characteristics to those of a CMOS circuit and see that adiabatic technology reduces heat dissipation and size.

Keywords Adiabatic · Power dissipation · DFAL · ECRL · INVERTER · 2 × 1 MUX1

1 Introduction

We might not be very much verse with the term 'adiabatic' which means no loss of energy in the form of heat dissipation [1]. Ever since the technologies are growing tremendously, demand for energy is increasing proportionally with it. Integrated chips thus come into a play where a large number of transistors are used. But with this, increase in power and energy is observed. Now, it becomes evident to develop a technology that does not dissipate energy and thus design on low-power consumption.

Therefore, we have adopted 'Adiabatic Technology' that utilizes energy efficiently [2].

H. Sharma (✉) · N. Sharma · S. D. Choudhary
Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida, India
e-mail: himanishrm7@gmail.com

N. Sharma
e-mail: nidhisharma@niet.co.in

S. D. Choudhary
e-mail: sdchoudhary@niet.co.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_5

It is the most convenient method to reduce energy consumption. Figure 1 shows the pull-up and pull-down network. In conventional CMOS, we have a pull-up and pull-down network where IN (input) is connected to gate terminal, and OUT (output) is connected to drain terminal. As the input voltage of CMOS varies from 0 to 5 V, the state varies accordingly. Thus, CMOS acts as a switch which charges and discharges the load capacitor C. Figure 2 shows the charging and discharging circuit for CMOS.

From high to low pulse, NMOS is off, whilst the PMOS stays on. This leads to the instant charging OUT to logic high and thus help in charging of load capacitor C [3, 4].

Energy drawn here is $C * V_{dd}{}^2$ but only$1/2C * V_{dd}{}^2$ is stored in load capacitor, and half is dissipated in PMOS network. A charge of $Q = C * V_{dd}$ is taken from the voltage source, an energy quantum of

$$E * V_{dd} = Q * V_{dd} \tag{1}$$

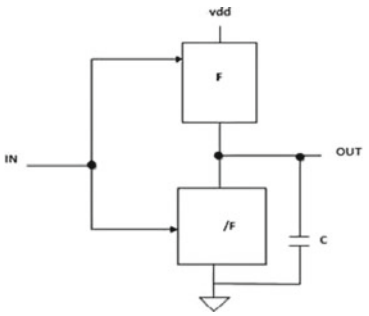


Fig. 1 Pull-up and pull-down network

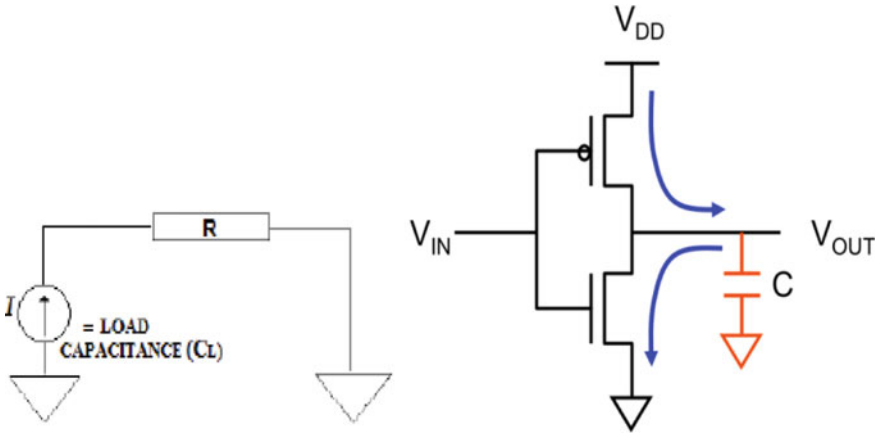


Fig. 2 CMOS charging and discharging circuit

$$= CL * V_{dd} \tag{2}$$

From low to high, dissipation of energy is carried out by the stored energy in load capacitor via NMOS.

$$E = 1/2\alpha C * V_{dd}^2 \tag{3}$$

where α is switching probability of the circuit.

The advantage of adiabatic process is that it follows 3R's: reduce, reuse, and recycle. It reduces energy dissipation and reuses the energy by recycling method. Thus, energy stored can be reused again without any loss.

2 Adiabatic Techniques

An adiabatic system mainly consists of digital core design made up of adiabatic circuits and the power clock generator. Two adiabatic logic families are discussed in the current paper—efficient charge recovery logic (ECRL) and diode free adiabatic logic (DFAL).

2.1 Efficient Charge Recovery Logic (ECRL)

Efficient charge recovery logic is formed by using two cross-coupled PMOS transistors and two NMOS structures as shown in Fig. 3. It uses AC power supply for recovering and reusing the supplied energy. In this technique, the constant load capacitance is always generated by the power clock which does not depend on input signal [5].

This methodology eliminates the pre-charge diode and evaluates at the same time. Hence, it dissipates less energy as compared to the adiabatic systems and full output swing can be obtained both in pre-charge and recovery phase. The circuit also suffers from adiabatic loss due to the threshold voltage of PMOS transistors both in pre-charge and recovery phase due to which it fails to give output for full swing. The effectiveness of this approach is shown by the design of inverter and 2:1multiplexer.

2.1.1 ECRL Inverter

The two inverters are connected to each other in back-to-back manner in ECRL inverter as shown in Fig. 4. When power clock (PWR) is given to Vdd, depending on the inputs 'IN' and 'IN̲', the outputs 'out' and '\out' change their values. The layout design and simulated waveform of ECRL inverter are shown in Figs. 5 and 6,

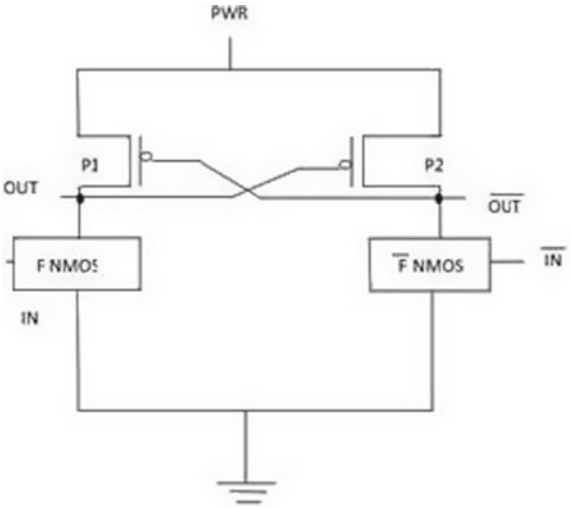


Fig. 3 Basic circuit of ECRL

respectively. When clock goes from low to high, if the input is 1, the 'out' is 0, and '\out' is 1. Now, we can use these output values as input for the next stage. And as the clock goes from high to low, the value of '\out' is given back to PWR. Hence, the energy is saved [6].

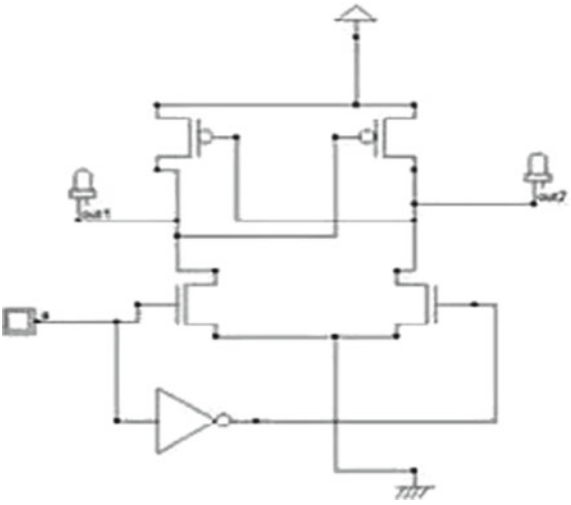


Fig. 4 ECRL inverter

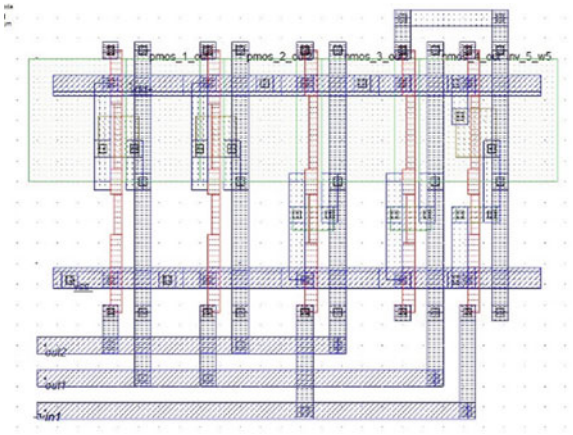


Fig. 5 Layout design of inverter using ECRL

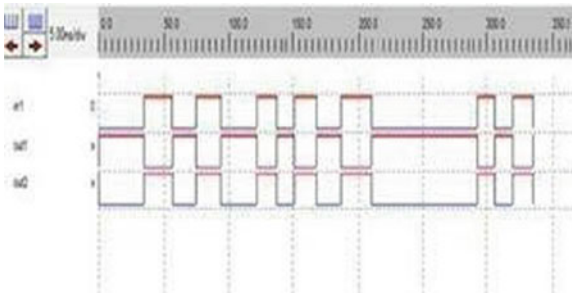


Fig. 6 Simulated waveform of inverter using ECRL

2.1.2 ECRL 2:1 Mutilpexer

The schematic diagram of ECRL 2:1 multiplexer is shown in Fig. 7. When a clock goes from low to high, and if select line(S) is 0, 'out' will be B, and when S is 1, '\out' changes to 'A'. Then, the output will be used as input for the next stage. Now when the clock goes from V_{dd} to zero, the circuit is able to recover the supplied charge as the value of high output is given back to power clock. The layout design and simulated waveform of 2:1 multiplexer using ECRL are shown in Figs. 7 and 8.

2.2 Diode Free Adiabatic Logic (DFAL)

Diode free adiabatic logic circuits were introduced in order to remove the problems associated with diode-based logic families like delay, complex structure, power dissipation. DFAL topology does not have any diode in its charging path and discharging path [7], thus it is diode free. An inverter and 2:1 multiplexer are constructed using this technology and explained below.

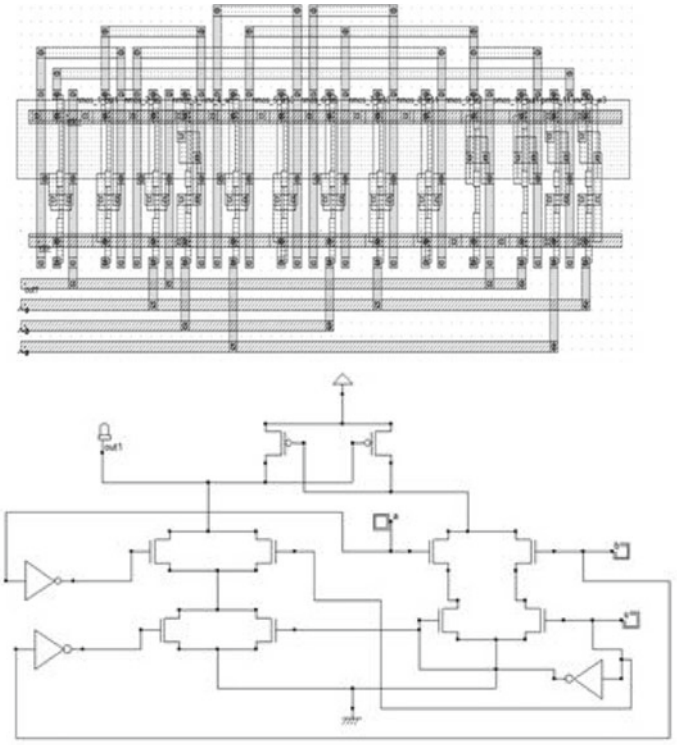


Fig. 7 ECRL 2:1 multiplexer and layout design of 2:1 MUX using ECRL

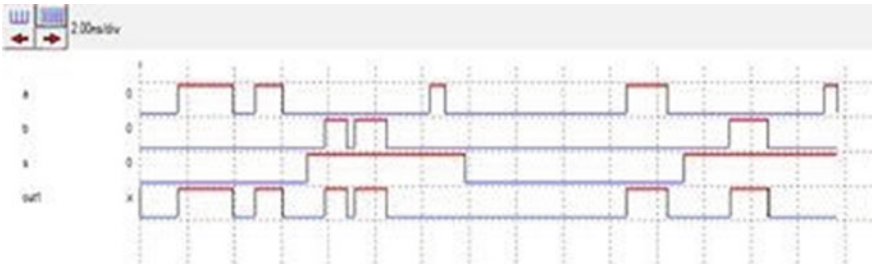


Fig. 8 Simulated waveform of 2:1 MUX using ECRL

2.2.1 DFAL Inverter

Figure 9 shows the circuit diagram and simulated waveforms of an inverter using DFAL technology. One characteristic that excel DFAL [8] from other adiabatic technique is that the load capacitance is charged and discharged using a power clock

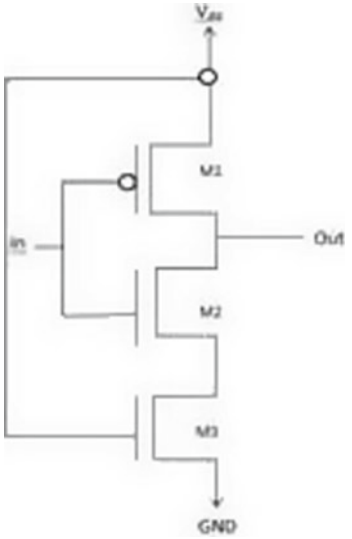


Fig. 9 DFAL inverter

that is sinusoidal and split level. This results in further minimization of power dissipation. The pull-down network consists of a NMOS transistor unlike conventional diode which had been used for discharging. This transistor is controlled by power clock, and it recycles charge from the output node; hence, adiabatic losses are further reduced. The layout design of inverter using DFAL is shown in Fig. 10.

2.2.2 DFAL 2:1 Multiplexer

Multiplexers are used for selecting one of n given inputs and to route it to the output. In Fig. 10 when select line is low, it will give '*a*' as output, whereas when select line(s) is high, it will give '*b*' as output (Fig. 11).

$$Y = as + bs \tag{4}$$

Now, using DFAL method for designing 2:1 MUX, power loss can be avoided as compare to the CMOS logic design. In DFAL method when select line is high, output of inverter tends to become low, thus it gets connected to NMOS. Hence, the transistor turns on. On the other hand, when select line is low, the output of inverter becomes high which gets connected to PMOS as a result transistor turns on. In this circuit, we have used an NMOS transistor in the pull-down network, which recycles charge from the output node. Figure 12 shows the layout design of 2:1 multiplexer using DFAL technique.

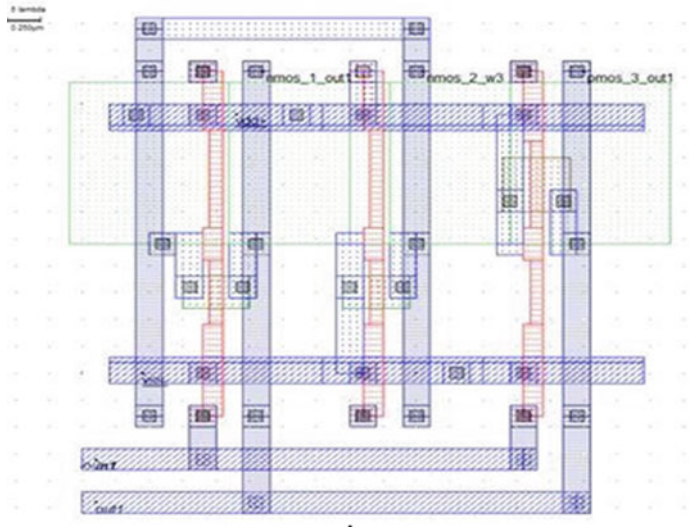


Fig. 10 Layout design of inverter using DFAL

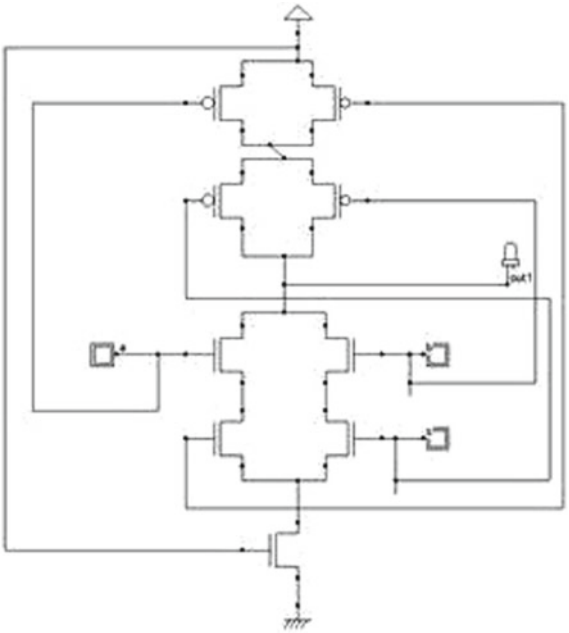


Fig. 11 DFAL 2:1 multiplexer

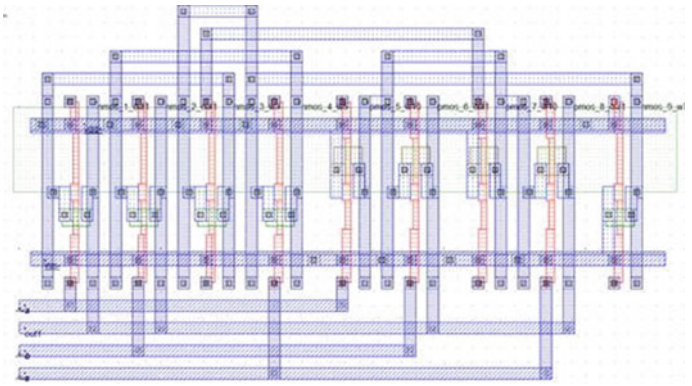


Fig. 12 Layout design of 2:1 multiplexer using DFAL

3 Power Consumption Comparison and Verification

The power consumption of circuits can be compared on the basis of following parameter:

Load capacitance and voltage supplied.

3.1 Variation in Load Capacitance

Figures 13 and 14 depict the variation between power dissipation per cycle with respect to load capacitance of the inverter and 2:1 MUX circuit of two adiabatic families, i.e. ECRL and DFAL. It shows that DFAL is better than ECRL as [9], it is showing more power dissipation with increasing load capacitance, whereas DFAL power is comparatively showing less effect with respect to load capacitance [10].

3.2 Supply Voltage Variation

Figures 15 and 16 depict the variation between power dissipation per cycle with respect to supply voltage of the inverter and 2:1 MUX circuit of two adiabatic families, i.e. ECRL and DFAL. It implies that the DFAL inverter is preferable to the ECRL inverter since the ECRL inverter's power dissipation increases as the supply voltage increases, whereas the DFAL's power dissipation increases just slightly.

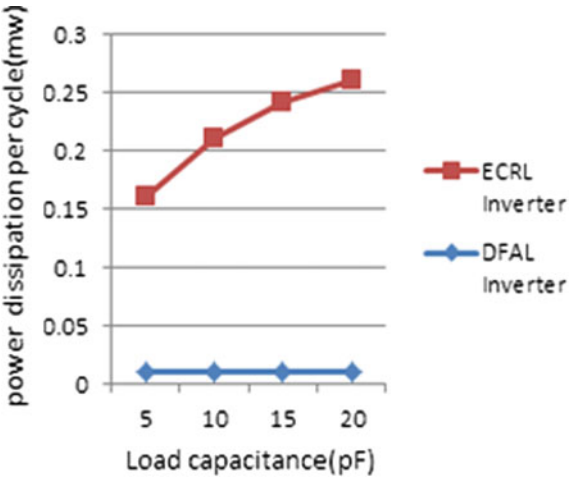


Fig. 13 Pictographical representation of an inverter showing power dissipation with respect to load capacitance at $V_{dd} = 5$ V and freq = 100 MHz

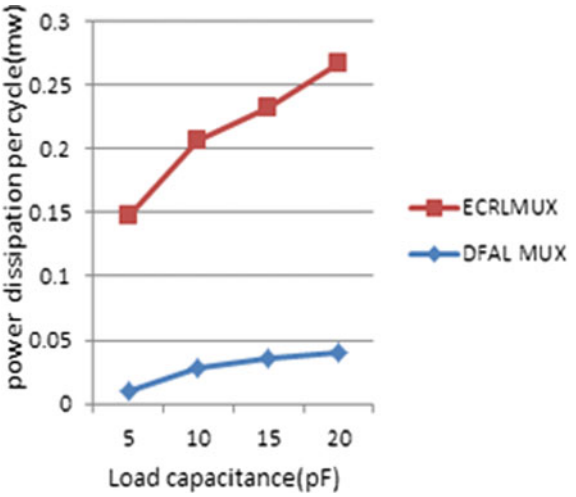


Fig. 14 Pictographically representation of 2:1 MUX showing power dissipation with respect to load capacitance at $V_{dd} = 5$ V and frequency = 100 MHz

4 Conclusion

An adiabatic design method is found to be superior for logic circuits than the more traditional CMOS design technique in this article. When compared to adiabatic families, the conventional CMOS technique dissipates more power. The inverter and 2:1 MUX circuits were designed using the ECRL and DFAL techniques, and their waveforms and layout diagrams were successfully simulated. In comparison with typical CMOS circuits, the layout diagrams demonstrate a reduction in area, and the simulated waveform shows the recycling of energy when the power is dropped. Two metrics, namely load capacitance variation and supply voltage variation, were used

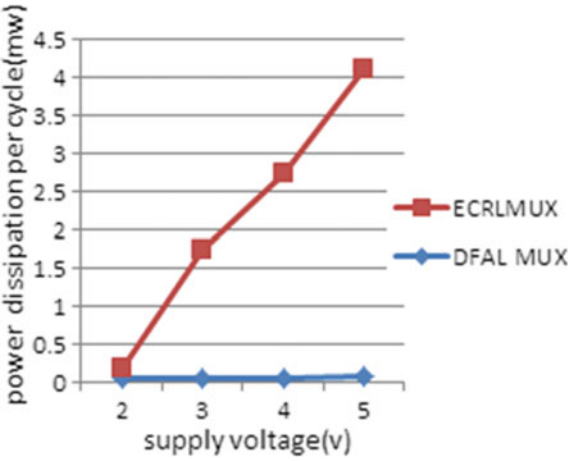


Fig. 15 Pictographically representation of inverter showing power dissipation with respect to supply voltage at load capacitance = 5 pf and frequency = 100 MHz

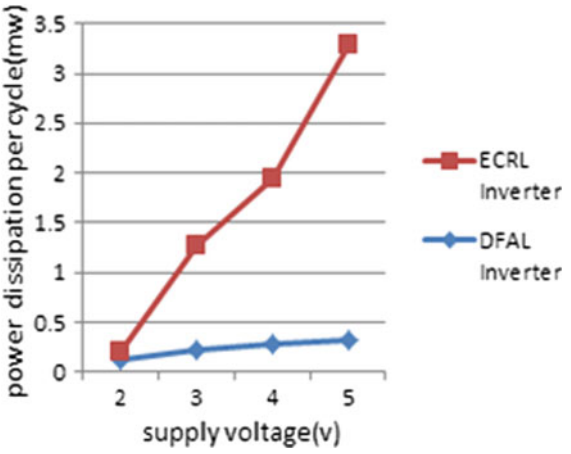


Fig. 16 Pictographical representation of 2:1 MUX showing power dissipation with respect to supply voltage at load capacitance = 5 pf and frequency = 100 MHz

to compare the two approaches in terms of power reduction. There is greater power dissipation consequence vs the parameters in ECRL than there is with as a result, the DFAL and ECRL circuits outperform the conventional CMOS circuit. RAM, ROM, SRAM, and resistors are examples of intricate logic circuits that will be developed in the future. Adiabatic techniques, which is more efficient and uses less power, can be used for fast signal processing.

References

1. Akers LA, Suman R (2002) Adiabatic circuits for low-power logic. IEEE papers, pp 286–289
2. Varga L, Kovacs F, Hosszu G (2001) An efficient adiabatic charge-recovery logic. In:

Proceedings of IEEE South East Conference, pp 17–20
3. Moon Y, Jeong DK (1996) An efficient charge recovery logic circuit. IEEE JSSC 31(04):514–522
4. Athas WC, Svensson LJ, Koller JG, Tzartzanis N, Chou E (1994) Low power digital system based on adiabatic switching principles. IEEE Trans VLSI Syst 2:398–407
5. Sharma K, Pahwa H, Karatangi SV, Rai A (2018) And design of SR flip flop using adiabatic technique. Int J Adv Res Sci Eng 7(Special issue no-5)
6. Upadhyay S, Mishra RA, Nagaria RK, Singh SP (2013) DFAL: diode-free adiabatic logic circuit. Hindawi Publishing Corporation ISRN Electronics
7. Maurya AK, Kumar G (2011) Energy efficient adiabatic logic for low-power VLSI application, international conference on communication system and network technologies
8. Kumar S, Hinnwar M (2014) Implementation of low power CMOS design using adiabatic improved efficient charge recovery logic. Int J Sci Eng Technol, 150–153
9. Kumar A, Sharma M (2013) Design and analysis of MUX using adiabatic ECRL and PFAL. In: ICCAI, pp 1341–1345
10. Chandrakasnand AP, Brodersen RW (1995) Low power CMOS digital design. Kluwer Academic, Norwell, MA, pp 28–31

Accuracy and Performance of Heart Failure Prediction System by Using Backpropagation Network (BPN)

Pawan Kumar Singh, Piyush Yadav, Pankaj Sharma, Virendra Pal Singh, and Durgesh Kumar

Abstract Heart failure is a dangerous and highly prevalent disease (about 2% in the adult population in developed countries, and more than 8% in patients older than 75 years). Roughly 3–5% of hospital admissions are related to accidents involving heart disease. In their clinical career, heart disease is the first reason for health-care practitioners to join. The expenditure in the developing world amounts to up to 2% of all health costs. To build an efficient disease management plan, it is necessary to analyze vast quantities of data, diagnose disease early, and to anticipate adverse effects early on. This would prevent the development of the illness, increase patients' quality of life, and lower hospital costs. Artificial intelligence and machine learning methods have been used in this way. This paper aims to present state-of-the-art methodologies for learning machines used for assessing cardiovascular failure, including the following: Naïve Bayes, REPTree, AdaBoost, logistical regression, J48, and the support vector machines. We trained and test above model on WEKA tool,

P. K. Singh (✉)
Department of Computer Science and Engineering, G. L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: pawan8832006@gmail.com

P. Yadav
Department of Electronics and Communication Engineering, G. L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: piyush.yadav@glbitm.ac.in

P. Sharma
Department of Computer Science and Engineering, Ajay Kumar Garg Engineering College, Ghaziabad, India
e-mail: pankajsharma.mnit@gmail.com

V. P. Singh · D. Kumar
Department of Information Technology, Ajay Kumar Garg Engineering College, Ghaziabad, India
e-mail: virendra.pal.singh@gmail.com

D. Kumar
e-mail: durgesh.durge@gmail.com

D. Kumar
Department of Computer Science and Engineering, GNIOT Group of Institutions, Greater Noida, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_6

and it has been observed that artificial neural network model (ANN) like backpropagation network is best model to predict and provide good accuracy and performance of the heart failure patients.

Keywords Heart failure dataset · Naïve Bayes · REPTree · AdaBoost · Logistic regression · J48 · Support vector machine · Backpropagation network and WEKA tool

1 Introduction

Latest findings have shown that cardiovascular disease is one of the leading causes of death worldwide. Cardiovascular disorders or cardiac diseases mean the disorder that affects the functioning of the heart directly or indirectly. Conditions that have blocked blood arteries that can lead to heart attacks and other heart diseases also occur. HF is a debilitating illness linked to high rates of morbidity and mortality. 26 million people are afflicted with HF in the world, and 3.6 million are diagnosed per year, according to the European Society of Cardiology (ESC). 17–45% of the HF patients die in the first year and the remainder die within 5 years. Relative to costs of HF administration, the vast majority are connected with recurring hospital admissions [1–3], account for around 1–2% of all health expenses. HF has become a disease outbreak in Europe and globally with increased frequency, higher cost of medicine, frequent hospitalization, lower quality of life (QoL) and early mortality, highlighting the need to diagnose early, and to provide timely treatment. The clinical procedure is accompanied by ancillary assessments such as blood tests, chest X-rays, electrocardiography and echocardiography, and carefully historical examinations and physical inspection [4]. The synthesis of evidence obtained from the aforementioned diagnostic protocol led to the development of various parameters for evaluating the occurrence of HF (e.g., Framingham, Boston, Göteborg, and the SSC) [5]. When an HF diagnosis has been established, specialists classify the HF gravity using a NYHA or US College of Cardiology guidance system (ACC/AHA), which allows them to select the treatment they need more (medication therapy, food guidelines, physical activity guidelines) [6]. While the complex Pathophysiology of HF is well understood, the amount and sophistication of evidence and knowledge to be evaluated and handled convert the correct and effective diagnosis of HF and the evaluation of therapeutic schemes into daunting and complex activities. The reasons for the huge increase in the application of machine learning techniques to analysis include the positive effects of early HF diagnosis, enabling experts to design an effective, potentially successful treatment plan, preventing a deterioration of the condition, affecting patients' health, improving the quality of the patient and contribute to lowered medical costs. Data mining strategies which are attractive to study groups include the classification processes. Clear definition of the level of illness or the etiology or subtypes enables accurate and targeted delivery of therapies and procedures and helps patient success to be assessed. In order to forecast heart

disease, machine learning methods such as Naïve Bayes, decision tree, ANN, and K-nearest neighbors are being used for heart disease prediction. The classification methods are used in comparative analytics [5]. I used a dataset from Kaggle in this analysis. The classification model is built using classification algorithms to predict heart disease in the WEKA tool. This study includes a discussion of algorithms used for heart failure prediction, as well as a comparison of existing systems.

2 Literature Review

The algorithm for diagnosing HF in a non-acute environment is as follows, according to the ESC guidelines [1]. In the first place, the risk of HF is measured based on the patient's previous health experience, signs, physical testing, and ECG rest. HF is extremely unlikely if all elements are natural. Should at least one factor be irregular, natriuretic peptides plasma should be tested. This assessment enables experts to identify patients with echocardiography. The diagnostic method for HF maybe I time reduction, (ii) assist, and (iii) the application of the computer teaching techniques to the available data with the same precision. HF identification is expressed as the issue of two-class classification when it is either HF that is present or not the contribution of the classifiers. The literature focuses much of the research on the use of heart rate variation (HRV) as a means of classifying a topic as standard or as an HF patient, according to Table 1.

3 Machine Learning Techniques Used

3.1 Naïve Bayes

A supervised algorithm is the classification Naïve Bayes. The Bayes theorem is a basic classification method. It takes strong (naive) attribute freedom. The Bayes theorem is a statistical principle to quantify the likelihood. The predictors are neither interrelated nor dependent to each other. The probability of maximization is contributed by all attributes separately. The paradigm of Naïve Bayes can be used and does not use Bayesian techniques. Naive Bayes classifiers are used in many complex real-world situations [16]:

$$P(X|Y) = \frac{P(Y|X) * P(X)}{P(Y)} \tag{1}$$

The following are $P(X/Y)$, $P(X)$ is the previous probability class, $P(Y)$ is the previous probability predictor, and $P(Y/X)$ is the probability predictor. The Naïve Bayes is a simple, easy to use algorithm that handles complex, nonlinear data. There

Table 1 Literature review methods of HF detection using HRV measures

Authors	Method	Data	Features	Evaluation measures
Asyali et al. [7]	Discriminatory linear regression Classification Bayesian	Data No. 54 standard subjects 29 CHF patients Data source: The PhysioBank interval RR libraries contain long-lasting (to 24 h) ECG recordings of beat annotation files	Features of the predictor HRV measures for the long run	Observed accord rate: 93.24%, Real positive sensitiveness: 81.82% (true negative) specificity: 98.08% Statistics of kappa: 0.832 (interval of 0.689–0.974 confidence of 95%)
			Response feature Normal, CHF	Validation NA
Isler et al. [8]	Feature selection (genetic algorithm) Minmax KNN standardization	Data number 54 normal individuals 29 subjects CHF Source of data including beat recordings of the MIT/BIH interval of long-term ECG records (ever at 24 h)	Predictor features short-term HRV compliance + entropy wavelet steps	*k = 5 Sensitivity 96.43% Specifications: 96.36% Accuracy: 96.39% k = 7 100% sensitivity Specificities: 94.74% Accuracy: 96.39%*
			Feature to respond Ordinary, CHF	Leave-one-out cross-validation
Elfadil et al. [9]	Supervised perception of multilayer	CHF data training data 53 normal sinus rhythm (NSR) and 17 recordings Checks 12 CHF and 12 common people Source of information Spontaneous simulation of Massachusetts data	Features of the predictor The R1 (band 1), R2 (band: two to one), R3 (bands: four to four), R4 (band: eleven to six), R5 (band: seventeen to twenty-first), R6 (bands: 25 to 32)	85.30% sensitivity Specificity: 82.00% Precision: 83.65%
			Feature to respond Ordinary CHF	Availability checks 12 CHF and 12 ordinary individuals
	Unsupervised normalization self-organizing map	Database No Database of the Technology Institute (MIT) Exercise: 1000 CHF + 1000 regular from 17CHF and 53 normal subjects simulated randomly Checks: 12 CHF and 12 ordinary subjects modeled randomly normal 1000 CHF and 1000	Features of the predictor R1 (band 1), R2 (bands: 2 through 3), R3 (bands: 4 through 10), R4 (bands: 11 through 16), R5 (bands: 17 through 24), R6 (bands: 25 to 32)	Sensitivity: 89.10% Specificity: 96.70% Accuracy: 92.90%
			Response feature Normal CHF	Availability checks 1000 CHF and 1000 ordinary individuals

(continued)

Table 1 (continued)

Authors	Method	Data	Features	Evaluation measures
Mellilo et al. [10]	Feature selection of CART	Number of dataset 72 subjects regular 44 subjects of CHF Data source: The normal sinus rhythm RR interview database and the MIT-BIH standard sinus rhythm database were used to exclude normal subjects CHF data were obtained from the heart failure RR interval database and the BIDMC heart failure congestive database	Predictor features Long-term HRV conduct	Specifications 89.74% Sensitivity: 100.0%
			Response feature Ordinary CHF	Availability 10 validation fold cross
Jovic et al. [11]	SVM, MLP, C4.5, Bayesian classifiers	Data No 25 regular subjects 25 subjects for CHF Data source BIDMC database for congestive cardiac failure Database of normal rhythm sinus MIT-BIH Database of normal RR rhythm sinus intervals	Features of the predictor Dimensions of correlation, index of space filling, measure of the central trend, entropy approximate (four features), standard NN (or RR) interval deviation SDNN, root of mean squared differences of N successive RR interval RMSSD, interval difference ratio over 20 ms, overall RR interval pNN20, HRV triangular index	SVM: 77.2% sensitivity Specificities: 87.4% MLP: 96.6% sensitivity Specificities: 97.8% C4.5: 99.2% sensitivity Specificities: 98.4% Bayesian: 98.4% sensitivity Specificities: 99.2%
			Feature to respond Normal CHF	Validation 10 × tenfold cross-validation
Yu et al. [12]	Genetic algorithm (GA) SVM feature selection	Data No.: 54 Standard topics 29 topics of CHF Data source: The research databases (chf2db) and the normal sinus rhythm (NSR) database (nsr2db), both available on the PhysioNET were provided by congestive heart failure (CHF)	Features of the predictor Features based on bispectral analysis	Sensitivity of RBF kernel: 95.55% Specificities: 100% Kernel linear 93.10% sensitivity Specificities: 98.14%
			Feature to respond Normal CHF	Validation Cross-validation leave-one-out

(continued)

Table 1 (continued)

Authors	Method	Data	Features	Evaluation measures
Liu et al. [13]	Selection of features Standardization feature Combination of features KNN and SVM	Data Number Thirty normal subjects of CHF 17 Data source The normal sinus rhythms RR interval database was found for normal subjects The congestive heart failure RR interval database was retrieved from the CHF group	Features of the predictor Short-term HRV action	100.00% SVM accuracy Accuracy: 100.00% 100.00% sensitivity KNN Precision: 91.49% Accuracy: 94.12% 84.21% sensitivity
			Feature to respond Normal, CHF	Validation Validation of the cross
Narin et al. [14]	Selection feature filter based on backward removal: SVM, KNN, LDA, MLP, and RBF	Data no. 54 regular individuals CHF 29 subjects Data source: Data used in this study were obtained from the MIT/BIH database in PhysioNET from the normal sinus and congestive cardiac insufficiency RR database	Features of the predictor Short-term measures for HRV + transformation measures for wavelet	SVM: 82.75%t sensitivity Specificities: 96.29% Precision: 91.56% KNN $k = 5$ 65.51% sensitivity Specificities: 96.29% Precision: 85.54% LDA polynomial 75.86% Sensitivity Specificity: 90.74%
			Feature to respond Normal, CHF	Validation: cross-validation leave-one-out
Masetic et al. [15]	Random forests algorithm	Congestive heart failure and PTB diagnostic ECG databases, either free on PhysioNet [21] Beth Israel Deaconess Medical Center (BIDMC)	Feature normal CHF response feature	The sensitivity, specificity, precise, F-measures, and ROC curves have been recorded and compared with the classification systems (C4,5, SVM, artificial neural networks (ANNs), and k-nearest neighbors (KNN). Due to its high precision, the authors chose random forests to classify a topic as normal or CHF
Aljaaf et al. [2]	Multi-level HF risk assessment by using the tree classification C4.5	Heart disease dataset of the Clinic Foundation	Feature CHF normal response	The proposed approach is 86.30% accuracy and 89.00, 86.50, 72.00, 90.90, and 100.00% predictor accuracy of every risk level Validation Cross-validation 10 times

Table 2 Detailed accuracy of Naïve Bayes

FP rate	Precision	Recall	F-measures	ROC area	Class
0.192	0.853	0.887	0.869	0.916	Absent
0.113	0.851	0.808	0.829	0.916	Present
0.157	0.852	0.852	0.851	0.916	Weighted avg.

Table 3 Confusion matrix of Naïve Bayes

a	*b*	Classified
133	17	*a* <- Absent
23	97	*b* <- Present

is, however, a lack of accuracy since it is founded on expectation and class independence. Once the WEKA tool has been trained and tested, the precision of this model with 10 times cross-validation has been shown to be 85.18%. The detailed accuracy matrix as shown in Table 2 and the uncertainty matrix in Table 3.

3.2 REPTree Classifier

A fast-decision-making tree learner is an algorithm. The algorithm C4.5 can also be used to generate classification (discrete outcome) and regression trees (continuous outcome). It creates a knowledge gain/variance regress/decision tree and lowers it with decreased error cutter (with back-fitting). After practicing and testing this model on a WEKA tool, 83.33% with tenfold cross-validation have been shown to be exact. The detailed accuracy matrix shown in Table 4 and the uncertainty matrix in Table 5. The Visualize classifier error of REPTree as shown in Fig. 1.

Table 4 Detailed accuracy of REPTree

FP rate	Precision	Recall	F-measures	ROC area	Class
0.325	0.787	0.960	0.865	0.878	Absent
0.040	0.931	0.675	0.783	0.878	Present
0.198	0.851	0.833	0.828	0.878	Weighted avg.

Table 5 Confusion matrix of REPTree

a	*b*	Classified
144	6	*a* <- Absent
39	81	*b* <- Present

Fig. 1 Visualize classifier error of REPTree

3.3 AdaBoost

Freund developed the boosting system for the first time [17]. The initial classifier is built from the original dataset in which each sample has a distribution rate equivalent to 1. The distribution ratios of the samples are different based on their estimation accuracy in the previous Di-1 dataset with the boosting process training dataset Di. If a sample has a lower Di-1 predictor accuracy score, a higher Di weight is given, and hence, a higher selection in Di is possible as described in Fig. 2. After the execution of AdaBoost model on WEKA tool. It has been observed that the accuracy is 85.92% achieved with tenfold cross-validation. The detailed accuracy is shown in Table 6, and confusion matrix is shown in Table 7. It has been also observed the visualize classify errors and visualize tree as shown in Fig. 4, and ensemble classifier is shown in Fig. 3.

3.4 J48 Classifier

In WEKA tool built from Java, c4.5 representation is J48. Decision tree uses tree concepts to resolve the issue. Leaf node is the class mark, while the internal node of tree is specified. The selection process is performed by information gain and gain index in the decision tree attribute. The classification is based on the idea of the gain of knowledge and the value of the gain of information. For a particular X attribute of a node, the knowledge gain is measured as

$$\text{Information Gain } (N, X) = \text{Entropy } (N) - \sum_{\text{Value at } X} \frac{|N|}{|N_i|} \text{Entropy}(N) \qquad (2)$$

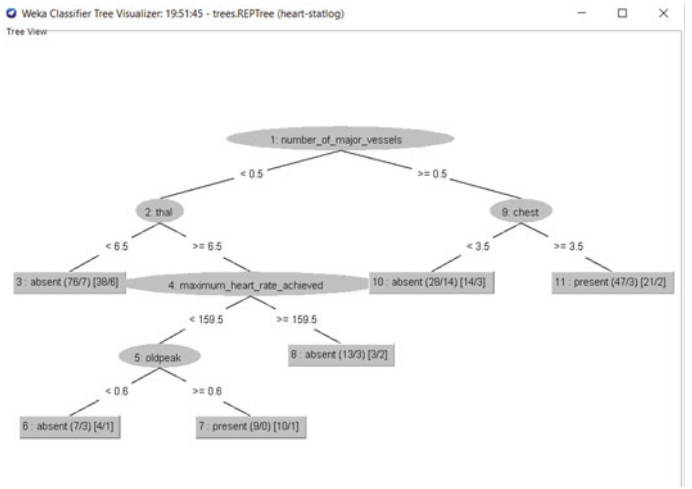


Fig. 2 Visualize tree of REPTree

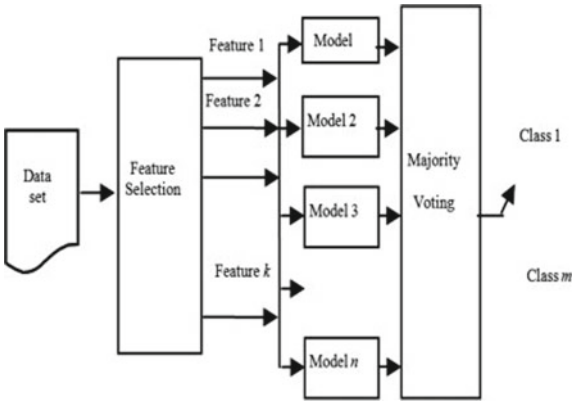
Table 6 Detailed accuracy of AdaBoost

FP rate	Precision	Recall	F-measures	ROC area	Class
0.167	0.868	0.880	0.874	0.933	Absent
0.120	0.847	0.833	0.840	0.933	Present
0.146	0.859	0.859	0.859	0.933	Weighted avg.

Table 7 Confusion matrix of AdaBoost

a	*b*	Classified
132	18	*a* <- Absent
20	100	*b* <- Present

Fig. 3 Ensemble classifier



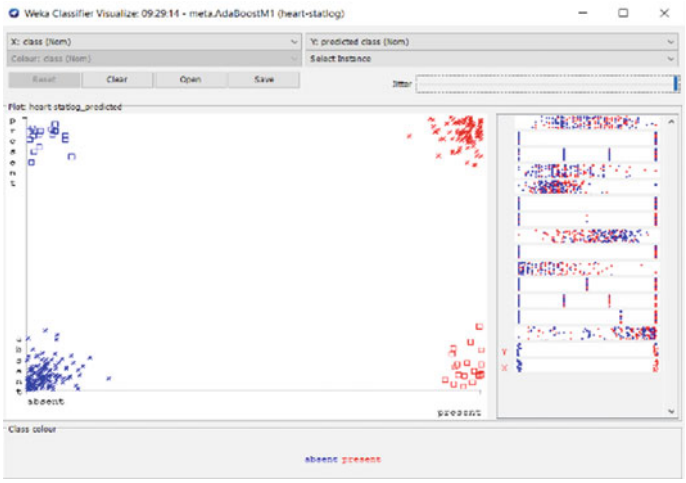


Fig. 4 Visualize error in AdaBoost

where N is the set of instances at that node, $|N|$: *cardinality*. Entropy of N is found as

$$\text{Entropy } (N) = \sum_{i=1}^{N} -p_i \log_2 p_i \tag{3}$$

After the execution of J48 classifier on WEKA tool along with dataset, it has been observed that accuracy is 91.48% achieved with tenfold cross-validation. The detailed accuracy is shown in Table 8, and Confusion matrix is shown in Table 9. The visualize classify errors and visualize tree are shown in Figs. 5 and 6.

Table 8 Detailed accuracy of J48

FP rate	Precision	Recall	F-measures	ROC area	Class
0.117	0.910	0.940	0.925	0.942	Absent
0.060	0.922	0.883	0.902	0.942	Present
0.91	0.915	0.915	0.915	0.942	Weighted avg.

Table 9 Confusion matrix of J48

a	*b*	Classified
141	9	*a* <- Absent
14	106	*b* <- Present

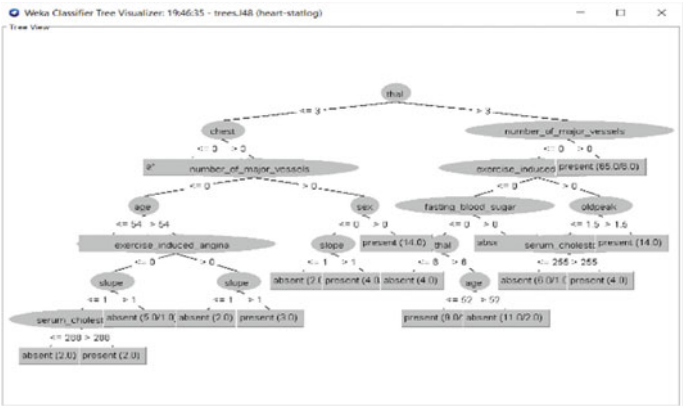


Fig. 5 Visualize tree of J48

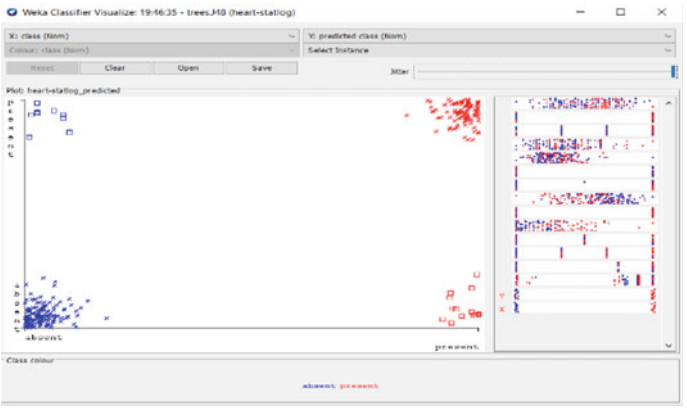


Fig. 6 Visualize classifier error of J48

3.5 Logistic Regression (LR)

For the development of the logistic regression classifier, statistics are used. The results of the input process data are dependent on the likelihood. In machine learning methods for handling binary input parameters, binary logistic regression is commonly favored. The sigmoid function is used to classify the class into a particular group. Logistic classification advantages are as follows.

- The logistic regression classification is highly adaptable
- Probabilistic model suitable for binary classification.

After the trained and test the model along with 10 cross-validations on WEKA tool. It has been observed that the accuracy of this model is 85.55% achieved with

Table 10 Detailed accuracy of logistic regression

FP rate	Precision	Recall	F-measures	ROC area	Class
0.183	0.858	0.887	0.872	0.929	Absent
0.113	0.852	0.817	0.834	0.929	Present
0.152	0.855	0.856	0.855	0.929	Weighted avg.

Table 11 Confusion matrix of logistic regression

a	*b*	Classified
133	17	*a* <- Absent
22	98	*b* <- Present

Fig. 7 Visualize error of logistic regression

tenfold cross-validation. The detailed accuracy is shown in Table 10, and confusion matrix is shown in Table 11. It has been also observed the visualize classify errors in Fig. 7.

3.6 Support Vector Machine (SVM)

Supporting vector classification is one of the basic and valuable methods in supervised learning. Due to its numerical capabilities, supported vectors are chosen for data processing in a very short time. This classifier deals with the principle of judgment limits. It is known as an hyperplane. The hyperplane is used to identify the input data within the goal category. But the maximum distance is selected from classification data points to match the decision boundary in one plane. User specified

Table 12 Confusion matrix of SVM

a	*b*	Classified
132	18	*a* <- Absent
22	98	*b* <- Present

Table 13 Detailed accuracy of SVM

FP rate	Precision	Recall	F-measures	ROC area	Class
0.183	0.857	0.880	0.868	0.848	Absent
0.120	0.845	0.817	0.831	0.848	Present
0.155	0.852	0.852	0.852	0.848	Weighted avg.

vector classification help can be framed to increase precision across the numerous kernel functions. The vector classification support system is suitable for organized and unstructured data. The issue of overfitting does not impact the vector classification help, which makes it accurate. After the trained and test the model along with 10 cross-validations on WEKA tool. It has been observed that the accuracy of this model is 85.182% achieved with tenfold cross-validation. The detailed accuracy is shown in Table 13, and confusion matrix is shown in Table 12. It has been also observed the visualize classify errors in Fig. 8.

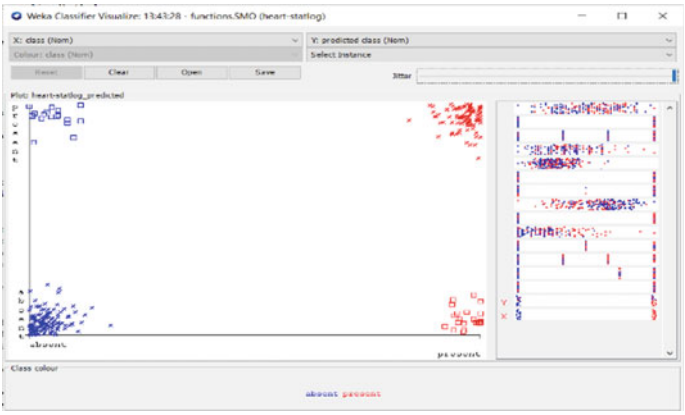


Fig. 8 Visualize error of SVM

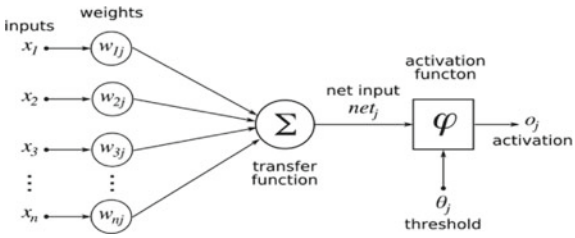


Fig. 9 Artificial neural network architecture

4 Proposed System

4.1 Artificial Neural Network (ANN)

An ANN is nonparametric method that allows it to be classified in the previous records based on input parameters to classify outcomes as fit or unfit. ANN is used to classify status of patients and prognosis based on uncertainty characteristics [8]. In addition, ANN is an artificial intelligence application [18, 19]. It is nonlinear simulations of intelligent computational techniques that have newly seen as advances registering and data handling tools that have taken a critical and progressed position in science and technologies. In normal, neural organizations have three sorts of layers as shown in Fig. 9:

- Input layer: input data
- Hidden layers: all computation
- related to model training is performed
- through intermediate layer.
- Output layer: outcome of given inputs.

4.2 Backpropagation Network

The core of net neural training is back propagation. It is the way to adjust net weights to the error rate in the earlier era (i.e., iteration). Proper weight adjustment allows the error rate to be reduced and the model to be more accurate. Backpropagation is a "backward spreading of mistakes" short form. It is a popular approach for the development of artificial neural networks. This approach helps measure the loss function gradient in terms to all network weights. The BPN model works the following: inputs X arrive through the preconnected path (Table 14).

- Entry is modeled with true weights W. Typically, the weights are chosen uniformly.
- Compute the output from the input layer to the hidden layer of each neuron to the output layer.
- The production errors are calculated.

Table 14 Attribute of heart failure

S. No.	Attribute name
1	age
2	sex
3	chest pain
4	resting_blood_pressure
5	serum_cholestoral
6	fasting_blood_sugar
7	resting_electrocardiographic_results
8	maximum_heart_rate_achieved
9	exercise_induced_angina
10	oldpeak
11	slope
12	number_of_major_vessels
13	thal

We load the dataset of heart failure patient as shown in Table 15 into BPN model implemented in WEKA tool as shown in Fig. 10 and select function like multilayer perceptron having 6 * 7 hidden layers trained and test the model along with tenfold cross-validation. We observed that the model provides 96.66% accuracy achieved. The detailed accuracy is shown in Table 17, and confusion matrix is shown in Table 16. The algorithm of BPN model is shown in Fig. 11.

The output variable shows whether an individual has heart failure (absent, present) as shown in Table 18.

4.3 Learning as an Optimization Problem

Firstly, the BPN model mathematical derivation of the relation between actual performance and accurate performance in a particular training example helps understand. A simple neural network with two input units, one output, and no hidden units is required. There is a linear output of every neuron, which is its weighted input number (unlike other neuronal mapping units that do not function on inputs).

At first, the weights are set arbitrarily before exercise. The neuron then learns from training examples consisting of a number of tuples, such as (attributes and target). Where attributes are the inputs to the network and t is the correct output. The initial network given attributes will compute an output y that likely differs from t (given random weights). For the measurement of the difference between the target result t and the computed output y, a loss function $L(t, y)$ is used. The squared error can be used as a loss function for regression analysis issues, and the categorical cross-entropy may be used for classification. For example, consider using the square error as a loss for a regression problem:

Table 15 Dataset of heart failure [21]

Sl. No.	Age	Gender	Chest	resting_ blood_pr esure	serum_c holestrol	fasting_ blood_su gar	resting_ cardiogr apbic_re sults	maximum_hea rt-beat_achieved	exercise_inclu de_engina	Old_pea k	slope	number of _major_ vessels	thal	Class
1	70	1	4	130	322	0	2	109	0	2.4	2	3	3	present
2	67	0	3	115	564	0	2	160	0	1.6	2	0	7	absent
3	57	1	2	124	261	0	0	141	0	0.3	1	0	7	present
4	64	1	4	128	263	0	0	105	1	0.2	2	1	7	absent
5	74	0	2	120	269	0	2	121	1	0.2	1	1	3	absent
6	65	1	4	120	177	0	0	140	0	0.4	1	0	7	absent
7	56	1	3	130	256	1	2	142	1	0.6	2	1	6	present
8	59	1	4	110	239	0	2	142	1	1.2	2	1	7	present
9	60	1	4	140	293	0	2	170	0	1.2	2	2	7	present
10	63	0	4	150	407	0	2	154	0	4	2	3	7	present
11	59	1	4	135	234	0	0	161	0	0.5	2	0	7	absent
12	53	1	4	142	226	0	2	111	1	0	1	0	7	absent
13	44	1	3	140	235	0	2	180	0	0	1	0	3	absent
14	61	1	1	134	234	0	0	145	0	2.6	2	2	3	present

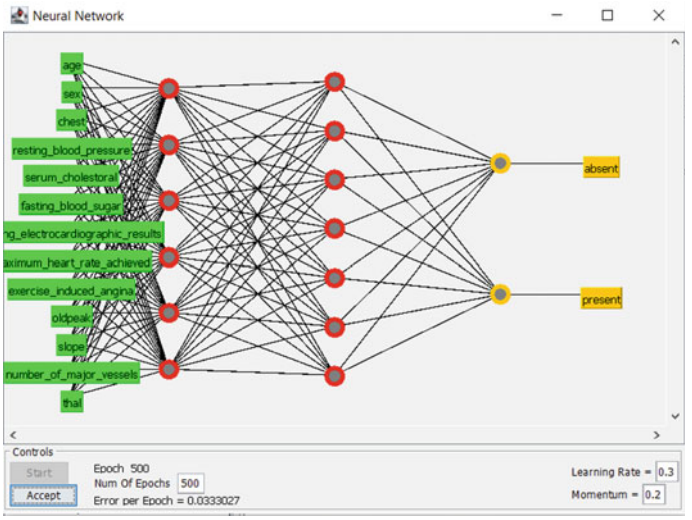


Fig. 10 Multilayer perceptron

Table 16 Detailed accuracy of BPN

FP rate	Precision	Recall	F-measures	ROC area	Class
0.42	0.967	0.973	0.970	0.961	Absent
0.027	0.966	0.958	0.962	0.961	Present
0.035	0.967	0.967	0.967	0.961	Weighted avg.

Table 17 Confusion matrix of BPN

a	*b*	Classified
146	4	*a* <- Absent
5	115	*b* <- Present

$$\text{Loss Function } (t, y) = (t - y)^2 = \text{ Error}(E) \tag{4}$$

where E is the discrepancy or error.

4.4 Gradient Descent

The procedure of gradient downturn includes the calculation of the weight derivative of the loss function. Typically, this is achieved with backpropagation. The squared error function is assumed to have one output neuron

```
Input:  Data set D, learning rate l, network      Output: Trained Neural Network
(1)  Initialize all weights and biases in network;
(2)  while terminating condition is not satisfied {
(3)      for each training tuple X in D {
(4)          // Propagate the inputs forward:
(5)          for each input layer unit j {
(6)              O_j = I_j; // output of an input unit is its actual input value
(7)          for each hidden or output layer unit j {
(8)              I_j = Σ_i w_ij O_i + θ_j; //compute the net input of unit j with respect to the
                     previous layer, i
(9)              O_j = 1/(1+e^{-I_j}); } // compute the output of each unit j
(10)         // Backpropagate the errors:
(11)         for each unit j in the output layer
(12)             Err_j = O_j(1 − O_j)(T_j − O_j); // compute the error
(13)         for each unit j in the hidden layers, from the last to the first hidden layer
(14)             Err_j = O_j(1 − O_j) Σ_k Err_k w_jk; // compute the error with respect to the
                     next higher layer, k
(15)         for each weight w_ij in network {
(16)             Δw_ij = (l) Err_j O_i; // weight increment
(17)             w_ij = w_ij + Δw_ij; } // weight update
(18)         for each bias θ_j in network {
(19)             Δθ_j = (l) Err_j; // bias increment
(20)             θ_j = θ_j + Δθ_j; } // bias update
(21)     } }
```

Fig. 11 Algorithm of backpropagation network

Table 18 Outcome

S. No.	Output variables	Heart failure prediction
1	Healthy "1"	Absent
2	Un-healthy "2"	Present

$$\text{Error } (E) = \text{Loss Fuction } (L)(t, y) \tag{5}$$

Here, E represents error or loss of NNM, 't' is target value, and 'y' is observed value.

$$O_j = \varphi(\text{net})_j = \varphi \sum_{k=1}^{n} w_{jk} O_j \tag{6}$$

Activation function is used in proposed NNM as follows

$$\varphi(u) = \frac{1}{1 + e^{-x}} \tag{7}$$

For each neuron, j and o_j are defined as

Where the activation function φ is nonlinear and differentiates as shown in Fig. 12 (regardless of whether the ReLU is not in one point).

Convenient derivative in NNM

$$\frac{\varphi(u)}{d(u)} = \varphi(u)(1 - \varphi(u)) \tag{8}$$

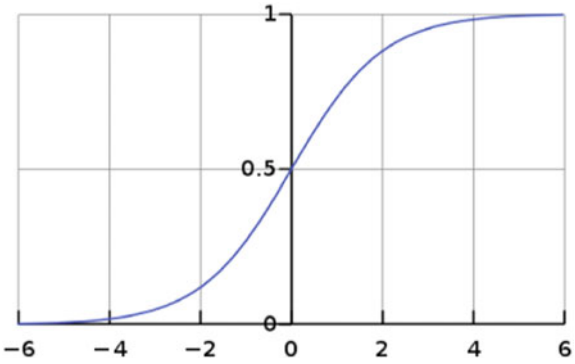


Fig. 12 Activation function

4.5 Derivative Error

In NNM, gradient descent is used to update weights of neurons w_{ij} with learning rate $\eta > 0$. The error E of NNM is back propagated by adjusting weights by improving or reducing neurons weights w_{ij} for adopting the model as per feed dataset. If $\frac{\partial E}{\partial w_{ij}} > 0$, as increases in w_{ij} increases E. Conversely, If $\frac{\partial E}{\partial w_{ij}} < 0$, an increase in w_{ij} decreases E. The new Δw_{ij} included to the previous weights followed by multiplication of learning rate by -1. It ensures that weight w_{ij} changes in a manner that consistently diminishes E. As such, in the equation immediately below, $\eta \frac{\partial E}{\partial w_{ij}}$ always changes w_{ij} in such a way that E is decreases.

$$\Delta w_{ij} = \eta \frac{\partial E}{\partial W_{ij}} = -\eta O_i \partial_j \tag{9}$$

4.6 Loss Function

The loss function guides estimate in any event one component onto an actual figure instinctively addressing few "costs" related with those characteristics. In backpropagation, NNM error is measured by loss function which represents differences between predicted and expected output values. NNM loss function's statement has accomplished binary state with the end goal for it to be potentially utilized in backpropagation network [20]. The first one is tending to be composed as an average $E = \frac{1}{n} \sum_x E_x$ over error function E_x for n training case, x. The explanations behind this supposition that will be that the BPN determines the error function's gradient for training dataset for each iteration. Final error is measured by summing up all BPN outcomes.

Let y, y' be a vector in R_n. Select an error function $E\left(y, y'\right)$ estimating the contrast for NNM outputs. The standard deviation between NNM outcome and expected

outcome is as follows:

$$E(y, y') = \frac{1}{2}(\|y - y'\|) \tag{10}$$

The error function over training models would then be able to be composed as an average of losses over case:

$$E(y, y') = \frac{1}{2n}(\|y - y'\|)^2 \tag{11}$$

5 Experimental Analysis

During the experimental analysis, we analyze all above models trained and tested in WEKA tool, and observed accuracy of Naïve Bayes classifier is 85.185%, accuracy of REPTree classifier is 83.33%, accuracy of AdaBoost classifier is 85.92, accuracy of J48 classifier is 91.48%, accuracy of logistic regression is 85.55%, accuracy of SVM classifier is 85.182%, and the proposed classifier, i.e., BPN is 96.66%. Accuracy prediction chart is shown in Fig. 13 along with accuracy table shown in Table 19. We also analyze the accuracy in terms of parameters like FP rate, precision, recall, F-measures, and ROC area as shown in Table 20 and compare all models as shown in Fig. 14. According to experimental investigation on the WEKA tool, the BPN model is the best classification model with the highest accuracy when compared to all other models.

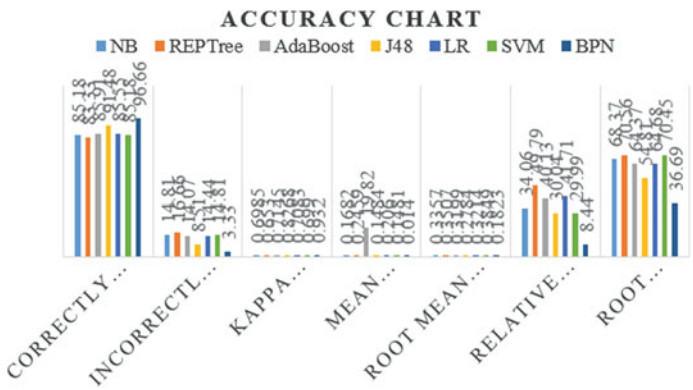


Fig. 13 Accuracy chart

Table 19 Comparison of models in terms of correctly instances and incorrectly instances

Classifier	Correctly classified instances	Incorrectly classified instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error
NB	85.18	14.81	0.6985	0.1682	0.3357	34.06	68.37
REPTree	83.33	16.66	0.653	0.2459	0.3507	49.79	70.56
AdaBoost	85.91	14.07	0.7145	19.82	0.3199	40.13	64.37
J48	91.48	8.51	0.8268	0.1484	0.2784	30.04	54.81
LR	85.55	14.44	0.7083	0.206	0.3214	41.71	64.68
SVM	85.18	14.81	0.699	0.1481	0.3849	29.99	70.45
BPN	96.66	3.33	0.932	0.014	0.1823	8.44	36.69

Table 20 Comparison of models in terms FP rate, precision, recall, F-measures, and ROC area

Classifier	FP rate	Precision	Recall	F-measures	ROC area	Class
NB	0.192	0.853	0.887	0.869	0.916	Absent
	0.113	0.851	0.808	0.829	0.916	Present
	0.157	0.852	0.852	0.851	0.916	Weighted avg.
REPTree	0.325	0.787	0.96	0.865	0.878	Absent
	0.04	0.931	0.675	0.783	0.878	Present
	0.198	0.851	0.833	0.828	0.878	Weighted avg.
AdaBoost	0.167	0.868	0.88	0.874	0.933	Absent
	0.12	0.847	0.833	0.84	0.933	Present
	0.146	0.859	0.859	0.859	0.933	Weighted avg.
J48	0.117	0.91	0.94	0.925	0.942	Absent
	0.06	0.922	0.883	0.902	0.942	Present
	0.91	0.915	0.915	0.915	0.942	Weighted avg.
LR	0.183	0.858	0.887	0.872	0.929	Absent
	0.113	0.852	0.817	0.834	0.929	Present
	0.152	0.855	0.856	0.855	0.929	Weighted avg.
SVM	0.183	0.857	0.88	0.868	0.848	Absent
	0.12	0.845	0.817	0.831	0.848	Present
	0.155	0.852	0.852	0.852	0.848	Weighted avg.
BPN	0.42	0.967	0.973	0.97	0.961	Absent
	0.027	0.966	0.958	0.962	0.961	Present
	0.035	0.967	0.967	0.967	0.961	Weighted avg.

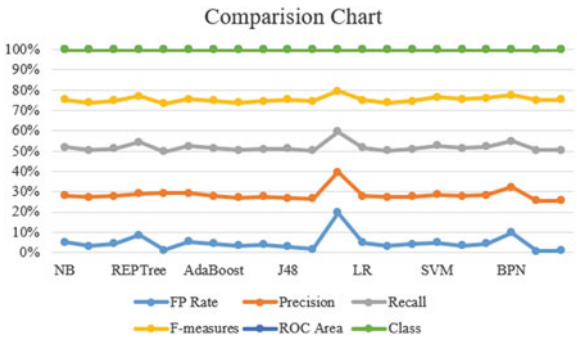


Fig. 14 Comparison chart of machine learning models

$$\text{Accuracy} = \frac{j\text{TP} + j\text{TN}}{j\text{TP} + J\text{TN} + j\text{FP} + j\text{FN}} \tag{12}$$

$$\text{Recall} = \frac{j\text{TP}}{j\text{TP} + j\text{FN}} \tag{13}$$

$$\text{Precision} = \frac{j\text{TP}}{j\text{TP} + j\text{FP}} \tag{14}$$

$$F\text{measure} = \frac{2 * \text{recall} * \text{precision}}{\text{recall} + \text{precision}} \tag{15}$$

6 Conclusion

Heart failure is one of the crucial diseases, and cure is very important part of patient life. Many studies are going on to cure human life for heart failure in different methods. In computer science, it plays a very important role by using machine learning technique to predict the accuracy and performance of heart patient life by applying heart patient datasets. In this paper, we took dataset from GitHub Websites and using ANN models, by applying BPN model to design and implement on WEKA tool to solve complex medical procedures. After implementation and execution of BPN model, we observed that number of epochs is 500, learning parameter is 0.3, and error per epochs is 0.33, and it has been seeing this model to predict the best accuracy as compared to all above models. The BPN is more effective in different medical fields (including prediction, diagnosis and treatment to helping specialists, and the general population).

References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS et al (2015) ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J, 2015 (ehw128) (2016). https://doi.org/10.1093/eurheartj/ehw128
2. Aljaaf AJ, Al-Jumeily D, Hussain AJ, Dawson T, Fergus P, Al-Jumaily M (2015) Predicting the likelihood of heart failure with a multi-level risk assessment using decision tree. In: Third international conference on technological advances in electrical, Beirut, Lebanon
3. Cowie MR (2012) The heart failure epidemic. Medicographia
4. Son C-S, Kim Y-N, Kim H-S, Park H-S, Kim M-S (2012) Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. J Biomed Inform 45:999–1008. https://doi.org/10.1016/j.jbi.2012.04.013
5. Roger VL (2010) The heart failure epidemic. Int J Environ Res Public Health 7:1807–1830. https://doi.org/10.3390/ijerph7041807
6. Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJV, Ponikowski P, Poole-Wilson, et al (2008) ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008 the task force for the diagnosis and treatment of acute and chronic heart failure 2008 of the European Society of Cardiology. Developed in collaboration with the heart failure association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). Eur Heart J 29:2388–2442. https://doi.org/10.1093/eurheartj/ehn309
7. Asyali MH (2003) Discrimination power of long-term heart rate variability measures
8. Işler Y, Kuntalp M (2007) Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. Comput Biol Med 37:1502–1510 (2007)
9. Elfadil N, Ibrahim I (2011) Self organizing neural network approach for identification of patients with congestive heart failure
10. Melillo P, Fusco R, Sansone M, Bracale M, Pecchia L (2011) Discrimination power of long-term heart rate variability measures for chronic heart failure detection. Med Biol Eng Comput 49:67–74. https://doi.org/10.1007/s11517-010-0728-5
11. Jovic A, Bogunovic N (2011) Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. Artif Intell Med 51:175–186. https://doi.org/10.1016/j.artmed.2010.09.005
12. Yu S-N, Lee M-Y (2012) Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. Comput Biol Med 42:816–825. https://doi.org/10.1016/j.compbiomed.2012.06.005
13. Liu G, Wang L, Wang Q, Zhou G, Wang Y, Jiang Q (2014) A new approach to detect congestive heart failure using short-term heart rate variability measures. PLoS One 9:e93399. https://doi.org/10.1371/journal.pone.0093399
14. Narin A, Isler Y, Ozer M (2014) Investigating the performance improvement of HRV indices in CHF using feature selection methods based on backward elimination and statistical significance. Comput Biol Med 45:72–79. https://doi.org/10.1016/j.compbiomed.2013.11.016
15. Masetic Z, Subasi A (2016) Congestive heart failure detection using random forest classifier. Comput Methods Programs Biomed 130:54–64. https://doi.org/10.1016/j.cmpb.2016.03.020
16. Fatima M, Pasha M (2017) Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl 9:1–16. https://doi.org/10.4236/jilsa.2017.91001
17. Breiman L (1999) Random forests—random features, Technical Report 567. University of California, Berkley
18. Abu Naser SS (2012) Predicting learners performance using artificial neural networks in linear programming intelligent tutoring system. Int J Artif Intell Appl 3(2):65–80

19. AbuEl-Reesh JY, Abu Naser SS (2017) A knowledge based system for diagnosing shortness of breath in infants and children. Int J Eng Inf Syst (IJEAIS) 1(4):102–115
20. Nielsen (2015) What assumptions do we need to make about our cost function ... in order that backpropagation can be applied? The first assumption we need is that the cost function can be written as an average ... over cost functions ... for individual training case ... The second assumption we make about the cost is that it can be written as a function of the outputs from the neural network
21. https://github.com/renatopp/arff-datasets/blob/master/classification/heart.statlog.arff

Analysis and Implementation of Microsoft Azure Machine Learning Studio Services with Respect to Machine Learning Algorithms

Soni Singh, K. R. Ramkumar, and Ashima Kukkar

Abstract In this paper we implemented Microsoft Azure machine learning studio using machine learning (ML) algorithm. Azure ML describes a graphical tool used to manage end to end operation of machine learning algorithms. With the help of Azure Tools, users of the machine learning group can conduct their own experiments to pre-process components on a given set of data with the help of different machine learning algorithms, after the experiment users check the accuracy of the model. Here, in this research we conducted two experiments on CVS using Bayesian algorithm and K-means clustering algorithm data set. After experimenting, we visualized the result using different aspects of the web-based application model. When we build an impressive model, it also helps our users to establish that model in Microsoft Azure. Azure ML uses various ML algorithms to predict models operating on preprocessing data sets. Using drag and drop we connect all the components together and run an experiment. Once we find an accurate result it is deployed in Microsoft Azure Studio for other people to use.

Keywords Azure ML · Azure ML services · Implementation of Azure ML in Bayesian linear regression algorithm and K-means clustering algorithm

1 Introduction

Machine learning is a method that implies predicting an outcome based on a given set of data. It uses various machine learning techniques such as supervised learning,

S. Singh · K. R. Ramkumar (✉) · A. Kukkar (✉)
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
e-mail: k.ramkumar@chitkara.edu.in

A. Kukkar
e-mail: ashima@chitkara.edu.in

S. Singh
e-mail: soni.singh@chitkara.edu.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_7

unsupervised learning and reinforcement learning techniques to solve complex problems. Machine learning algorithms are not very easy to implement. ML Studio provides various services and components in a simplified manner which can be used very well by machine learning [1]. Microsoft Azure is a cloud-based platform that helps people performing machine learning operations. Microsoft Azure runs on a cloud-based platform. With the help of Microsoft Azure, we can work with huge amounts of data and can be accessed from all over the world. Use of Azure only requires a web browser and an Internet connection. Azure ML provides graphical tools to control the ML process, it is a set of data preprocessing components, a set of various machine learning algorithms, and an API to deploy a model to applications [2]. Tools used in ML to manage ML operations from start to finish. Using these tools, those using the ML method can apply a data preprocessing component to raw data, run experiments on the available data with the help of machine learning algorithms, and then evaluate the resulting model. When we find a better model, Azure ML Studio also deploys that model to Microsoft Azure [2, 3]. ML Studio gives a simple tool for managing all machine learning operations. One of the main objectives of Azure ML is to make it easy for people to use all machine learning algorithms with the help of this tool [3]. Azure ML offers a lot of machine learning, and people using machine learning can easily build their own models using Azure machine learning tools.

1.1 Concept of Azure Machine Learning

Azure ML studio is an online graphical tool used to solve complex ML problems. For using this tool, for using Azure tool it needs only a browser and an Internet connection [4]. By using this tool people using ML team can apply on preprocessing component to raw data, run experiment on the prepare data with the help of ML algorithm and evaluate the result of that model [5]. Once we find an accurate and correct model then ML studio also used for its users to establish that model in Microsoft Azure [4, 5] (Fig. 1).

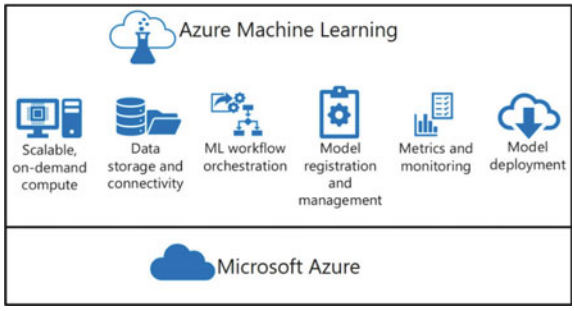


Fig. 1 Azure machine learning

The best part of Azure ML is that you can train a model over a local machine and then deploy it on cloud [6]. Azure gives many computing services like Azure ML compute and Azure Databricks to make your model more accurate and best [7].

2 Literature Survey

Das [2] In this research the author proposed a prediction model for companies selling products based on customer potential and their buying habits. They analyze existing customers and their purchase orders for predictions. For the experiment they use the naive Bayes classifier, K-nearest neighbor (KNN) and support vector machines for customer-based purchase orders. They compare the accuracy and efficiency of models based on their prediction rates for the available datasets. Salmi et al. [3] In this research the author work on the classification method using the naïve Bayes classifier for predicting the cancer disease. They calculated the accuracy of the model and also measure the precision score and recall factor. They also present the F1-score of the model using the naïve Bayes classifier for measuring the patients whoever is suffering from cancer. The proposed model accuracy is 95.24% with high assumptions and less complexity. The model is high classification accuracy model with accurate prediction. Osisanwo et al. [4] In this paper the author study about various supervised machine learning algorithm for comparing the other supervised machine learning algorithms for classification-based problem. In this research seven machine learning algorithm are used for prediction. For implementing these algorithms diabetic data are used. The data set consist of 786 instances and 8 attributes for analysis. The final result shows that support vector machine learning classifier is providing the accurate result comparing with other machine learning algorithms. The accuracy also calculated using the Mean Absolute Error (MAE) parameter. Barga et al. [8] In this research the author defines Microsoft Azure ML in relation to data scientists and machine learning developers. Developing a prediction model with the help of Azure ML. It describes the use of Microsoft azure services in relation to ML. This makes machine learning algorithms easy to implement and provides many solutions for society. Machine learning studio also helps to perform multiple tasks easily and make predictions based on the data set. Julian et al. [5] In this research the author study about benefits, component and uses of Microsoft azure ML studio. An Azure ML SDK tool is used to build a model to provide end to end solutions by developing a prediction model.

3 Services of Azure Machine Learning

Azure ML has an ability to self-train its model and it provide tool for python and other open-source packages that allows its user to create and train model with machine learning in Azure ML service workspace. The Python provide many packages like PyTorch, MXNet, Scikit-learn, etc. [4].

Once you create the model with the help of these tools then you need to create a Docker that should be tested locally [8]. After creating a successful model, you can deploy that model as web service using Azure Container service, then the Azure portal will guide you to manage that deployed web model. Azure ML will also help you to measure the performance of model, instant tracking and restructuring of revised versions [5, 6]. Azure ML provide all the necessary tools to machine learning developers and data scientist for the smooth operation of workflow as well as:

- **Azure ML designer**: It provides drag and drop module to create your experiment and then connect these modules together with pipelines. Azure design to creation of data, train, track, test and measure your machine learning model without writing any code. It does not need any program to write you just simply connect all the data sets and module together to build a model [8].
- **Jupyter notebook**: With the help of jupyter notebook you can create your own notebook for your machine learning with the help of our SDK for Python samples [9].
- **R scripts**: R script is like a notebook where you can write your own code with the help of SDK for R, you can also use R module in your design.
- **Open-source frameworks**: Azure ML provides many open-source frameworks like MXNet, Scikit-learn, TensorFlow and PyTorch, etc. [9].
- **ML flow or Kubeflow**: ML flow is used to measure, track and establishment of models or Kubeflow is used to create end to end workflow pipelines.
- **Secure Communication**: Azure ML storage account provides compute targets, and many other services that can be used securely into a virtual network to build and train models [8, 9].
- **Other Azure ML services**: Azure ML provide many other services on Azure platform like Azure Kubernetes services, Azure Databricks, Azure HDInsight, Azure Container and Azure data Lake Analytics. It also works with open-source tools like Git and MLflow [9].

4 Components of Azure Machine Learning

Azure ML describes a graphical tool which is used to manage the operation of machine learning algorithm from starting to end. With the help of azure tool, the users of machine learning group can perform their experiment on pre-process component to given set of data, with the help of different machine learning algorithm [6, 10]. After performing the experiment, the user checks the accuracy of model. When we create an impressive model, then it also helps its users to establish that model in Microsoft Azure [5].

Azure ML is a tool which is used to make ML algorithm easier for implementation. It provides several components which shown in Fig. 2.

Azure machine learning gives different components they are as follows:

- It is a collection of preprocessing set of components.

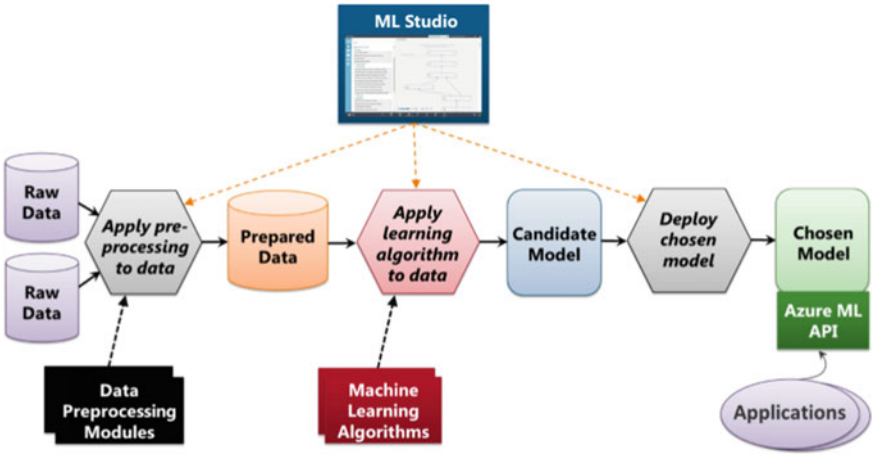


Fig. 2 Components of Azure machine learning

- It shows a collection of different ML algorithms.
- It also helps in API to select model used for an application when its established in Azure ML.
- Azure ML experiment performed on given preprocessed data and then evaluate the correct model based on the result of the experiment [6].

Azure ML is a graphical representation tool used to handle multiple machine learning operations. It works on a collection of data preprocessing components, a collection of different machine learning algorithm, and various API to dispose a design to use for an application [7].

5 Microsoft Azure ML Studio

Microsoft Azure ML studio is a graphical tool used for performing various machine learning operations with the help of given pre-processes data to establish a model [11].

Figure 3 describes the ML studio design and process in which it allows its user to simply drag and drop datasets, after drag and drop the user can connect these modules together graphically. Once it connects all the modules together then run experiment to compute the results.

Example Let us consider that a data scientist can use Microsoft Azure ML to connect a data set with other pre-process data with the help of ML algorithm that he selected for the implementation of his experiment. When he completes his experiment, he evaluates the result of that experiment [10]. Once he fully satisfied with that selected model then he establishes that model in Microsoft Azure, where other users can

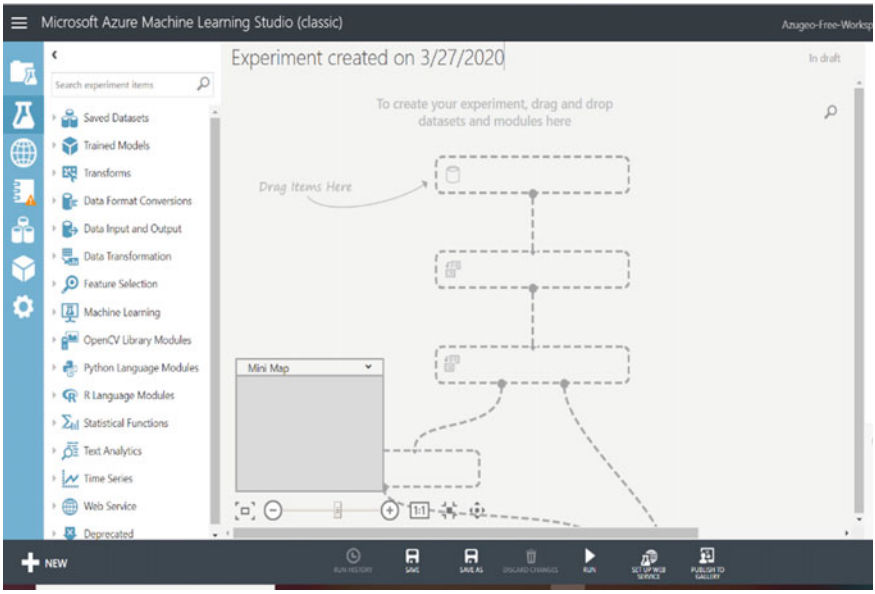


Fig. 3 Microsoft Azure ML studio

access that model. Azure ML studio gives a graphical tool which helps to manage the complete process of machine learning.

6 Case Study: Implementation of Bayesian Algorithm In-Flight Delay Data Set Using Microsoft Azure Machine Learning

Naïve Bayes classification algorithm is very simple, fast and reliable to use. It is a probabilistic machine learning algorithm that can be used for large diversity of classification problems. The naïve Bayes classification is based on Bayes' Theorem for assumption on that one specific feature in a model is independent of each other. That is change in value of one feature does not affect or change the value of any other feature used in algorithm [12]. This is totally based on probabilistic model, so through this algorithm the prediction made really easy and fast.

Azure ML uses different ML algorithm for the prediction of model it works on preprocessing data set. By using drag and drop we connect all the component together and run an experiment. Once we find an accurate result then it deployed in Microsoft azure studio for other people to use it [13].

Step by Step Experiment of Flight delay data in Microsoft Azure machine learning using Bayesian Linear Regression algorithm:

- **Drag Flight Delay Data from given Sample data**: First we drag flight delay data set from the given sample data for perform experiment.
- **Drag Select column for selecting columns**: After selecting data we drag the select columns option from the given option for selecting column from the chosen data set for performing operation [11].
- **Drag Split data for training**: Then we drag the split data option from the given option to split our data in two part, one for training and other for score with the help of Bayesian Linear Regression algorithm [10, 11].
- **Drag Bayesian Linear Regression algorithm**: After splitting the data we drag Bayesian Linear Regression algorithm for performing experiment.
- **Drag Train model for data training**: In this we train our model with the help of Bayesian Linear Regression algorithm [7, 10].
- **Drag Score Model for selected data**: After training our model we select the score model for any given data set option to perform experiment.
- **Evaluate Model**: In this we select evaluate model option to evaluate our model performance [14].

Table 1 describes the flight delay data in csv format that used for experiment using machine learning algorithm.

Figure 4 describes flight delay experiment by using Bayesian Linear Regression algorithm in Microsoft Azure ML studio. In which we simply drag and drop all the modules and connect all the selected module together then run experiment. After run an experiment we calculate the result [15].

In Fig. 5 describes the visualized result of flight delay in selected Score Model in year and month. It shows graph and its values for the selected column.

Figure 6 describes the Evaluation model and their Root Mean Square Error and other in-flight delay data.

Table 2 shows the evaluation of model using the Bayesian algorithms. It shows the mean absolute error, root mean square error and coefficient of determination. The table shows that mean square error is very less and it has accurate prediction using the flight delay data set.

7 Case Study: Implementation of K-Means Clustering Algorithm in Restaurant Rating Data Set Using Microsoft Azure Machine Learning

K-Means Clustering algorithm is one of the most usable algorithms in machine learning to classify data based on their feature, with the help of Microsoft Azure K-Means Clustering is easy to implement [15]. In this experiment we have taken a Restaurant rating data for which we run the experiment and analyze the result [16, 17].

Step by Step Experiment of Restaurant rating data in Microsoft Azure machine learning using K-Means Clustering algorithm:

Table 1 Flight delay data set

Year	Month	Day of month	Day of week	Carrier	Origin airport ID	Dest airport ID	CRS dep time	Dep delay	Dep Del15	CRS Arr time	Arr delay	Arr Del15	Canceled
2013	4	19	5	DL	11433	13303	837	−3	0	1138	1	0	0
2013	4	19	5	DL	14869	12478	1705	0	0	2336	−8	0	0
2013	4	19	5	DL	14057	14869	600	−4	0	851	−15	0	0
2013	4	19	5	DL	15016	11433	1630	28	1	1903	24	1	0
2013	4	19	5	DL	11193	12892	1615	−6	0	1805	−11	0	0
2013	4	19	5	DL	10,397	15016	1726	−1	0	1818	-19	0	0
2013	4	19	5	DL	15016	10397	1900	0	0	2133	−1	0	0
2013	4	19	5	DL	10397	14869	2145	15	1	2356	24	1	0
2013	4	19	5	DL	10397	10423	2157	33	1	2333	34	1	0
2013	4	19	5	DL	11278	10397	1900	323	1	2055	322	1	0
2013	4	19	5	DL	14107	13487	1540	−7	0	2043	−13	0	0
2013	4	19	5	DL	11433	11298	835	22	1	1035	41	1	0
2013	4	19	5	DL	11298	11433	1115	40	1	1450	20	1	0
2013	4	19	5	DL	11433	12892	1935	−2	0	2140	−7	0	0
2013	4	19	5	DL	10397	12451	1625	71	1	1738	75	1	0
2013	4	19	5	DL	12451	10397	1830	75	1	1955	57	1	0
2013	4	19	5	DL	12953	10397	1000	−1	0	1234	10	0	0
2013	4	19	5	DL	11433	12953	725	−3	0	918	−10	0	0
2013	4	19	5	DL	10397	14771	1725	31	1	1953	38	1	0
2013	4	19	5	DL	13204	10397	2030	8	0	2201	25	1	0

(continued)

Table 1 (continued)

Year	Month	Day of month	Day of week	Carrier	Origin airport ID	Dest airport ID	CRS dep time	Dep delay	Dep Del15	CRS Arr time	Arr delay	Arr Del15	Canceled
2013	4	19	5	DL	13487	12889	655	−3	0	827	−2	0	0
2013	4	19	5	DL	12889	13487	909	7	0	1415	16	1	0
2013	4	19	5	DL	10397	14027	1150	0	0	1337	19	1	0

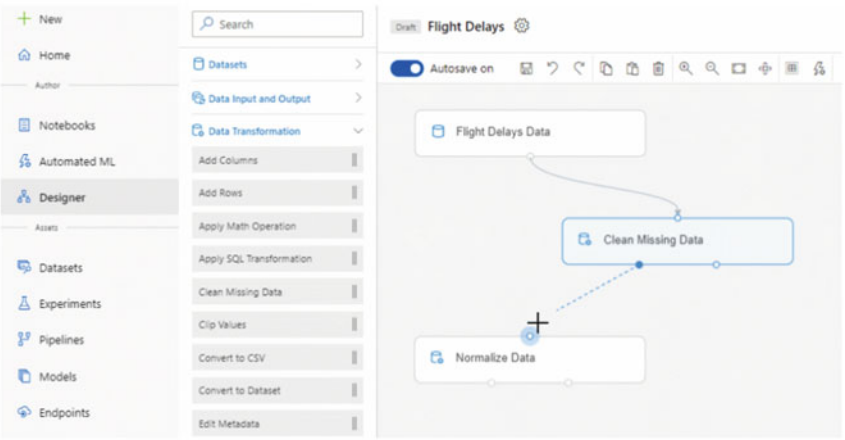


Fig. 4 Experiment of flight delay data using Azure ML

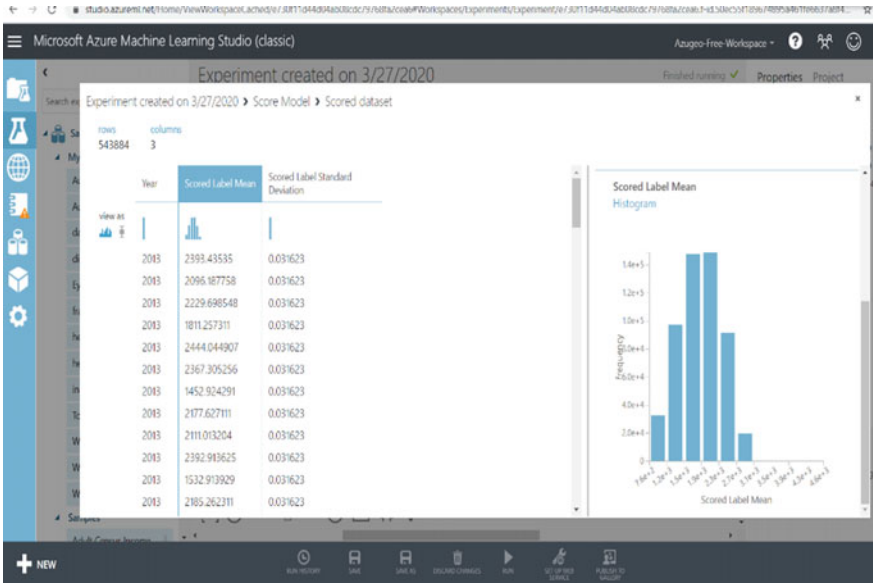


Fig. 5 Visualized the result of score model in-flight delay data using Azure ML

- **Drag Restaurant rating Data from given Sample data**: First we drag restaurant rating data set from the given sample data for perform experiment.
- **Drag Select column for selecting columns**: After selecting data we drag the select columns option from the given option for selecting column from the chosen data set for performing operation [18, 19].

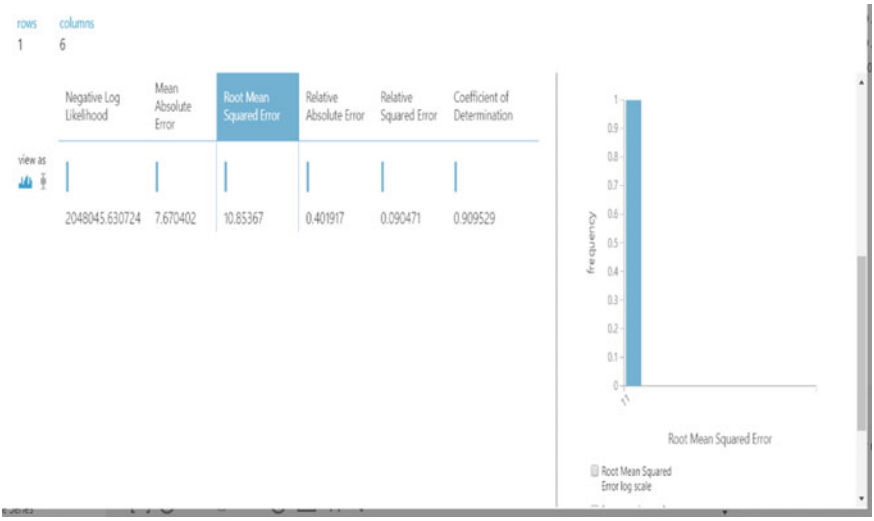


Fig. 6 Visualized the result of evaluate model in-flight delay data using Azure ML

Table 2 Flight delay data model evaluation

Model evaluation using the Bayesian algorithm		
Mean absolute error	Root mean square error	Coefficient of determination
7.670	10.856	0.909

- **Drag Split data for training**: Then we drag the split data option from the given option to split our data in two parts, one for training and other for score with the help of K-Means Clustering algorithm [20, 21].
- **Drag K-Means Clustering algorithm**: After splitting the data we drag K-Means Clustering algorithm for performing experiment.
- **Drag Train Clustering model for data training**: In this we train our model with the help of K-Means Clustering algorithm.
- **Drag Assign Data to Cluster for selected data**: After training our model we select the Assign Data to Cluster for any given data set option to perform experiment [20].
- **Evaluate Model**: In this we select evaluate model option to evaluate our model performance.

Table 3 describes the Restaurant Rating data set in csv format that used for experiment using machine learning algorithm.

Figure 7 describes Restaurant rating experiment by using K-Means Clustering algorithm in Microsoft Azure ML studio [22]. In which we simply drag and drop all the modules and connect all the selected module together then run experiment. After run an experiment we calculate the result [20].

Table 3 Restaurant rating data set

User ID	Place ID	Rating
U1077	135085	2
U1077	135038	2
U1077	132825	2
U1077	135060	1
U1068	135104	1
U1068	132740	0
U1068	132663	1
U1068	132732	0
U1068	132630	1
U1067	132584	2
U1067	132733	1
1U1067	132732	1
U1067	132630	1
U1067	135104	0
U1103	132560	1
U1103	132584	1
U1103	132732	0
U1103	132630	1
U1103	132613	2
U1103	132667	1
U1103	135104	1
U1103	132663	1
U11033	132733	2
U1107	132660	2
U1107	132584	2

Figure 8 describes the visualized result of created cluster of restaurant rating in selected Assign Data in Cluster with the help of selected data. It shows graph of two clusters which are Cluster 1 and Cluster 2.

Figure 9 describes the Evaluation model and their Average Distance to Cluster Center and other graphs in Restaurant Rating data [16].

Figure 10 describes the Evaluation model. This model defines the Result Description graph in a combined evaluation of all created Clusters with respect to their frequency and other graphs in Restaurant Rating data.

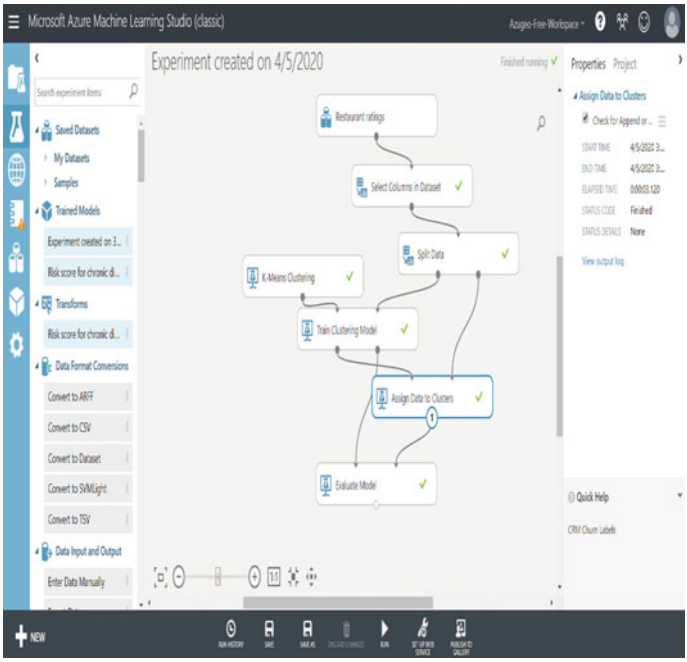


Fig. 7 Experiment of restaurant ratings using Azure ML

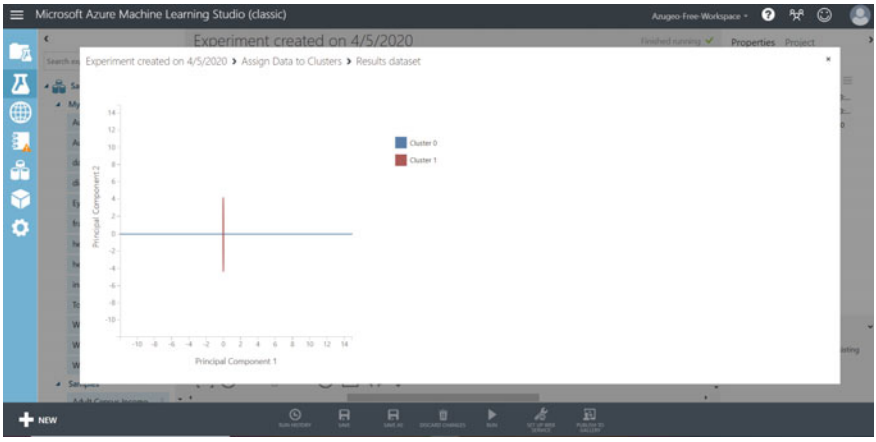


Fig. 8 Visualized the result of assign data in cluster of restaurant rating data using Azure ML

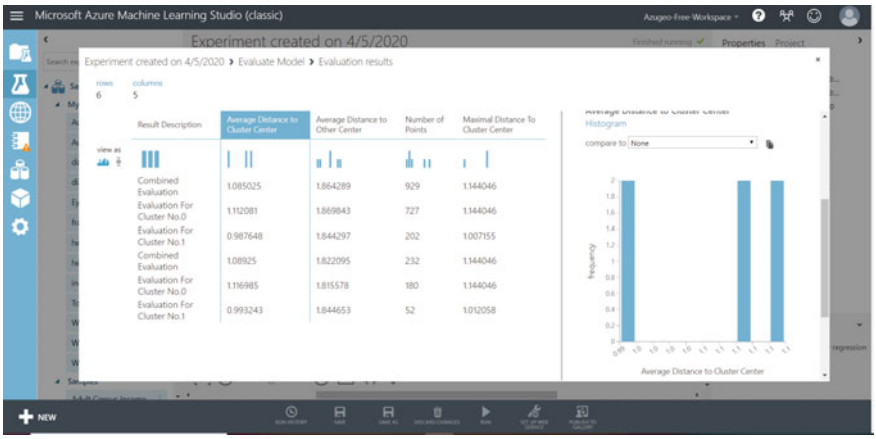


Fig. 9 Visualized the result of evaluate model in restaurant rating data using Azure ML

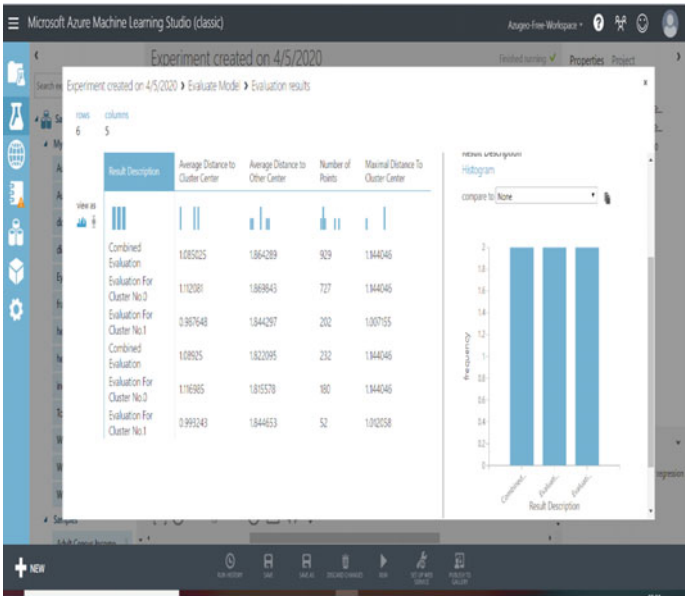


Fig. 10 Visualized the result of combined evaluation of all clusters in restaurant rating data using Azure ML

8 Advantages of Azure Machine Learning

- Azure ML tool makes easier to implement machine learning algorithm.
- Azure machine learning supports low code or code free implementation of different machine learning algorithm [23, 24].

- It is a graphical tool interface in which you can locally train your model then implement in Microsoft Azure ML studio which can be used by another ML team.
- It provides cloud-based operation in which you can easily access your data from anywhere [21, 25].
- With the help of Azure ML, you can train, track and measure performance of your model without writing any code [26].

9 Conclusion

The purpose of machine learning algorithms is to provide solutions to many problems. In today's time it is very useful everywhere. Currently we have huge amount of data for which we perform ML operations. Unless it is easy to use for the people performing ML algorithms, we cannot easily use our big data to provide solutions for that particular problem, and also not able to improve the quality of model that can be used in applications. The main goal of Azure machine learning is to make it easier for people to use machine learning algorithms. Data scientist in this era still needed this type of cloud-based tool to implement their operations, as there is little expertise in machine learning required for people using this Azure ML tool. Azure ML plays a big role in making machine learning very popular and easy to use for everyone who uses this machine learning algorithm. In future Microsoft Azure ML studio helps in easy implementation of all ML algorithms that helps in providing many solutions toward our society.

References

1. https://www.dgp.toronto.edu/~hertzman/411notes.pdf.
2. Das TK (2015) A customer classification prediction model based on machine learning techniques. In: International conference on applied and theoretical computing and communication technology (iCATccT), pp 321–326. IEEE
3. Salmi N, Rustam Z (2019) Naïve Bayes classifier models for predicting the colon cancer. IOP Conf Ser Mater Sci Eng 546:52–68
4. Osisanwo FY, Akinsola JET, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J (2017) Supervised machine learning algorithms: classification and comparison. Int J Comput Trends Technol (IJCTT) 48:128–138
5. Soh J, Singh P (2020) Introduction to Azure machine learning. In: Data science solutions on Azure. Apress, Berkeley, CA
6. https://data-flair.training/blogs/machine-learning-tutorial/
7. https://stackoverflow.com/questions/10059594/a-simple-explanation-of-naive-bayes-classification
8. Barga R, Fontama V, Tok WH (2015) Introducing Microsoft Azure machine learning. In: Predictive analytics with Microsoft Azure machine learning. Apress, Berkeley, CA
9. https://www.saedsayad.com/naive_bayesian.html
10. Machine learning in action by Peter Harrington

11. cs229.stanford.edu
12. Sternberg A, Soares J, Carvalho D, Ogasawara E (2017) A review on flight delay prediction. arXiv preprint
13. Choi S, Kim YJ, Briceno S, Mavris D (2016) Prediction of weather-induced airline delays based on machine learning algorithms. In: 35th digital avionics systems conference (DASC). IEEE, pp 1–6
14. https://docs.microsoft.com/en-us/azure/machine-learning/
15. Hossain FT, Hossain MI, Nawshin S (2017) Machine learning based class level prediction of restaurant reviews. In: Region 10 humanitarian technology conference (R10-HTC). IEEE, pp 420–423
16. Bzdok D, Krzywinski M, Altman N (2017) Machine learning: a primer. Nat Methods 14:1119
17. Chappell D (2015) Introducing Azure machine learning. A guide for technical professionals, sponsored by Microsoft Corporation
18. https://en.wikipedia.org/wiki/Unsupervised_learning
19. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, vol 7. Springer, New York, pp 485–586
20. https://www.outsource2india.com/software/articles/machine-learning-applications-how-it-works-who-uses-it.asp
21. Mund S (2015) Microsoft Azure machine learning. Packt Publishing Ltd. (2015)
22. Jordan MI, Bishop CM (2014) Neural networks. In: Tucker AB (ed) Computer science handbook, 2nd edn. (Section VII: Intelligent Systems). Chapman & Hall/CRC Press LLC, Boca Raton, FL
23. Bzdok D, Krzywinski M, Altman N (2018) Machine learning: supervised methods. Nat Methods 15:1–5
24. https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1
25. Hinton GE, Sejnowski TJ (eds.) (1999) Unsupervised learning: foundations of neural computation. MIT Press
26. Barnes J (2015) Microsoft Azure essentials Azure machine learning. Microsoft Press

Design and Analysis of Dual-Band Compact Antenna for 5G Communication Applications

Pankaj Jha, Seema Nayak, Samarth Golus, Prakriti Jaiswal, and Ayushi Kumari

Abstract A compact dual-band antenna is proposed in this paper for WLAN and 5G applications due to rising demand of high channel capacity and speedy transmission. The square loop radiator accomplished with multiple stubs with the partial ground is used to obtain dual operating bands. The stubs are responsible for impedance matching. The proposed antenna achieved 10 dB impedance bandwidth from 3.35 to 3.8 GHz and from 5.15 to 5.25 GHz. The maximum realized gain is 2.32 dB at 3.6 GHz frequency. The simulated radiation efficiency is more than 92% in the first operating band and more than 70% in the second operating band.

Keywords Deflective ground · Metamaterial inspired · Impedance matching · 5G communication · WLAN

1 Introduction

The materialize technology world demanding the 5G antenna with features previously unseen. Communication has been a challenge for wireless technology for its high-speed data transfer demand stagey and faster as well as a dynamic system. In recent few years, there is a rapid growth of wireless communication in various applications, and this evolution gives rise to the use of patch antenna. The antenna is the heart of the wireless communication system. Microstrip antenna is preferred for its ease of design, low profile, low cost in fabrication and compact structure. Operating frequency, antenna size, polarization, manufacturing cost and bandwidth are all factors to consider when designing a suitable antenna for 5G networks. Many researchers are working in the designing of dual-band or multiband antenna using different technologies [1]. In [2], metamaterial inspires CSRR structure is used to obtain multiband operation. In [3], multiband operation is achieved with the help of multiple slots in the radiator. In [4], inverted F-shaped radiator is used with the unsymmetrical ground to achieve multiband operation. In [5], unsymmetrical feed

P. Jha (✉) · S. Nayak · S. Golus · P. Jaiswal · A. Kumari
Department of Electronics and Communication Engineering, IIMT College of Engineering, Greater Noida, India
e-mail: pankaj.maahi@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_8

with rectangular slots in the radiator is used to design a wideband antenna for sub-6 GHz frequency. In [6], effective ground structure is used to shift the operating frequency towards 5G operating band (3.5 GHz). In [7], a rectangular patch with insists feed is used to design 5G antenna where unsymmetrical feed is analysed to optimize the operating band. In [8], CSRR-based metamaterial inspires metallic patch antenna is proposed where CSRR is multiband operation. In [9], circular slots in the circular radiator in compiles with CSRR in-ground are responsible for multi-band operation. In [10], a polygon shape radiator with CSRR is used to design a dual-band antenna where CSRR is demonstrating metamaterial characteristics.

The proposed dual-band antenna is compact and achieved |S11| from 3.35 to 3.8 GHz and from 5.15 to 5.25 GHz with adequate gain and efficiency which is suitable for 5G and WLAN application.

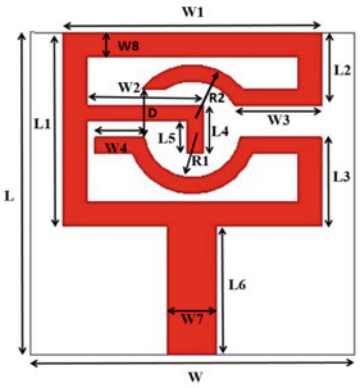
2 Antenna Design and Analysis

The projected 5G microstrip antenna is plan on FR4 substrate (thickness = 1.6 mm and $\varepsilon r = 4.4$,) depicted in Fig. 1a_r and 1b_r. The antenna prototype is illustrated in Fig. 1c_r. The pertaining dimensions of proposed antenna $W = 20$ mm and $L = 20$ mm. The feed is optimized to achieve 50 Ω impedance with dimension $L6 = 8$ mm and $W7 = 3$ mm. The other pertaining dimensions associated with proposed 5G antenna are $R1 = 3$ mm, $W1 = 16$ mm, $G_W = 20$, $W2 = 7.2$ mm, $L3 = 5.5$ mm, $W3 = 5.4$ mm, $L4 = 3$ mm, $W4 = 3$ mm, $L1 = 16$ mm, $R2 = 4$ mm, $L2 = 4.5$ mm, $L5 = 2$ mm, $W8 = 2$ mm, $G_L = 6$ mm and $D = 3$ mm.

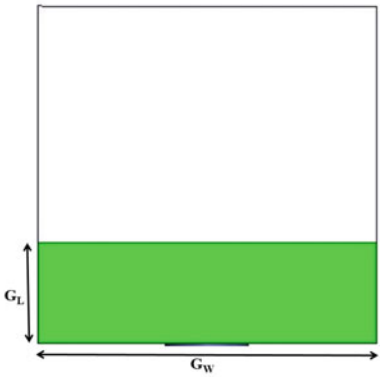
Initially, a square loop radiator is designed with 16 mm outer length and 14 mm inner length using partial ground. To achieve 50 Ω impedance, matching feed (8 mm × 3 mm) is taken, which is connected with a SMA connector to minimize the power loss. To modify in the obtained 10 dB bandwidth, three stubs are accomplished in a loop-shaped radiator. The upper two stubs are responsible for impedance matching in lower operating (3.57 GHz) band, and horizontally inverted L-shaped stub is responsible for impedance matching in higher operating band (5.2 GHz). The simulated |S11| is depicted in Fig. 1d_r where |S11| in dB is varying from 3.35 to 3.8 GHz and from 5.15 to 5.25 GHz. The realized gain (Peak) at both operating band is positive, and its maximum value is 2.32 dB as illustrated in Fig. 1e_r.

3 Results and Discussion

The proposed compact antenna is designed on HFSS-13 EM software. The bandwidth (10 dB) varies from 3.35 to 3.8 GHz and from 5.15 to 5.25 GHz as depicted in Fig. 2a_r. The peak gain (realized) is obtained 2.3 dB as depicted in Fig. 2b_r. The simulated and simulated results are close to each other. The dual-band antenna has more than 92% radiation efficiency in the first operating band and more than 70% radiation



(a_r)

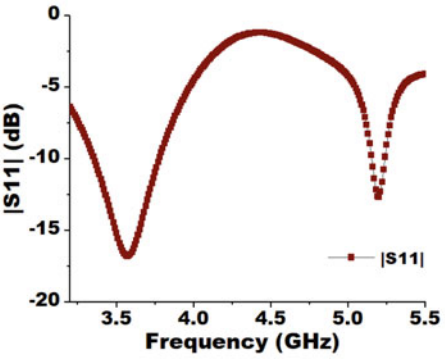


(b_r)

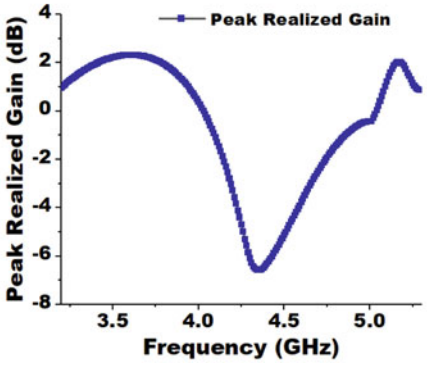
(c_r)

Fig. 1 Proposed antenna geometry. $\mathbf{a_r}$ Radiator, $\mathbf{b_r}$ ground, $\mathbf{c_r}$ prototype, $\mathbf{d_r}$ the simulated |S11| and $\mathbf{e_r}$ realized gain (Peak) of proposed antenna

Fig. 1 (continued)



(d_r)

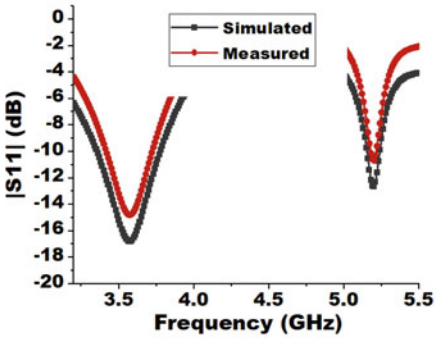


(e_r)

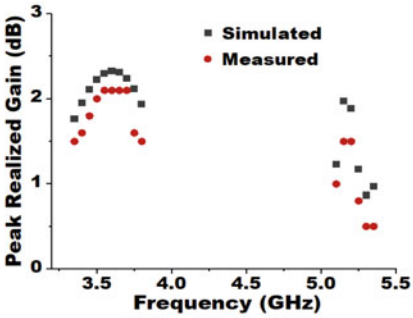
efficiency in the second operating bands depicted in Fig. 2 c_r. The radiation pattern (normalized) in xz and yz direction is illustrated in Fig. 2d_r at 3.57 GHz and 5.2 GHz resonant frequency. The proposed antenna is compared with the other published antenna in Table 1, and it reveals that the proposed antenna is having smaller size with better gain compared with other references in lower operating band.

4 Conclusion

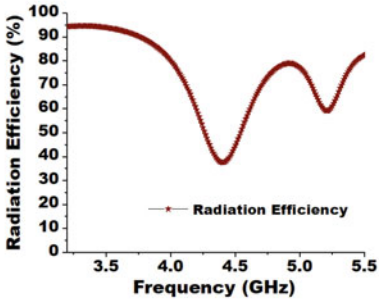
A square loop antenna with multiple stubs is designed and analysed with partial ground. The loop antenna accomplished with multiple stubs in radiator is responsible for dual operating bands. The proposed antenna achieved 10 dB impedance bandwidth from 3.35 to 3.8 GHz and from 5.15 to 5.25 GHz. The maximum realized gain is 2.32 dB at 3.6 GHz frequency. The pertaining dimensions of proposed microstrip antenna $W = 20$ mm and $L = 20$ mm. The proposed antenna with smaller dimensions is having positive gain with satisfactory radiation efficiency. The proposed antenna can be applicable in 5G/WLAN applications.



(a_r)

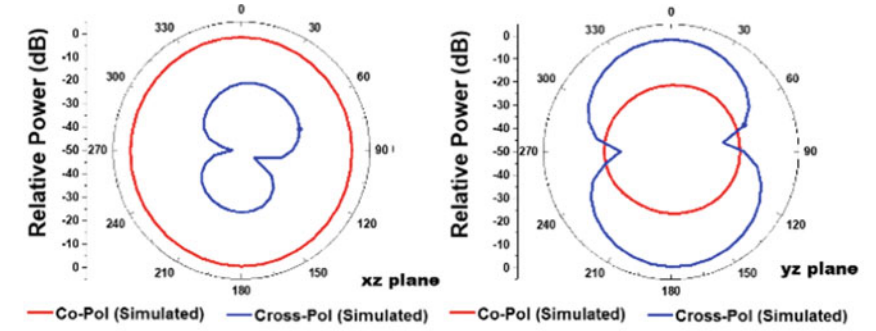


(b_r)

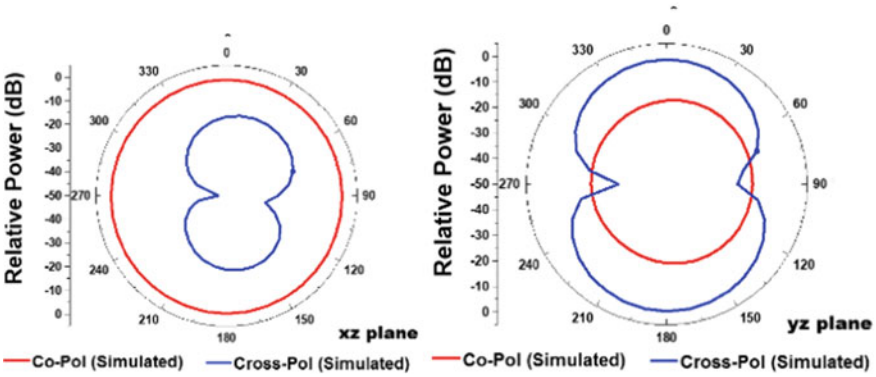


(C_r)

Fig. 2 $\mathbf{a_r}$ and $\mathbf{b_r}$ the simulated and measured |S11| and realized gain (Peak) $\mathbf{c_r}$ Simulated radiation efficiency and $\mathbf{d_r}$ are simulated radiation pattern at 3.3 GHz and 5.2 GHz of proposed antenna



(c$_r$)



(d$_r$)

Fig. 2 (continued)

Table 1 Comparison table with recently published work

References	Size (mm^2)	Operating Bandwidth (GHz)	Gain (dB)
2	25.2 × 23.7	2.4–8.3	0.70
4	33 × 20	2.3–2.5 3.5–5	2.1
5	40 × 30	3–5.64	1.73
6	27 × 30	3.30–3.70	2.60
8	29 × 26	2.38–2.48, 3.37–3.79, 4.36–6.06	3.27
9	30 × 20	2.65–2.9, 3.17–3.75, 5.55–6.1	1.4
10	30 × 30	2.43–2.61, 3.43–5.83	1.64
Proposed work	20 × 20	3.35–3.8, 5.15–5.25	2.32

References

1. Balanis CA (2015) Antenna theory: analysis and design. Wiley
2. Rajkumar R, Kiran KU (2018) Gain enhancement of compact multiband antenna with metamaterial superstrate. In: Optical and microwave technologies. Springer, Singapore, pp 91–98
3. Prasad KD, Ali T, Biradar RC (2017) A compact slotted multiband antenna for L-band and WLAN applications. In: 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), pp 820–823
4. Al-Khaldi M (2017) A highly compact multiband antenna for Bluetooth/WLAN, WiMAX, and Wi-Fi applications. Microwave Opt Technol Lett 59:77–80
5. Jha P, Singh S, Yadava RL (2021) Wideband Sub-6 GHz Micro-strip antenna: design and fabrication. In: Advances in smart communication and imaging systems. Springer, Singapore, pp 109–115
6. Er-Rebyiy R, Zbitou J, Tajmouati A, Latrach M, Errkik A, El Abdellaoui L (2017) A new design of a miniature microstrip patch antenna using Defected Ground Structure (DGS). In: International conference on wireless technologies, embedded and intelligent systems (WITS). IEEE, pp 1–4
7. Kaeib AF, Shebani NM, Kaeib AF, Shebani NM, Zarek AR (2019) Design and analysis of a slotted microstrip antenna for 5G communication networks at 28 GHz. In: 19th international conference on sciences and techniques of automatic control and computer engineering (STA). IEEE, pp 648–653
8. Wang S, Li K, Kong F, Du L (2021) A miniaturized triple-band planar antenna combing single-cell metamaterial structure and defected ground plane for WLAN/WiMAX applications. J Electromagn Waves Appl 35:357–370
9. Setia V, Sharma KK, Koul SK (2019) Triple-band metamaterial inspired microstrip antenna using split ring resonators for WLAN/WiMAX applications. In: IEEE Indian conference on antennas and propagation (InCAP). IEEE, pp 1–4
10. Murugeshwari B, Daniel RS, Raghavan S (2019) A compact dual band antenna based on metamaterial-inspired split ring structure and hexagonal complementary split-ring resonator for ISM/WiMAX/WLAN applications. Appl Phys A 125:1–8
11. Zeng J, Luk KM (2019) Single-layered broadband magneto electric dipole antenna for new 5G application. IEEE Antennas Wirel Propag Lett 18:911–915
12. Ghosh CK, Parui SK (2010) Design, analysis and optimization of a slotted microstrip patch antenna array at frequency 5.25 GHz for WLAN-SDMA system. Int J Electr Eng Inf 2:102–112

Design and Analysis of Graphene-Based Metasurface Absorber for Temperature and Refractive Index Sensing in THz Spectrum

Ramkrishna and Rajveer S. Yaduvanshi

Abstract This paper has proposed a model of graphene-based absorber in THz spectrum. This absorber consists of dielectric metasurface and array of micro-rods of graphene. The analysis of absorber is done using its equivalent electrical circuit to understand the variation of RLC parameters in THz spectrum. The proposed absorber can be designed for applications based on temperature sensing and refractive index (RI) estimation simultaneously. The result obtained after simulation shows that reflection coefficient is quite small at 5, 10, and 500 THz, and the Q-factor can be calculation through its mathematical definition for the proposed TA at room temperature of 295 K. The most fundamental and specific application of TA for a unit cell is bipolar resonance in its cell structure. The properties of TA specifically absorption can be altered by variation in frequency of excitation, quality factor, and cell structure under resonance, respectively. The TA acts as temperature sensor as the electrical property of graphene is sensitive to change in temperature in its surrounding. So, TA works quite well like a μ sensor by virtue of its high Q-factor which computes the inductance as the optical wave is not reflected back by the graphene metamaterial-based absorber. Thus, the proposed TA can be designed effectively for sensor or transducer and also for devices based on demodulation, filter circuit, and some electroluminescent devices in terahertz spectrum.

Keywords MDM · TA · Graphite · R · L · C circuit · Dynamic impedance · Quality factor · MATLAB · Reflection coefficient

Ramkrishna (✉)
Department of Electronics and Communication Engineering, AIACTR, GGSIPU, Dwarka, Delhi, India
e-mail: yadavramkrishna@gmail.com

R. S. Yaduvanshi
Department of Electronics and Communication Engineering, NSUT, Dwarka, Delhi, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_9

1 Introduction

Nearly, all the electromagnetic radiation (EMR) can be absorbed by the elimination process of transmitted and reflected radiation simultaneously by metamaterials (MMs) known to be as precise absorbers (TA) at a given frequency [1, 2]. TAs are well known for decades about their application prospects in the spectrum from microwave to visible such as harvesting-energy, imaging, emitting, sensing, etc. [2–8]. From the design perspective of TA, it has been explored that metal-dielectric-metal (MDM) multi-layered structure is chosen specifically three-layered structure, but its complex geometry and fabrication cost both can limit is application as sensor [6–13]. The progress in design of PA in terahertz (THz) spectrum has been due to fact of poor electromagnetic response of nature of substrate for this THz spectrum [7–16]. The most appropriate class of two-dimensional (2-D) artificially designed metamaterial is metasurface characterized by periodic subwavelength and composite planar structure based on single-layer pattern of either conductor, dielectric, or hybrid dielectric-conductor (DD) structure [17, 18]. The bottomless width of subwavelength of metasurface can be useful for obtaining significant electromagnetics spectacles and planar designs of THz functional devices such as polarized conversion, refraction and reflection anomalously, reduction of sensor cross-segment (SCS), converging lenses and absorbers [19–23]. Intensive and significant developments are proposed in Pas based on several metasurfaces due to their several application prospects in THz spectrum [14, 24–34]. For instance, Wang, et al., have showed mathematically a multi-band PA employing a resonator with four split ring type with various dimensions [26]. In [29], the author Luo et al., has presented a TA with a resonator having dual-layer PDMS structure adhered on the metal-ground that is designed to obtain a dual-band absolute absorption in the THz frequency spectrum. Similarly, Yan et al. have presented numerically an ultra-high-quality factor dual-band TA with dielectric grating slit waveguide utilized in refractive index sensing [31]. Since the absorption spectrum mostly are fixed of TAs once it comes to an end of its design and fabrication because of this its application becomes practically limited. Also, the hybrid MDM or DM structure of most PAs tends to increase both complexity in structure and its fabrication cost. Instead, PAs having only simple structure find its more applications in THz spectrum like switch, detector, and sensor [35–41]. So, it is important to design a PAs having simple metasurface structure which can be tuned actively [42–45]. For the purpose of designing high-performance tunable Pas, all dielectric metasurface could be ideal. So, TAs of tunable type are used for all dielectric graphene metasurface for demonstration purpose in THz spectrum. The structure of unit cell pf proposed PA consists of only of a subwavelength graphene micro-rod structure etched into a graphene substrate. The research shows that the absorption peak of PAs is about 99.9% at 1.757 THz with environmental temperature being at 295 K. The corresponding Q-factor for of the PAs is approximately about 53.24 as an influence of parameter like geometric structure of unit cell on the absorption characteristics of studied PAs. The sensitivity of PA was found to be about 4.2 GHz/K for detection of changes in external or environmental temperature. However, the PAs

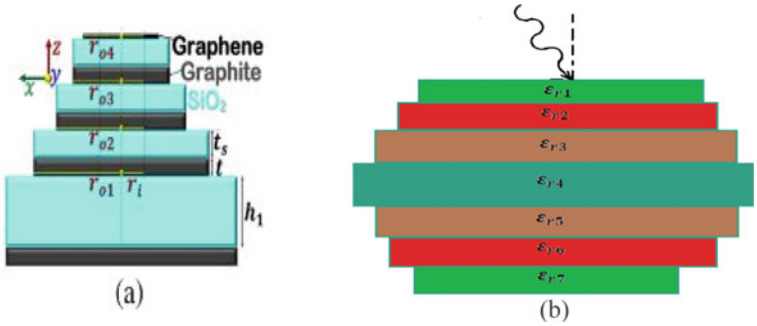


Fig. 1 **a** Front view of absorber, **b** conceptual diagram of proposed absorber

are also used for the detection of variations in RI of surrounding, and the sensitivity of PA was found experimentally about 1043.3 GHz/RIU at 295 K and 920 GHz/RIU at 300 K, respectively. The high value of quality factor of PAs along with its absolute absorption properties makes applicable for various applications in chemical and biological areas of multifunctional optoelectronic sensing devices. The TAs proposed and designed by us are all dielectric type which has very simple geometric structure as compared to that of all multi-layered structured TAs designed earlier and also our designed TAs because of its simplicity in design makes it suitable for integration in recent photoelectric systems. The fabrication of our proposed all dielectric TAs is also quite simpler than other TAs designed earlier which is another significant advantage with our designed TAs and delimits the constraints for optical spectrum with significant applications in detection in THz spectrum (Fig. 1).

In healthcare infrastructure, all device is now practically designed in the form of sensors-based application, for instance refractive index (RI) sensors and temperature sensor. With reference to above applications of sensors in healthcare monitoring and control domain, our purpose is to design a sensor which absorbs light or any electromagnetic wave at different terahertz frequency. For the purpose of design of this sensor, we use graphene with combination of silicon dioxide (SiO_2) as an intermediate substrate having the probability of changing its dielectric strength by different level of doping concentration. The absorber has multi-metamaterial configuration with various dielectric strength which can be synthesized as an electrical RLC circuit with combination of RL, and RC is series and parallel.

Explanation of Absorption Phenomena in Absorber

When the light is injected on the top aperture of absorber, then it enters the absorber then some part of light is refracted or transmitted into the dielectric layer below the top surface, and some part of incident light is reflected back toward the top metal surface. This metal surface does not allow the reflected light to escape out of the absorber and sends back to the absorber. The arrangement of dielectric substrate within the absorber is such that the dielectric strength of metamaterial reduces the top metallic surface of absorber to from the top metallic surface of absorber to the base metallic surface of absorber. At each interface of dielectric surface, maximum

portion of light is refracted into the next dielectric layer, and this process keeps on until the light reaches the bottom part of the absorber where the metallic base reflects back the incident light back into the absorber into the absorber in upward direction, thus preventing the light to escape out of the absorber from bottom. The above phenomena are analogous to resonance effect in the proposed RLC circuit. The light is thus trapped in the absorber and keeps transmitting and reflecting within the absorber among different dielectric layers of silicon dioxide substrate which can be coincidence as oscillator of light.

Explanation of Equivalent Circuit Absorber

With respect to the phenomena that happens in absorber with light as input, the proposed RLC circuit can be considered as equivalent circuit of absorber with voltage signal as input analogous to light input in absorber.

Our objective is obtaining the resonance condition in the proposed circuit so that the reactive energy oscillates between the capacitive and inductive element. And as the terahertz frequency value is increased, the active power loss reduces and becomes negligible, and the reactive power (or energy) stored in the reactive element tends to increase and becomes maximum under resonance condition. This is achieved mathematically by deriving the dynamic impedance, quality factor of the proposed RLC circuit. The input impedance has real and reactive part. The dynamic impedance is real part input impedance, which is dependent upon frequency, and its magnitude decreases with the increase in frequency in terahertz spectrum.

How Oscillation of Energy Happens Between Inductor and Capacitor

Initially, both the inductor and capacitor are uncharged having no energy. The capacitor then charged with the help of an external energy source; then, the energy stored in capacitor is $\frac{Q^2}{2C}$, and both plates of capacitor carry equal and opposite charge.

At $t = 0$, an inductor is connected in parallel to the charge capacitor. For ideal case, the resistance of inductive element is neglected which will be considered in practical cases.

The electron from negatively charged plate of capacitor will flow toward the inductor, as a result a current will flow apposite to flow of electrons. This current flow through inductor will generate flux in the inductor core $\emptyset = LI$; the flux will cause an induced emf across the inductor terminals with same polarity as the of capacitor, and when the emf induced in inductor becomes equal to that of capacitor, then the flow of charge stops. At this moment, the energy stored in capacitor becomes zero, and the energy stored in inductor is maximum, and the current is also at its maximum value.

At $t = 3$ T/4, the inductor starts transferring energy back to capacitor such that the polarity of capacitor is reversed as compare to its previous polarity. Now, the direction of current will reverse because the energy transfer takes place from inductor to capacitor. This process of energy transfer between capacitor and inductor and vice versa is termed as oscillation of energy and charge in undamped lc circuit because no energy loss takes place.

2 Analysis of Absorber and Its Equivalent Circuit

The absorber consists of several plates of graphene with a dielectric material (SiO_2) placed between two successive plates of graphene. The equivalent electric circuit of the proposed absorber is given Fig. 2.

The graphene plate has two carbon coil for which above circuit is considered. This circuit has two resonant modes and so two resonant frequencies in THz spectrum. One resonant mode exists due to tuning between the impedance Z_2 and Z_3, respectively, as parallel resonance. The second resonant mode exists due to tuning between the impedance Z_p (parallel equivalent of Z_2 & Z_3 inductive in nature) and Z_1, which is capacitive. The input impedance of absorber is represented by $z_{in}(s)$ derived below.

The main purpose of deriving the input impedance is to obtain its real and imaginary part. With increase in THz frequencies, the real part of input impedance can be reduced that represents radiation of input energy can be minimized, and maximizing the imaginary part of input impedance represents that the energy stored in the reactive element of the circuit can be maximized.

$$z_1(s) = R_1 + \frac{1}{sC_1} \tag{1}$$

$$z_2(s) = R_2 + sL_2 \tag{2}$$

$$z_3(s) = R_3 + \frac{1}{sC_3} \tag{3}$$

$$z_P(s) = z_2(s)z_P(s) = z_2(s) \| z_3(s) = \frac{z_z(s)z_3(s)}{z_2(s) + z_3(s)}$$

$$z_{\text{in}}(s) = z_1(s) + z_P(s),\ z_{\text{in}}(s) = z_1(s) + \frac{z_2(s)z_3(s)}{z_2(s) + z_3(s)}$$

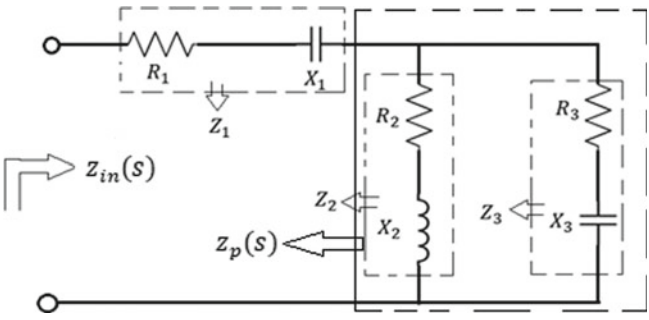


Fig. 2 Equivalent circuit of proposed absorber

$$z_{\text{in}}(s) = R_1 + sc_1 + \frac{(R_2 + sL_2)\left(R_3 + \frac{1}{sc_3}\right)}{R_i + sL_2 + R_3 + \frac{1}{sc_3}} \tag{4}$$

$$z_{\text{in}}(s) = R_1 + sc_1 + \frac{R_2R_3 + sL_2R_3 + \frac{L_2}{c_3} + \frac{R_2}{sc_3}}{R_2 + R_3 + sL_2 + \frac{1}{Sc_3}}$$

$$z_{\text{in}}(s) = \frac{(R_1 + sc_1)\left(R_2 + R_3 + sL_2 + \frac{1}{sc_3}\right) + R_2R_3 + sL_2C_3 + \frac{L_2}{c_3} + \frac{R_2}{sc_3}}{R_2 + R_3 + sL_2 + \frac{1}{sc_3}}$$

$$z_{\text{in}}(s) = \frac{R_1R_2 + R_1R_3 + sL_2R_1 + sL_2R_3 + R_2R_3 + \frac{R_1}{sc_3} + \frac{R_2}{sc_1} + \frac{R_3}{sc_1} + \frac{R_2}{sc_3} + \frac{1}{c_1c_3} + \frac{L_2}{c_1} + \frac{L_2}{c_3}}{R_2 + R_3 + sL_2 + \frac{1}{sc_3}}$$

$$z_{\text{in}}(s) = \frac{R_1R_2 + R_1R_3 + R_2R_3 + \frac{1}{c_1c_3} + \frac{L_2}{c_1} + \frac{L_2}{c_3} + s(L_2R_3 + L_2R_1) + \frac{1}{s}\left(\frac{R_1}{c_3} + \frac{R_2}{c_1} + \frac{R_3}{c_1} + \frac{R_2}{c_3}\right)}{R_2 + R_3 + sL_2 + \frac{1}{sc_3}}$$

$$z_{\text{in}}(s) = \frac{s^2c_3(L_2R_1 + L_2R_3) + sc_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right) + R_1 + R_2 + \frac{(R_2+R_3)C_3}{c_1}}{s^2L_2C_3 + sc_3(R_2 + R_3) + 1}$$

$$z_{\text{in}}(s) = \frac{s^2(C_3L_2R_1 + C_3L_2R_3) + sC_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right) + R_1 + R_2 + \frac{(R_2+R_3)C_3}{c_1}}{s^2L_2C_3 + sc_3(R_2 + R_3) + 1}$$

Put $s = j\omega and s^2 = -\omega^2$

$$z_{\text{in}}(s) = \frac{R_1 + R_2 + \frac{(R_2+R_3)C_3}{c_1} - \omega^2(c_3L_2R_1 + c_3L_2R_3) + j\omega c_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right)}{(1 - \omega^2L_2C_3) + j\omega C_3(R_i + R_3)} \tag{5}$$

$$z_{\text{in}} = \frac{a + jb}{c + jd}$$

where $a = R_1 + R_2 + \frac{(R_2+R_3)C_3}{c_1} - \omega^2(C_3L_2R_1 + C_3L_2R_3)$

$b = \omega c_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_3c_1}\right)$,

$$c = [1 - \omega^2L_2C_3],\ d = \omega C_3(R_2 + R_3).$$

Convert real and imaginary part separately

$$z_{\text{in}}(j\omega) = \frac{(a + jb)(c - j\text{d})}{c^2 + d^2} = \frac{ac - jad + jcb + bd}{c^2 + d^2}$$

$$z_{\text{in}}(j\omega) = \frac{ac+bd}{c^2+d^2} + j\frac{(cb-ad)}{c^2+d^2}$$

For the resonance imaginary part of input impedance $z_{\text{in}}(j\omega) = 0$

$$\frac{cb-ad}{c^2+d^2} = 0, cb-ad = 0, cb = ad$$

$$\left(1-\omega^2 L_2 C_3\right)\left\{\omega C_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right)\right\}$$
$$= \left\{R_1 + R_2 + \frac{(R_2+R_3)c_3}{c_1} - \omega^2(C_3L_2R_1 + C_3L_2R_3)\right\}\{\omega C_3(R_1+R_2)\}$$

$$\omega^2\left[\frac{L_2C_3\left(R, R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right)}{R_2 + \frac{R}{3}} - (C_3L_2R_1 + C_3L_2R_3)\right]$$
$$= \frac{R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1}c_3}{R_2+R_3} - \left\{\frac{R_1 + R_2 + (R_2+R_3)c_3}{c_1}\right\}$$

$$\omega_R = \sqrt{\frac{R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3} - \left\{R_1 + R_2 + \frac{(R_2+R_3)C_3}{c_1}\right\}}{\frac{L_2c_3\left(R, R_2+R_1R_3+R_2R_3+\frac{L_2}{c_1}+\frac{L_2}{c_3}+\frac{1}{c_1c_3}\right)}{R_2+R_3} - (C_3L_2R_1 + C_3L_2R_3)}} \quad (6)$$

$$Q = \frac{\frac{cb}{c^2+d^2}}{\frac{ac+bd}{c^2+d^2}} = \frac{cb}{ac+bd}$$

$$Q = \frac{(1-\omega^2L_2C_3)\omega C_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right)}{\left(R_1 + R_2 + \frac{(R_2+R_3)C_3}{c_1} - \omega^2(C_3L_2R_1 + C_3L_2R_3)\right)(1-\omega^2L_2C_3) + \omega c_3(R_2+R_3)\omega c_3\left(R_1R_2 + R, R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right)} \quad (7)$$

3 Complex Reflection Coefficient (Γ)

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = S_{11}$$

$$Z_0 = \sqrt{\frac{R_s + j\omega L_s}{\frac{1}{R_s} + j\omega C_s}} = \sqrt{\frac{R_s^2 + j\omega L_s R_s}{1 + j\omega C_s R_s}} = \sqrt{\frac{\left(R_s^2 + j\omega L_s R_s\right)((1 - j\omega C_s R_s)}{(1 + j\omega C_s R_s)((1 - j\omega C_s R_s)}}$$

$$Z_0 = \sqrt{\frac{R_s^2 + \omega^2 L_s R_s^2 C_s + j\left(\omega L_s R_s - \omega C_s R_s^3\right)}{\left(1 + \omega^2 C_s R_s\right)}}$$

$$Z_0 = \sqrt{\frac{R_s^2 + \omega^2 L_s R_s^2 C_s}{\left(1 + \omega^2 C_s R_s\right)}}\left[1 + \frac{j\left(\omega L_s R_s - \omega C_s R_s^3\right)}{2\left(R_s^2 + \omega^2 L_s R_s^2 C_s\right)}\right] \tag{8}$$

$z_L(j\omega) = \frac{ac+bd}{c^2+d^2} + j\frac{(cb-ad)}{c^2+d^2} = (P + jX)$, where , $\mathrm{P} = \frac{\mathrm{ac+bd}}{\mathrm{c^2+d^2}}$ & $\mathrm{X} = \frac{\mathrm{(cb-ad)}}{\mathrm{c^2+d^2}}$

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = S_{11} = \frac{P + jX - Z_0}{P + jX + Z_0}, \Gamma = S_{11} = \frac{(P - Z_0) + j(X)}{(P + Z_0) + j(X)}.$$

Magnitude of complex reflection coefficient

$$|S_{11}| = |\Gamma| = \left|\frac{(P - Z_0) + j(X)}{(P + Z_0) + j(X)}\right| = \frac{\sqrt{(P - Z_0)^2 + (X)^2}}{\sqrt{(P + Z_0)^2 + (X)^2}} \tag{9}$$

Figure 3 shows the variation of reflection coefficient (S_{11}), input impedance (Z_{in}), real and imaginary part of Z_{in}, and quality factor (Q) of the equivalent electrical circuit in the 10 THz spectrum. The reflection coefficient decreases with increase in frequency which signifies that the absorber is working as per its desired phenomena. The real part of input impedance (Z_{in}) which is responsible for power loss or radiation is also decreasing as represented by its graphical demonstration. The imaginary part of input impedance is increasing with increase in THz frequency thereby stating that the energy stored in the reactive element of equivalent electrical circuit of absorber is increasing. Also, the quality factor of the circuit is observed to be increasing with the increase in the frequency thereby stating that the absorber is working quite well as represented by its equivalent electrical network. Similar graphical demonstration has been shown for 500 and 5 THz frequency. All the parameters plotted show that the circuit is behaving corresponding to the phenomena of how an absorber would work at different frequency in the THz spectrum.

Figure 4 shows the variation of reflection coefficient (S_{11}), input impedance (Z_{in}), real and imaginary part of Z_{in}, and quality factor (Q) of the equivalent electrical circuit in the 500 THz spectrum. The reflection coefficient decreases with increase in frequency which signifies that the absorber is working as per its desired phenomena. The real part of input impedance (Z_{in}) which is responsible for power loss or radiation is also decreasing as represented by its graphical demonstration. The imaginary part of input impedance is increasing with increase in THz frequency thereby stating that the energy stored in the reactive element of equivalent electrical circuit of absorber is increasing. Also, the quality factor of the circuit is observed to be increasing with the increase in the frequency thereby stating that the absorber is working quite well as represented by its equivalent electrical network.

Figure 5 shows the variation of reflection coefficient (S_{11}), input impedance (Z_{in}), real and imaginary part of Z_{in}, and quality factor (Q) of the equivalent electrical circuit in the 5 THz spectrum.

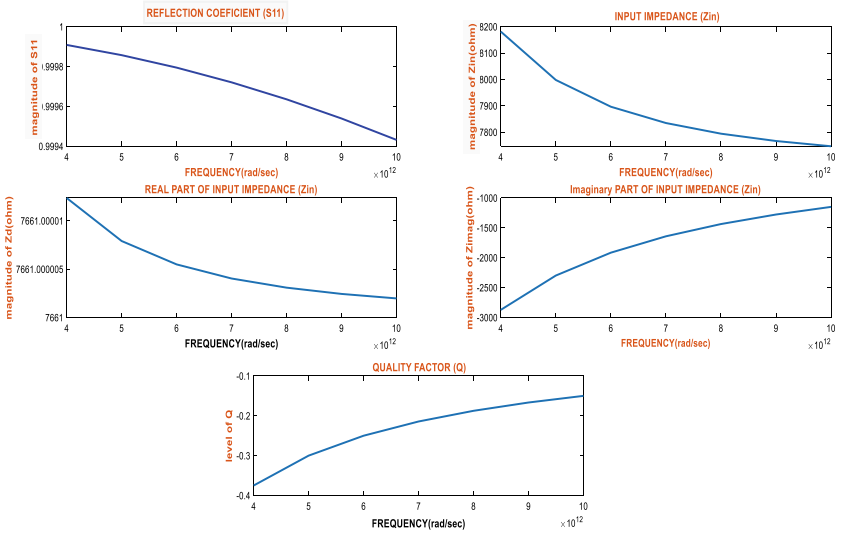
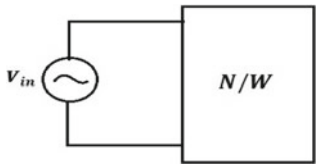


Fig. 3 Reflection coefficient, dynamic Impedance vs frequency (rad/s) spectrum at 10 THz

4 Results and Discussions

Input impedance of the network is



$$z_{\text{in}}(j\omega) = \frac{ac+bd}{c^2+d^2} + j\frac{(cb-ad)}{c^2+d^2}$$

$$z_{\text{in}}(j\omega) = \sqrt{\frac{(ac+bd)^2+(cb-ad)^2}{(c^2+d^2)^2}} \angle \tan^{-1}\left(\frac{cb-ad}{ac+bd}\right) \tag{10}$$

$$\emptyset = \tan^{-1}\frac{cb-ad}{ac+bd}$$

Power factor $= \cos\phi$.
Power factor $\cos\phi = \frac{\text{Active power}}{\text{apprant power}}$

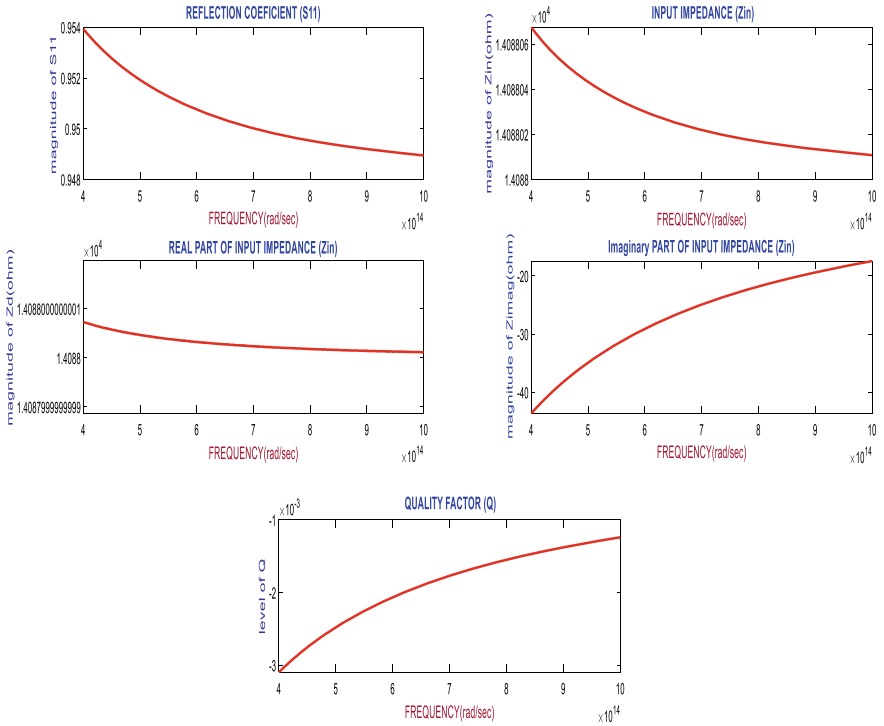
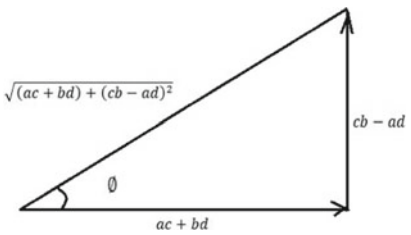


Fig. 4 Reflection coefficient, dynamic impedance vs frequency (rad/s) spectrum at 500 THz

$$\tan\phi = \frac{cb - ad}{ac + bd}$$

$$\cos\phi = \frac{ac + b\mathrm{d}}{\sqrt{(ac + bd)^2 + (cb - ad)^2}}$$



Active power $ac + bd = 0$ (only zero power factor)

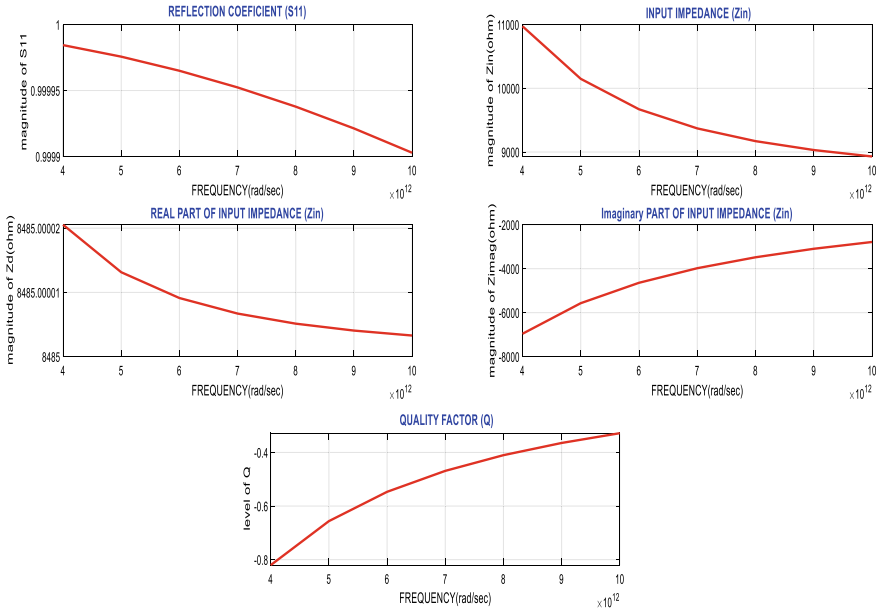


Fig. 5 Reflection coefficient, dynamic impedance vs frequency (rad/s) spectrum at 5 THz

$$\left\{R_1 + R_2 + \frac{(R_2 + R_3)c_3}{c_1} - \omega^2(c_3L_2R_1 + C_3L_2R_3)\right\}\left(1 - \omega^2L_2c_3\right) + \omega c_3(R_2 + R_3)\left\{\omega c_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{\lg} + \frac{1}{c_1c_3}\right)\right\} = 0$$

$$\omega^4L_2^2C_3^2(R_1 + R_3) + \omega^2\left[\left\{c_3^2(R_2 + R_3)\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1}\frac{+L_2}{c_3} + \frac{1}{c_1c_3}\right)\right\} - c_3L_2(R_1 + R_3) - L_2C_3\left\{R_1 + R_2 + \frac{(R_2 + R_3)C_3}{c_1}\right\}\right] + R_1 + R_2 + \frac{(R_2 + R_3)c_3}{c_1} = 0$$

Apparent power is maximizing $f = \sqrt{(ac + bd)^2 + (cb - ad)^2} = mssax$

$$f = \sqrt{\left[\left\{R_1 + R_2 + \frac{(R_2 + R_3)C_3}{C_1} - \omega^2C_3L_2(R_1 + R_3)\right\}\{l - \omega^2L_2C_3\}\right]^2 + \left[\left\{\omega C_3\left(R_1R_2 + R_1R_3 + R_2R_3 + \frac{L_2}{c_1} + \frac{L_2}{c_3} + \frac{1}{c_1c_3}\right)\right\}\omega C_3(R_2 + R_3)\right]^2} \quad (11)$$

When the light is incident on the absorber, the maximum part of input power becomes reactive, and the real part power is minimum. This is due to the fact that the dynamic impedance of input impedance is purely real and inversely proportional to terahertz frequency, so as the THz frequency increases real power of input light decreases, and the reactive power increases and oscillates with resonance frequency. But practically, the magnitude of reactive power decays as exponential sinusoidal wave due to inherent resistance of absorber (Table 1).

$$
\begin{aligned}
z_{\text{in}}(j\omega) &= \frac{ac+bd}{c^2+d^2} + j\frac{(cb-ad)}{c^2+d^2} \\
f(\omega) &= c^2 + d^2 \\
f(\omega) &= \left(1-\omega^2 L_2 C_3\right)^2 + \omega^2 c_3^2 (R_2+R_3)^2 \\
f(\omega) &= 1 + \omega^4 L_2^2 c_3^2 - 2\omega^2 L_2 C_3 + \omega^2 c_3^2 (R_2+R_3)^2 \\
f(\omega) &= \omega^4 L_2^2 c_3^2 + \omega^2 \left\{c_3^2 (R_2+R_3)^2 - 2L_2 C_3\right\} + 1 \\
\frac{\mathrm{d}f(\omega)}{\mathrm{d}\omega} &= 0 \\
\frac{\mathrm{d}f(\omega)}{\mathrm{d}\omega} &= 4\omega^3 L_2^2 C_3^2 + 2\omega\left\{C_3^2 (R_2+R_3)^2 - 2L_2 C_3\right\} \\
4\omega^3 L_2^2 C_3^2 &+ 2\omega\left\{c_3^2 (R_2+R_3)^2 - 2L_2 C_3\right\} = 0 \\
\omega^2 &= \frac{2L_2 - C_3 (R_2+R_3)^2}{2L_2^2 C_3} \\
\omega &= \pm\sqrt{\frac{2L_2 - C_3 (R_2+R_3)^2}{2L_2^2 c_3}} \qquad (12) \\
\frac{\mathrm{d}^2 f(\omega)}{\mathrm{d}\omega^2} &= 12\omega^2 L Z_3^{c^2} + 2\left\{C_3^2 (R_2+R_3)^2 - 2L_2 C_3\right\}
\end{aligned}
$$

For maxima, $\frac{\mathrm{d}^2 f(\omega)}{\mathrm{d}\omega^2} < 0$, For minima, $\frac{\mathrm{d}^2 f(\omega)}{\mathrm{d}\omega^2} > 0$

$$\frac{\mathrm{d}^2 f(\omega)}{\mathrm{d}\omega^2} < 0$$

Table 1 Frequency variation (Terahertz) vs electrical parameters

Frequency (THz)	Synthesis electrical element					
	L_2 (H)	$R_2(\Omega)$	R_3 (Ω)	$R_1(\Omega)$	C_1 (F)	C_3 (F)
5	1.0047	6899	4955	3530	25×10^{-10}	143×10^{-10}
10	0.4654	3881	4131	3530	129×10^{-10}	145×10^{-10}
500	3.0638	9999	10,558	3530	129×10^{-11}	145×10^{-11}

$$12\omega^2 L_2^2 c_3^2 + 2\left\{c_3^2(R_2+R_3)^2 - 2L_2C_3\right\} < 0$$

$$6L_2^2 c_3\left(\frac{2L_2 - C_3(R_2+R_3)^2}{2L_2^2 c_3}\right) + \left\{c_3(R_2+R_3)^2 - 2L_2\right\} < 0$$

$$6L_2 - 3C_3(R_2+R_3)^2 + C_3(R_2+R_3)^2 - 2L_2 < 0$$

$$L_2 < \frac{C_3(R_2+R_3)^2}{2} \tag{13}$$

At 5, 10, and 500 THz, the electrical parameters like R, L, C have been estimated mathematically for the absorber. We have derived mathematically the criteria at which the absorber functions quite will and absorbs the incoming light.

5 Conclusions

The absorber designed shows that it absorbers the optical wavelength in terahertz spectrum on the top surface with maximum transmission and minimum reflection coefficient, respectively. This is also evident from simulation result and graphical analysis of reflection coefficient (S_{11}). From the graphical analysis of input impedance of absorber, it is evident that as frequency in terahertz spectrum increases, the dynamic impedance (real part of input impedance) decreases thereby reducing the power loss through the radiation and imaginary part of input impedance increases, which is responsible for oscillation of power within the absorber, respectively, the improvement in quality factor of equivalent analogous circuit.

References

1. Landy NI, Sajuyigbe S, Mock JJ, Smith DR, Padilla WJ (2008) Perfect metamaterial absorber. Phys Rev Lett 100:207402
2. C.M. Watts, X. Liu, W.J. Padilla, Metamaterial electromagnetic wave absorbers, Adv. Mater. 24 (2012) OP98–OP120.
3. Iwaszczuk K, Strikwerda AC, Fan K, Zhang X, Averitt RD, Jepsen PU (2012) Flexible metamaterial absorbers for stealth applications at terahertz frequencies. Opt Express 20:635–643
4. Fan K, Suen J, Wu X, Padilla WJ (2016) Graphene metamaterial modulator for freespace thermal radiation. Opt Express 24(22):25189–25201
5. Wang B-X, Xiang Z, Wang G-Z, Huang W-Q, Wang L-L (2015) A novel dual-band terahertz metamaterial absorber for a sensor application. J Appl Phys 117:014504
6. Wang B-X, Wang G-Z, Wang L-L, Xiang Z (2016) Design of a five-band terahertz absorber based on three nested split-ring resonators. IEEE Photon Technol Lett 28:307
7. Wang B-X, Wang G-Z, Wang L-L (2016) Design of a novel dual-band terahertz metamaterial absorber. Plasmonics 11:523–530

8. Suen JY, Fan K, Padilla WJ, Liu X (2017) All-dielectric metasurface absorbers for uncooled terahertz imaging. Optica 4(6):601–604
9. Wang B-X (2017) Quad-band terahertz metamaterial absorber based on the combining of the dipole and quadrupole resonances of two SRRs. IEEE J Sel Top Quant Electron 23:4700107
10. Luo H, Cheng YZ (2018) Ultra-thin dual-band polarization-insensitive and wideangle perfect metamaterial absorber based on a single circular sector resonator structure. J Electron Mater 47(1):323–328
11. Wang B-X, Tang C, Niu Q, He Y, Chen R (2019) A broadband terahertz metamaterial absorber enabled by the simple design of a rectangularshaped resonator with an elongated slot. Nanoscale Adv 1:3621
12. Cheng Y, Luo H, Chen F, Gong R (2019) Triple narrow-band plasmonic perfect absorber for refractive index sensing applications of optical frequency. OSA Continuum 2(7):2113–2122
13. Wang B-X, Tang C, Niu Q, He Y, Chen T (2019) Design of narrow discrete distances of dual-/triple-band terahertz metamaterial absorbers. Nanoscale Res Lett 14:64
14. Li J, Chen X, Yi Z, Yang H, Tang Y, Yi Y, Yao W, Wang J, Yi Y (2020) Broadband solar energy absorber based on monolayer molybdenum disulfide using tungsten elliptical arrays. Mater Today Energy 16:100390
15. Wang B-X, He Y, Lou P, Huang W-Q, Pi F (2020) Pentaband terahertz light absorber using five localized resonance responses of three patterned resonators. Res Phys 16:102930
16. Wang B-X, He Y, Lou P, Xing W (2020) Design of a dual-band terahertz metamaterial absorber using two identical square patches for sensing application 2:763–769
17. Yu N, Aieta F, Genevet P, Kats MA, Gaburro Z, Capasso F (2012) A broadband, background-free quarter-wave plate based on plasmonic metasurfaces. Nano Lett 12:6328–6333
18. He Q, Sun S, Xiao S, Zhou L (2018) High-efficiency metasurfaces: principles, realizations, and applications. Adv Optic Mater 1800415
19. Chen K, Feng Y, Monticone F, Zhao J, Zhu B, Jiang T, Zhang L, Kim Y, Ding X, Zhang S, Alù A, Qiu CW (2017) A reconfigurable active huygens' metalens. Adv Mater 29(17):1606422
20. Yang JJ, Zhi Cheng Y, Dong Q, Gong RZ (2018) Study of energy scattering relation and RCS reduction characteristic of matrix-type coding metasurface. Appl Sci 8(8):1231
21. Fahad AK, Ruan CJ, Chen K (2019) A wideband terahertz transmissive polarization manipulator based on metasurfaces. Electronics 8:1068
22. Fan J, Cheng Y (2020) Broadband high-efficiency cross-polarization conversion and multi-functional wavefront manipulation based on chiral structure metasurface for terahertz wave. J Phys D Appl Phys 53:025109 (10pp)
23. Yang D, Zhang C, Ju X, Ji Y, Lan C (2020) Multi-resonance and ultra-wideband terahertz metasurface absorber based on micro-templateassisted self-assembly method. Opt Express 28(Issue 2):2547–2556
24. Cong L, Tan S, Yahiaoui R, Yan F, Zhang W, Singh R (2015) Experimental demonstration of ultrasensitive sensing with terahertz metamaterial absorbers: a comparison with the metasurfaces. Appl Phys Lett 106:031107
25. Wang BX, Wang GZ, Sang T (2016) Simple design of novel triple-band terahertz metamaterial absorber for sensing application. J Phys D 49:165307
26. Wang X, Wang Q, Dong G, Yanan H, Lei M, Ke B (2017) Multiband terahertz metasurface absorber, Mod. Phys. Lett. B 31 (No. 36) 1750354
27. Gao J, Lan C, Zhao Q, Li B, Zhou J (2018) Experimental realization of Mie-resonance terahertz absorber by self-assembly method. Opt Express 26(10):13001–13011
28. Cheng YZ, Zuo X, Huang ML, Wang T, Gong R (2019) Design of a photo-excited broadband tunable terahertz absorber. J Infrared Millim Waves 38(1):97–102
29. Luo H, Cheng Y (2019) Dual-band terahertz perfect metasurface absorber based on bi-layered all-dielectric resonator structure. Opt Mater 96:109279
30. Wang Y, Chen Z, Xu D, Yi Z, Chen X, Chen J, Tang Y, Wu P, Li G, Yi Y (2020) Triple-band perfect metamaterial absorber with good operating angle polarization tolerance based on split ring arrays. Res Phys 16:102951

31. Yan F, Qi L, Tian H, Wang Z, Li L (2020) Ultrahigh Q-factor dual-band terahertz perfect absorber with dielectric grating slit waveguide for sensing. J Phys D Appl Phys. https://doi.org/10.1088/1361-6463/ab7bb5
32. Geng Z, Su W, Wang X, Jiang Y, Liu Y (2019) Numerical design of a metasurface-based ultra-narrow band terahertz perfect absorber with high Qfactors. Optik 194:163071
33. Li J, Chen Z, Yang H, Yi Z, Chen X, Yao W, Duan T, Wu P, Li G, Yi Y (2020) Tunable broadband solar energy absorber based on monolayer transition metal dichalcogenides materials using Au nanocubes. Nanomaterials 10:257
34. Cen C, Chen Z, Xu D, Jiang L, Chen X, Yi Z, Wu P, Li G, Yi Y (2020) High quality factor, high sensitivity metamaterial graphene—perfect absorber based on critical coupling theory and impedance matching. Nanomaterials 10:95
35. Chen, F. Cheng, Y.Z., Luo, H.: A broadband tunable terahertz metamaterial absorber based on single-layer complementary gammadion-shaped graphene. Materials 13:860
36. Zhao X, Wang Y, Jacob S, Duan G, Kevin C, Zhang J, Chen C, Richard D, Averitt, Zhang X (2020) Optically modulated ultra-broadband all silicon metamaterial terahertz absorbers. ACS Photonics 6(4):830–837
37. Zou HJ, Cheng YZ (2019) Design of a six-band terahertz metamaterial absorber for temperature sensing application. Opt Mater 88:674–679
38. Li WY, Cheng YZ (2020) Dual-band tunable terahertz perfect metamaterial absorber based on strontium titanate (STO) resonator structure, Opt Commun. 462:125265
39. Luo H, Cheng Y (2020) Thermally tunable terahertz metasurface absorber based on all dielectric indium antimonide resonator structure. Opt Mater 102:109801
40. Chen F, Cheng Y, Luo H (2020) Temperature tunable narrow-band terahertz metasurface absorber based on InSb micro-cylinder arrays for enhanced sensing application. IEEE Access 8(1):82981–82988
41. Song Z, Chen A, Zhang J (2020) Terahertz switching between broadband absorption and narrowband absorption. Opt Express 28(2):2037–2044
42. Liu H, Ren G, Gao Y, Zhu B, Wu B, Li H, Jian S (2016) Tunable terahertz plasmonic perfect absorber based on T-shaped InSb array. Plasmonics 11(2):411–417
43. Cole MA, Powell DA, Shadrivov IV (2016) Strong terahertz absorption in all-dielectric Huygens' metasurfaces. Nanotechnology 27:424003
44. Liu X, Fan K, Shadrivov IV, Padilla WJ (2017) Experimental realization of a terahertz all-dielectric metasurface absorber. Opt Express 25(1):191–201
45. Ming X, Liu X, Sun L, Padilla WJ Degenerate critical coupling in all-dielectric metasurface absorbers. Opt Express 25(Issue 20):24658–24669 (2017)
46. Oszwalldowki M, Zimpel M (1988) Temperature dependence of intrinsic carrier concentration and density of states effective mass of heavy holes in InSb. J Phys Chem Solid 49:1179–1185
47. Zhu J, Han J, Tian Z, Gu J, Chen Z, Zhang W (2011) Thermal broadband tunable terahertz metamaterials. Opt Commun 284:3129–3133
48. Li W, Kuang D, Fan F, Chang S, Lin L (2012) Subwavelength B-shaped metallic hole array terahertz filter with InSb bar as thermally tunable structure. Appl Opt 51:7098–7102
49. Hien NT, Le LN, Trang PT, Tung BS, Viet ND, Duyen PT, Thang NM, Viet DT, Lee YP, Lam VD, Tung NT (2015) Characterizations of a thermo-tunable broadband fishnet metamaterial at THz frequencies. Comput Mater Sci 103:189–193
50. Iyer PP, Pendharkar M, Palmstrom CJ, Schuller JA (2017) Ultrawide thermal freecarrier tuning of dielectric antennas coupled to epsilon-near-zero substrates. Nat Commun 8:472
51. Mu Q, Fan F, Chen S, Xu S, Xiong C, Zhang X, Wang X, Chang S (2019) Tunablemagneto-optical polarization device for terahertz waves based on InSb and its plasmonic structure. Photon. Res. 7:325–331
52. Zhou H, Cheng Y, Hu D, Li D, Hui X, Zhang F, Chen M, Mu X (2019) Terahertz biosensing based on bi-layer metamaterial absorbers toward ultra-high sensitivity and simple fabrication. Appl Phys Lett 115:143507
53. Yang D, Zhang C, Li X, Lan C (2019) InSb-enhanced thermally tunable terahertz silicon metasurfaces. IEEE Access 7:95087–95093

54. Aslinezhad M (2020) High sensitivity refractive index and temperature sensor based on semiconductor metamaterial perfect absorber in the terahertz band. Opt Commun 463:125411
55. Smith DR, Vier DC, Koschny T, Soukoulis CM (2005) Electromagnetic parameter retrieval from inhomogeneous metamaterials. Phys Rev E 71:036617
56. Carver KR, Mink JW (1981) Microstrip antenna technology. IEEE Trans Antenn Propag 29:2–24

Optimizing and Validating Performance of 40 Gbps Optical System

Shradha Gupta, Shilpa Choudhary, Kanojia Sindhuben Babulal, and Sanjeev Sharma

Abstract The purpose of this paper is to optimize the performance of the RZ modulated 40 Gbps optical system by using RSM optimizer for pre- and post-compensation of dispersion. There are many input factors whose variation causes improvement of optical signals such as laser power, RZ duty cycle and EDFA gain (pre-and post-compensation), which were optimized to determine their effects on the signal characteristics. The output signal quality was characterized using noise power and *Q*-value. Using OPTSIM simulator a 40 Gbps optical network design and run over the 31 combinations as defined by RSM algorithm. These provide the optimized parameters. These optimized input parameters are again validated by running the optical system on these values. The pre-position of DCF shows the best performance with a percentage error of less than 5% for *Q*-value and 0.3% for noise power. This compensation technique is applied for RZ modulated system and validate by *Q*-value and noise power with percentage error. The optimized results matched with the simulated result and thus validate.

Keywords Dispersion compensation · Optimization · Pre- and post-compensation · *Q*-value · Noise power · Validation · Dispersion compensating fiber · RZ modulation

S. Gupta (✉) · S. Sharma
Department of Applied Science and Humanities, G.L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: shradhamay@gmail.com

S. Choudhary
Department of Electronics and Communication Engineering, G.L. Bajaj Institute of Technology and Management, Greater Noida, India

K. S. Babulal
Department of Computer Science & Technology, Central University of Jharkhand, Ranchi, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_10

1 Introduction

With the increase in demand for higher bitrate, the number of users who want more access to connectivity is also increasing extremely large. Many kinds of research are working for the best option either with different modulation for free-space-optical communication [1] or with photonics communications through DWDM system [2–8]. In both cases, optical communication comes as better communication technology. To meet these requirements optical fibers with high bit rate transmission are meeting these requirements. Commercially to achieve these, many optical fibers with attractive features and extremely larger bandwidth are available [9].

Pulse propagation through optical fiber was studied by solving the nonlinear Schrodinger equation in the presence of nonlinear effects like SPM, GVD, SRS and Self steepening effect by both the scalar and vector approach while considering without and with the polarization concepts, respectively [10]. Nonlinear effects like self-phase modulations (SPM), Stimulated Raman Scattering (SRS) and self-steepening effects were considered while solving the equation. The propagation characteristic of the pulses in the fiber media is analyzed and verified by simulation. These linear and nonlinear mainly affect the system performance and degrade the received signal. These factors cause the temporal and spatial broadening of signal through fiber and cause an error while receiving at the receiver ends. Thus, these are the key factors that decide the optimizing parameters [11].

The transmission degradation while traveling through fiber causes dispersion, which can be managed by different types of single-mode fiber and other nonlinear passive optical devices like FBG [12, 13]. A common method for managing dispersion is to combine two or more types of single-mode fiber to produce/compensate for the desired dispersion over the entire span length. Managing the total chromatic dispersion slightly above zero helps to avoid the nonlinear four-wave mixing effect [10]. However, the negative dispersion can be achieved by using dispersion compensating fiber (DCF) or fiber Bragg gratings. Work with DCF for dispersion compensation was also reported for the WDM system [14]. For which they optimized the 5 channel WDM by using power input before DCF by controlling the gain of amplifier for three positions of DCF. For three different setups, output was recorded in terms of Q-value and eye diagram for different lengths of DCF. They have done an exhaustive approach to explore the optimum dispersion map. In a multichannel optical system, the inter-channel nonlinear interference noise comes into play from two, three and four-pulse collisions. Their effect was studied and classified as pulse collisions either complete or incomplete [15]. Ahmed et al. compensated the dispersion using the post position of Dispersion Compensating fiber (DCF) across the single-mode fiber by [16] and electronically for more than 40 Gbps WDM system by RZ and CSRZ modulation formats. Fiber Bragg grating is used for compensating dispersion by analysis of group delay response and reflection spectrum [17].

For 10 Gbps the dispersion compensation using different positions of DCF was done [18]. Performance optimization with duty cycle selection for RZ-DQOSK

modulation for 10 Gbps optical system is done for different fibers. They compensated the dispersion by post placing of Dispersion Compensating fiber in combination with three different types of single-mode fibers. They come with the best-suited fiber with a suitable duty cycle for long-distance optical systems for a 10 Gbps data rate. A similar analysis was done for the DWDM system using five different fibers to mitigate the effect Four-Wave Mixing (FWM) and Stimulate Brillouin Scattering (SBS) nonlinear effect in WDM optical communication system [19]. NRZ, RZ and Duobinary modulation formats were compared for the WDM system of 10 and 40 Gbps bitrate for the DWDM system [20]. The specialty of modulation formats for passive optical network systems was done by Li et al. for WDM [21]. Authors in [21] also concluded that CS-RZ-DPSK) shows better performance against the non-linear effect.

Fiber Bragg Grating has been used by Shabaneh for compensation in optical transmission links [22]. Symmetric Bragg Grating was placed and analyzed concerning *Q*-factor and noise power. They found that grating length is proportional to signal power. Similar work is also done [23] with and without a low pass Gaussian filter [24, 25]. Many such types of performance analyses have been done using an optical simulator. Thus, in this paper, a novel work is done with the addition/use of optimization tools for optimizing the optical system performance. Such type of work has not been done before for optical links.

To optimize the performance of the 40 Gbps optical network, the work starts with optimizing the parameters of single-channel components. In this paper, the optimized value of duty cycle of RZ pulse, laser power, Erbium-doped fiber amplifier (EDFA) gain by pre- and post-compensating dispersion due to single-mode fiber was first to find out then again validate by a simulated optical system.

2 Methodology

In this paper, to optimize the optical system an extensive work is done and verified in three steps. First, a dispersion-compensated 40 Gbps optical system by both pre- and post-compensation is analyzed for different combinations of (four) test parameters. Second with the help of Response Surface Methodology (RSM) optimization of system output was carried out and corresponding input parameters were found out. And finally, the optimized output parameters were verified by simulation. During the final simulation, the input values were fixed at optimized input parameters to obtain results. The output simulated results verified with the output optimized results and validates by eye diagram and percentage error.

2.1 Optical System

Simulation setup for 40 Gbps optical system used in this paper is shown in Fig. 1. The four input parameters which has been considered in this paper are duty cycle of RZ pulse, laser power, Gain of Erbium-Doped Fiber Amplifier (EDFA) for Single Mode Fiber (SMF) and gain of EDFA for Dispersion Compensating Fiber (DCF).

The simulation run on Optiwave–Optisystem. RZ pulse generator with duty cycle varies from 0.1 to 0.9 is modulated with a laser pulse of central frequency at 1550 nm whose laser power also varied from 1 to 10 mW. The output of the modulator is feed to the optical fiber with a loop contains DCF of 12 km chosen to compensate the 100% dispersion caused by SMF, i.e., the standard length of 60 km. The DCF is placed before SMF for pre-compensation and after SMF for post-compensation. The total span run is 8, i.e., total length of the fiber link is 576 km. The loss (attenuation) caused by fiber is compensated by placing an amplifier of variable gain. The EDFA gain of SMF varied from 7 to 15 dB and EDFA gain for DCF varies from 5 to 9 dB. The total combination of 4 input parameters was 31 as it is defined for RSM Logic.

Secondly, based on the table II data, the optimization is done using the RSM logic with the result (Figs. 2 and 3). The red marked upper horizontal values are optimized input parameters. The vertical blue values are output values using RSM logic in the same figures. In brief, Figs. 2 and 3 provide the optimized input parameters and output values using RSM logic.

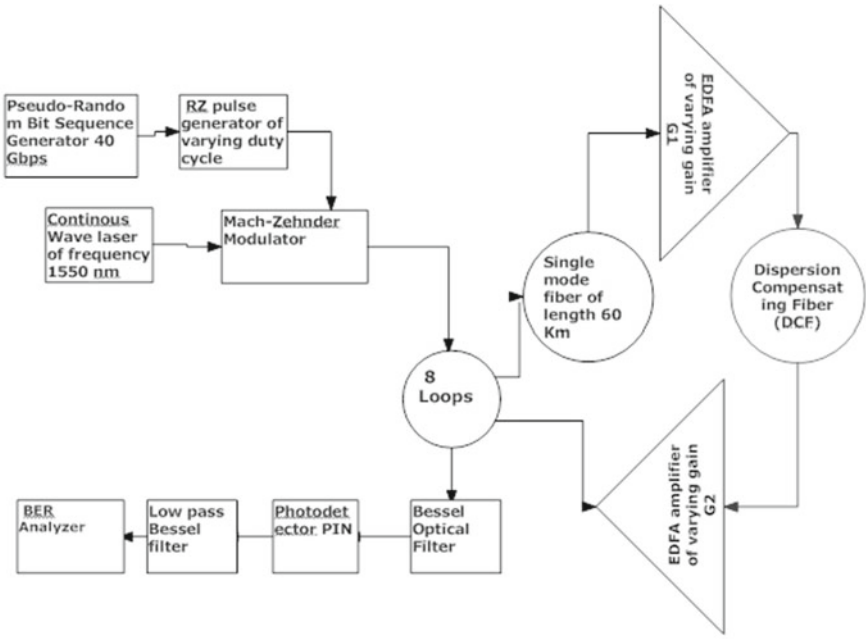


Fig. 1 Block diagram of simulated optical system

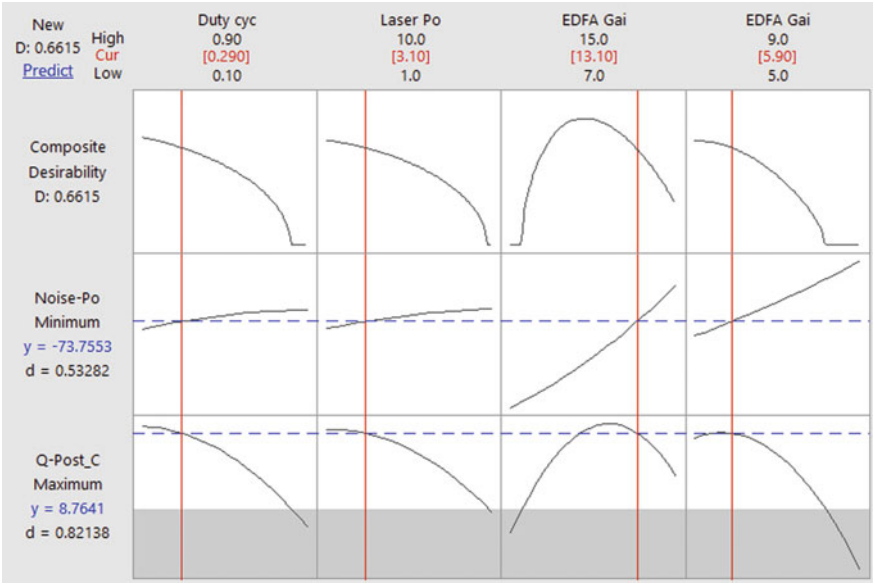


Fig. 2 Optimization plot: post-compensation results

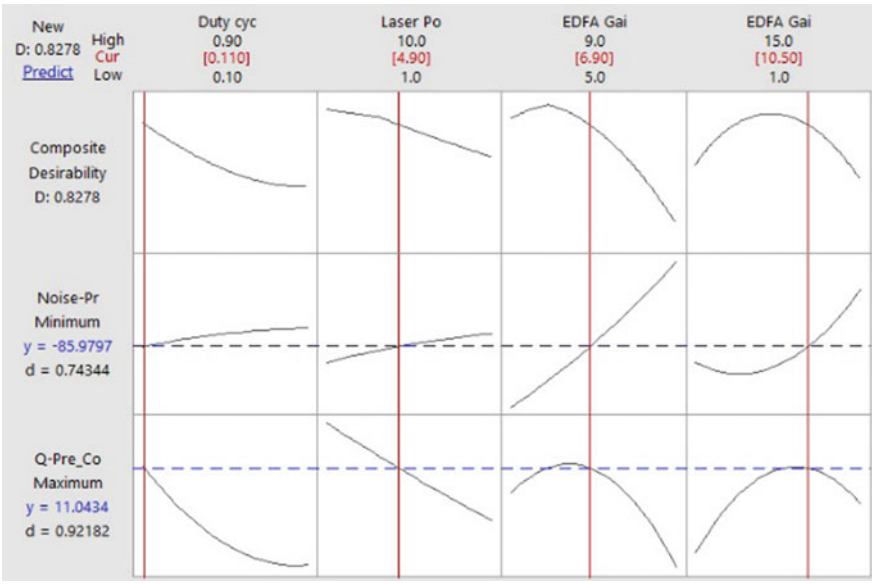


Fig. 3 Optimization plot: pre-compensation results

Lastly, the validation is done by again using the optical simulator whose output is obtained for optimized input values. The input values of the optical simulator are fixed at the input parameters as decided by RSM logic. For validation the output values obtain from the optical simulator are compare with RSM output values.

3 Results and Discussion

The Coding level and range of all the input variables is mentioned in Table 1. The simulation results (output parameters are *Q*-value and Noise power) for all the 31 combinations of input variables are shown in Table 2.

3.1 RSM Optimizer

With the help of RSM logic, optimization of system parameters was carried out. Optimized results for post- and pre- compensation are as shown in Figs. 2 and 3 for post-compensation the optimized values for noise power is −73.7553 dBm and *Q*-value is 8.7641 at input values of 0.290, 3.10, 13.10 and 5.90 for Duty cycle, laser power, EDFA Gain DCF and EDFA SMF Simultaneously (Fig. 2). Similarly, for pre-compensation the optimized results for *Q*-value and Noise power are 11.0434 and −85.9797 dBm simultaneously as shown in Fig. 3. These optimized values are obtained at the input values of 0.110, 4.90, 6.90 and 10.

3.2 Validation

There Now the optical simulator run for the values given in Table 3. These are optimized input and output parameters were obtained.

Validation of RSM optimize results was carried out with the help of OptiSystem–Optiwave and the percentage error was found to be 5.0 and 0.28 for *Q*-value and noise power, respectively, for pre-compensation. For post-compensation percentage

Table 1 Coding level of input parameters

Factors or input parameters	Symbol	Coded levels				
		−2	−1	0	1	2
Input 1	Duty cycle	0.1	0.3	0.5	0.7	0.9
Input 2	Laser power (mW)	1	3	5	7	10
Input 3	EDFA Gain1 SMF (dB)	7	9	11	13	15
Input 4	EDFA Gain2 DCF (dB)	5	6	7	8	9

Table 2 Simulation results for 31 coded levels for pre-compensation and post-compensation

Exp. run	Duty cycle Input 1	Laser power Input 2 (mW)	EDFA Gain1 SMF Input 3 (dB)	EDFA Gain2 DCF Input 4 (dB)	*Q*-value post-comp	*Q*-value pre-comp	Noise power post-comp (dBm)	Noise power pre-comp (dBm)
					O/P 1	O/P 2	O/P 3	O/P 4
1	0.7	7	13	8	0	0	−50.45	−49.59
2	0.5	5	11	7	8.872	4.478	−76.786	−75.97
3	0.5	5	11	7	8.872	4.478	−76.786	−75.97
4	0.5	5	15	7	0	0	−44.77	−45.092
5	0.7	3	13	6	0	0	−69.26	−68.97
6	0.3	3	13	6	8.623	7.309	−73.84	−72.58
7	0.7	7	13	6	0	0	−64.98	−65.54
8	0.3	3	9	8	6.91	10.183	−88.63	−88.18
9	0.3	3	13	6	8.684	7.31	−72.99	−72.58
10	0.5	5	11	5	6.574	6.9	−90.76	−90.98
11	0.5	5	11	7	8.872	4.478	−76.786	−75.97
12	0.5	5	7	7	0	0	−99.17	−100.09
13	0.3	7	9	8	10.67	10.02	−85.34	−84.79
14	0.5	5	11	7	8.872	4.478	−76.786	−75.97
15	0.5	5	11	9	0	0	−59.089	−60.43
16	0.7	7	9	6	4.216	4.489	−94.74	−95.21
17	0.3	7	13	8	0	0	−53.21	−53.98
18	0.1	5	11	7	8.589	11.98	−82.95	−81.05
19	0.3	7	13	6	3.289	3.027	−69.36	−69.002
20	0.5	1	11	7	8.684	10.181	−81.84	−82.55
21	0.5	10	11	7	4.198	0	−73.78	−73.15
22	0.3	3	13	8	0	0	−57.005	−56.98
23	0.5	5	11	7	8.872	4.478	−76.786	−75.97
24	0.5	5	11	7	8.872	4.478	−76.786	−75.97
25	0.7	3	9	8	9.573	7.031	−84.52	−84.54
26	0.7	7	9	8	8.64	2.923	−81.159	−81.11
27	0.7	3	13	8	0	0	−52.85	−53.37
28	0.3	7	9	6	2.75	3.535	−97.44	−97.35
29	0.5	5	11	7	8.872	4.478	−76.786	−75.97
30	0.7	3	9	6	2.568	3.097	−96.781	−97.27
31	0.9	5	11	7	4.0523	2.345	−74.24	−74.66

Table 3 Comparison of results between simulated and optimized results

	Duty cycle	Laser power	EDFA DCF gain	EDFA SMF gain	*Q*-value simulation	*Q*-value optimized (RSM)	Noise power simulation	Noise power optimize	% Error *Q*-value	%Error Noise power
Pre-compensation	0.110	4.9	6.90	10.50	10.5128	11.0434	−85.74	−85.9797	5.05%	0.28%
Post-compensation	0.290	3.10	5.9	13.10	8.44852	8.7641	−73.82	−73.76	3.74%	0.08%

error was found to be 3.74 and 0.08 for *Q*-value and noise power, respectively. The percentage error in both cases was found to be within the tolerable limit.

The *Q*-value and noise power for pre-compensation are 10.51 and −85.74, respectively. Whereas for post-compensation the simultaneous values are 8.45 and −73.82. The pre-compensation is showing better performance. The optimized parameters for post-compensation have low laser power and comparatively high laser power as compare to pre-compensation. For post-compensation, this low laser power is the reason for poor performance in terms of noise factor and *Q*-value w.r.t pre-compensation. As the low laser power is not enough to avoid Amplified Spontaneous Emission (ASE). The spontaneous emission is amplified by gain and increases the noise. In the case of two or more gain, the overall performance is decided by high gain at the input stage and thus affecting its performance in post-compensation. For 40 Gbps optical system pre-compensation is performing better to deal with dispersion degrading effect while transmission for more than 700 km optical link.

4 Conclusion

The performance of optical communication is depending on many input parameters and optimization is a better technique to increase its efficiency by deciding the input parameters. Through this paper, we have shown and validate that the input parameter can be optimized to compensate for the dispersion effect and improve the output data. For dispersion compensation, the placement of Dispersion Compensating fiber before the standard fiber is the best combination to effectively compensate the dispersion.

References

1. Jeyarani J, Sriramkumar D, Caroline BE (2018) Performance analysis of free space optical communication system employing WDM-PolSK under turbulent weather conditions. J Optoelectron Adv Mater 20(9–10):506–514
2. Zhou Q, Ahmed H (2018) Optical solitons in DWDM system with modified simple equation method. J Optoelectron Adv Mater 20(1–2):46–51
3. Sharma SZ (2021) Design of a tunable 8-channel DWDM covers all transmission windows using defect mode nonlinear photonic crystals. Opt Eng 60(1):017105
4. Sharma S, Kumar V, Gupta S (2021) Tunable transmittance using temperature dependence ZnS-based ID photonic crystals. In: Advances in smart communication and imaging systems, Springer, pp 325–330
5. Sharma S, Kumar A, Singh KhS (2020) Design of a tunable DWDM multiplexer using four defect layers of GaAs nonlinear photonic crystals. Optik—Int J Light Electron Opt 212:164652
6. Sharma S, Kumar R, Singh KhS, Kumar V, Kumar A (2014) Single channel tunable wavelength division demultiplexer using one dimensional defect mode nonlinear photonic crystal. Optik—Int J Light Electron Opt 125:4895–4897
7. Sharma S, Gupta S, Suther B, Singh KhS (2020) Design of a tunable transmission mode filter using 1D Ge based nonlinear photonic crystal. AIP Conf Proc 2220:050016. https://doi.org/10.1063/5.0001128

8. Sharma S, Kumar R, Singh KS, Kumar V, Kumar A (2014) Performance analysis of dense wavelength division demultiplexer based on one dimensional defect mode nonlinear photonic crystal. J Optoelectron Adv Mater 16:1015–1019
9. Nishimura M (2005) Optical fibers and fiber dispersion compensators for high-speed optical communication. J Opt Fiber Commun Rep 2:115–139. https://doi.org/10.1007/s10297-004-0024-y
10. Ramprasad V, Meenakshi M (2006) A study on the propagation characteristics of pulses in optical fiber communication systems. In: 2006 IEEE Workshop on signal propagation on interconnects, Berlin, pp 263–266. https://doi.org/10.1109/SPI.2006.289240
11. Ellis AD, McCarthy ME, Al-Khateeb MAJ, Sorokina M, Doran NJ (2017) Performance limits in optical communications due to fiber nonlinearity. Adv Opt Photonics 9(3):429–503
12. Tiwari U, Saini TS, Saini S, Bharadwaj M (2018) Performance evaluation of the UDWDM-PON with 10 Gbps bit rate using FBG based dispersion compensation. J Optoelectron Adv Mater 20(3–4):113–117
13. Sharma S, Kumar A (2021) Analysis of silica based single-mode fiber doped with germanium at different transmission window. SILICON. https://doi.org/10.1007/s12633-020-00884-6
14. Hamza MY, Hayee I (2016) Performance improvement of 40 Gb/s WDM systems by optimization of dispersion map. In: 2016 IEEE 10th International conference on application of information and communication technologies (AICT), Baku (2016), pp 1–5. https://doi.org/10.1109/ICAICT.2016.7991755
15. Dar R, Feder M, Mecozzi A, Shtaif M (2016) Pulse collision picture of inter-channel nonlinear interference in fiber-optic communications. J Lightwave Technol 34(2):593–607. https://doi.org/10.1109/JLT.2015.2428283
16. Ahmed N, Hayee MI, Zhang Q (2010) Electronic post-compensation of fiber nonlinearity for 40 Gbit/s WDM systems. IEEE/OSA J Opt Commun Networking 2(7):456–462. https://doi.org/10.1364/JOCN.2.000456
17. Dar AB, Jha RK (2017) Chromatic dispersion compensation techniques and characterization of fiber Bragg grating for dispersion compensation. Opt Quant Electron 49:108. https://doi.org/10.1007/s11082-017-0944-4
18. Hu B, Jing W, Wei W, Zhao R (2010) Analysis on dispersion compensation with DCF based on Optisystem. In: 2nd International conference on industrial and information systems, Dalian, pp 40–43. https://doi.org/10.1109/INDUSIS.2010.5565685
19. Rachnarong V, Noppanakeepong S (2012) A study on the effect of four wave mixing and stimulated Brillouin Scattering of DWDM transmission system in optical fiber. Eng Appl Sci Res 36(4):261–271
20. Bobrovs V, Ivanovs G (2008) Comparison of different modulation formats and their compatibility with WDM transmission system. Latv J Phys Tech Sci 2:3–16. https://doi.org/10.2478/v10047-008-0005-7
21. Li L, Zhang J, Yin A (2011) Analysis the speciality of modulation format in WDM-PON system. J Comput 6(12):2534–2540. https://doi.org/10.4304/jcp.6.12.2534-2540
22. Shabaneh AA (2020) Investigative modeling of symmetric Fiber Bragg Grating as dispersion compensation for optical transmission system. Optica pura aplicada 53(4)
23. Dahir A, Yu Z (2020) Dispersion compensation by using FBG and low pass Gaussian filter. In: 2020 5th International conference on computer and communication systems (ICCCS), Shanghai, China, 2020, pp 803–806. https://doi.org/10.1109/ICCCS49078.2020.9118555
24. Mishra R, Bharti R (2017) An analysis of 10 Gbits/S optical transmission system using Fiber Bragg Grating (FBG) and double EDFA. Int J Adv Res Comput Sci 8(8)
25. Mohammadi SO, Mozaffari S, Mahdi Shahidi M (2011) Simulation of a transmission system to compensate dispersion in an optical fiber by chirp gratings. Int J Phys Sci 6(32):7354–7360

Medical Image Analysis Using Soft Computing Feature Selection and Classification of Skin Cancer

Birendra Kumar Saraswat, Shipra Srivastava, Samender Singh, Arun Kumar Takuli, and Aditya Saxena

Abstract Adaptive resonance theory ARTI neural network (unsupervised learning) is being implemented for the classification of feature values with the objective of carrying out a comparative study with the performance of other technique including multi-layer feed forward network transform (supervised learning). A neural network is applied for the detection/identification of melanoma skin cancer. The images were collected from melanoma database is first processed and some distinguishable features were extracted by a MATLAB GUI. Two classes of the images were used: (a) benign, (b) malignant. A collection of interpreted binary input data, which is digitized from analog data after normalization is, then clustered adaptively with neural network ART-1. The goal of the project is to classify the feature dataset using supervised and unsupervised neural networks. To collect the feature set 1 implemented MATLAB image processing toolbox to create my own MATLAB functions. I created my own ARTI codes in MATLAB 701 (AKTI) to analysis which features are giving better efficiency using different vigilance parameter. Along with these supervised and unsupervised classification and pattern recognition, I am also trying to classify my own data-set of extracted features on Weka machine learning benchmarking software which is nothing but implements machine learning algorithms in JAVA.

Keywords Adaptive resonance theory · Unsupervised learning · Supervised learning · Weka · Texture feature

B. K. Saraswat (✉) · A. Saxena
Department of Computer Science and Engineering, G.L.A. University, Mathura, India
e-mail: birendrasaraswat@gmail.com

S. Srivastava
Department of Information Technology, Greater Noida Institute of Technology, Greater Noida, India

S. Singh
Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India

A. K. Takuli
Department of Computer Science and Engineering, Hi-Tech Institute of Engineering and Technology, Ghaziabad, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_11

1 Introduction

This is a paper developed on medical image processing and classification of skin cancer. It is developed in MATLAB 7.0.1, and the classifier used is neural network. The algorithm used for neural network is adaptive resonance theory (ART1) [1]. Over the past decade, in biology and medicine, medical imaging is better than the most other methods of visualization and interpretation [2]. System designing, implementing, and validating of difficult medical applications requires strong collaborative work between physicians and engineers because inadequate image quality contribute to problem of collecting, evaluating and defining medical device features [3]. Moreover, most of the today's research work on enhancing imperfect picture quality is equipped. In this project, I need the images of skin cancer from the image scanner. The input image will be first preprocessed and enhanced [4]; this image will be segmented from the background. Certain useful features will be extracted from the enhanced image, which are the prime requirements for pattern matching [5–7]. After then the classification of the image will be done using soft computing where a decision will be made whether the input image belongs to certain category [8]. Comparing the input image with certain stored image patterns that are already diagnosed will do this. In my current project, I have studied the segmentation process and some useful features to be extracted for the pattern recognition from previous works of other researchers [9–12]. I used texture models for the skin cancer cell images appearance within the affected area [13].

1.1 Adaptive Resonance Theory for Pattern Recognition

I used adaptive resonance theory (ART) from some review works by L. G. Heins and D. R. Tauritz in May/June 1995 introducing motivation and ART mechanism [14]. An independent learning model is the fundamental ART form. It generally comprises of a comparative field and a neuron field, a vigilance parameter and a reset module [15]. Individual learning is a machine learning method where a model makes observations. The fact that there is no a priori production distinguishes it from the administered learning. A data collection of input items is gathered in individual learning [16]. Individual learning usually handles objects that are input as a series of random variables. A model of joint density is then constructed for the data set. Adaptive resonance theory (ART) is also clustering style among other styles, and may cause the number of clusters to differ with size problem [17]. The key distinction between ART and other clustering approach is that ART allows the user to monitor the similarities among groups of the same clusters by means user-defined variables called the vigilance parameter. Theory of adaptive resonance provides a neural architecture class that combines the pattern space and input evaluation of vector weight templates. The consistency problem of plasticity was not solved by traditional artificial neural networks. ART nets and algorithms retain the requisite plasticity to learn new patterns

and avoid the pattern changes previously appreciated. The network of ART solves stability—plasticity dilemma, is a classifier of vector. For several pattern recognition activities, ART networks are often used such as automated target recognition and seismic signal processing. Carpenter and Grossberg developed the first version of ART which was known as "ART1". It is the first ART family member. It is a network of two-layers discovering the pattern clusters in arbitrary Boolean pattern sets.

2 Design of the Project

2.1 Collection of Melanoma Skin Cancer Images

Sample images of 20 affected patients were taken and trained into the network. First, the image samples were taken and certain features were extracted from them based on their texture and boundaries. Then, these values were taken in a dataset and trained with the neural network. The next section deals with the parameters that were extracted.

2.2 Image Processing

Boundary extraction: The boundary was extracted using canny edge detector algorithm. The algorithm returns a binary image with the edges and the boundary detected.

After this, the image is converted to a label matrix using the bwlabel function. Now, the boundary parameters are extracted.

2.3 Region of Interest

Extracting the edge of one image, it is basically dependent on the first order derivative. Laplacian order is not efficient so Gaussian distribution is used to separate the edge and the direction. We cannot find the contour. Region of interest is the one specified region over which the feature elements are extracted. In output image, this gives a thin line.

2.4 Boundary Extraction

The boundary was extracted using canny edge detector algorithm. The algorithm returns a binary image with the edges and the boundary detected.

After this, the image is converted to a label matrix using the bwlabel function. Now the boundary parameters are extracted.

2.5 Region of Interest

Extracting the edge of one image is basically dependent on the first order derivative. Laplacian order is not efficient so Gaussian distribution is used to separate the edge and the direction. We cannot find the contour. Region of interest is the one specified region over which the feature elements are extracted. In output image, this gives a thin line.

2.6 Feature Extraction

Boundary parameters: These are—'Area', 'Major Axis Length', 'Minor Axis Length', 'Eccentricity', Perimeter, Solidity, and Diameter. The boundary parameters' values for the above 14.bmp image of class malignant are as follows, respectively:

452 7.5 352 117 825 12.5 68

2.7 Texture Parameters

I took coarse texture for recognition of a cancer cell. Many images are formed sensing reflected light from a scene, but transmitting light or other radiation forms some images. In diagnosing skin cancers as benign or malignant from X-rays, it is important to search for calcifications, which appear as tiny light spots, especially in malignant skin cancers. They will always produce edges with abrupt changes in gray level at their boundaries. An algorithm first find sharp left edges between two pixels then look for sharp right edges in a surrounding region that is consistent with the possible size range for calcification.

2.8 Texture Analysis

Describing an area means quantifying its material on texture. I prefer to use statistical approach in support for statistical property based texture analysis of the strength histogram. One of those groups of a measure like this is based on statistical moments. Mean, standard deviation, smoothness, third moment, uniformity, and entropy. The texture parameters' values for the said Image belonging to the class malignant are as follows, respectively:

6 96 0.02 6 6 0.2

The texture parameters' values for the 16.bmp Image belonging to the class malignant are as follows, respectively:

8 109 0.03 8 6 0.3

The two classes for classification of the images are given numeric value 2 for benign, and 4 for malignant. The entropy, average contrast, and average intensity are higher for coarse region than the other, because the values of the pixels in that region are more random than other regions.

3 Dataset Preparation for Art 1

A set of data, which were extracted using MATLAB 7.0.1 function, and GUI-based approach are now ready to be prepared as an ASCII dataset. Because adaptive resonance theory ART1 in neural network requires binary input in ASCII form. Here, my original data file is skincancer .txt resided in MATLAB work folder. The normalized data are in can1.txt file in the work folder. This can1.txt will be used for binary conversion in ART1.m program.

3.1 Data Description

Sources: The Melanoma skin cancer database was obtained from Image and Vision Computing, vol. 17, no. 1, 1999, 65–74. by L. Xu, M. Jackowski, A. Goshtasby, C. Yu, D. Roseman, S. Bines, A. Dhawan, A. Huntley.

Table 1 Attribute Information

S. No.	Attribute information	Domain
1	Area	0.1–0.5
2	Eccentricity	0.0005–0.01
3	Major axis	0.07–0.35
4	Minor axis	0.03–0.15
5	Mean	0.002–0.01
6	Standard deviation	0.047–0.1
7	Smoothness	0.000
8	Third moment	0.002–0.01
9	Uniformity	0.0037
10	Entropy	0001–0.0004
11	Perimeter	0.31–1.00
12	Solidity	0.0001–0.02
13	Diameter	0.01–0.09

3.2 Information

- Number of Instance: 20
- The ten instances are giving wrong attribute values, so they are deleted from the servers, abandoning ten instances
- Attribute number: 13 plus the class value
- Attribute 1 to 13 will be used to represent instances
- There is one of two possible groups in each instance: Benign and malignant
- Class distribution
- Benign: 5 (50%)
- Malignant: 5 (50%).

3.3 Attribute Information

See Table 1.

3.4 Database

The cancer database file has ten instances. Since this database does not have enough instances, 100% accuracy will not be achieved in this case. However, even with this limited database, the decision maker will be able to give an 85% overall accuracy rate. The number of required instances to achieve 100% accuracy and the number of available instances in this case is listed below (Table 2):

Table 2 Cancer database file

Class	Required	Actual
Instance	Instance	
2 = benign	100	5
4 = malignant	100	5

4 Pattern Recognition Using Adaptive Resonance Theory

4.1 Media Learning in Art

Two types of learning in ART, viz.

1. ART1 network uses fast learning. Here, contribution is binary. The main changes during the resonance period occur more rapidly.
2. ART2 network uses slow learning. The resonance period occur slowly during the main changes.

4.2 Art 1

Construction/Architecture

There are computational units and supplemental units in The ART 1 network. The architecture is shown in given Fig. 1.

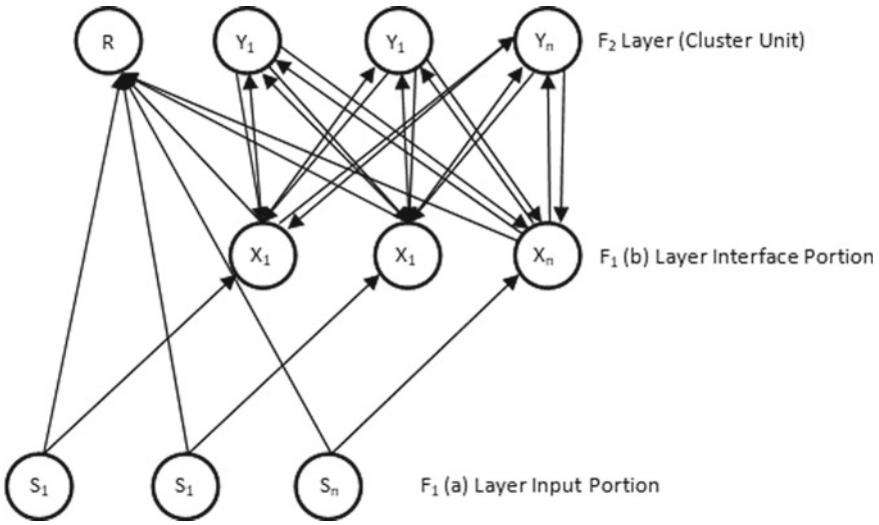


Fig. 1 Construction/Architecture ART 1

In order to avoid the given problems, in addition to the rest unit, two additional units (called gain control units, G_1 and G_2) are present in addition to reset unit, R.

Algorithm
The parameters used in the algorithms are:

N	–	Number of input vector components
M	–	Maximum number of potential clusters
b_{ij}	–	bottom-up weights (from Fi(b) to F2 unit)
t_{ij}	–	top-down weights (from F2 to F; (b) units)
P	–	Vigilance parameter
S	–	Binary input vector
X	–	Activation vector for interface (F1 (b) layer (binary))
$\|\|x\|\|$	–	norm of vector x (sum of the components x_i)

4.3 Description

Integer vector input will then be appended to the F1 (a) and given via the weighted interconnection path with F1 (b) node, sending the activation signal to layer F2. The net entry is determined for each unit of F2. The winner is the unit with the maximum net input, with the $d = 1$. All additional units will be activated as empty. That single winning unit must learn the current input pattern. The signal sent by F2 to F1(a) is referred to as top-down weights by weighted interconnections. Even if the F1 (a) and F2 units get non-null weights, the 'X' units are 'on.'

The vector norm $||x||$ will give the number of components in which the winning unit's top-down weight vector (t_{ij}) and the input vector S are both 'l'. The values are reset for the winning cluster units that are changed depending on the x to S norm ratio ($||x||/||S||$). The entire circle can be replicated until whether a match is identified or the neurons are interrupted. The known ratio Match Ratio in known as ($||x||/||S||$). At the conclusion of each presentation of a sequence, All cluster units are returned to inactive states at the end of each presentation of a series but are available for the further involvement.

The **training algorithm** of ART 1 network is as follows.
Step 1: Initialize parameters
"$L > 1$ and $0 < \rho < 1$
Initialize weights
$0 < b_{ij} < L/(L - 1 + n)$, $t_{ij}(0) = 1$
Step 2: While stopping condition is false, perform steps 3–14
Step 3: For each training input, do steps 4–13
Step 4: Set activation for all F2 units to zero

Set activations of F1 (a) units to input vector s
Step 5: Compute the norm of s:
$\|s\| = \Sigma S_i$
Step 6: Send input signal from F_1 (a) to F_1 (b) layer
$X_i = S_i$.
Step 7: For each F_2 node that is not inhibited
If $y_j \neq -1$, then, $y_j = \Sigma b_{ij} X_i$
Step 8: While reset is true, perform steps 9–12
Step 9: Find J such that $y_J \geq y_j$ for all nodes j
If $y_j = -1$, then all the odds are inhibited and this pattern cannot be clustered
Step 10: Recompute activation x for F_1 (b)
$X_i = S_i t_{ji}$
Step 11: Compute the norm of vector x:
$\|x\| = \Sigma X_i I$
Step 12: Test for reset
If $(\|x\|/\|s\|) < \rho$, then.
$y_j = -1$ (inhibit node J) (and continue executing Step 8 again)
If $(\|x\|/\|s\|) \geq \rho$, then proceed to Step 13
Step 13: Update the weights for node J
$b_{ij\ (\text{new})} = L_{xi}/(L - 1 + \|x\|)$
$t_{ji\ (\text{new})} = X_i$"
Step 14: Test to stop Condition
The units cannot be moved, the weight cannot modify the stop circumstances, and the 220 search periods cannot be used up. If a tie occurs, J should be the smallest index in the winner selection. Even t_{ij} is 0 or 1 when set to nil for the duration of the study, due to a stable learning method it cannot be set to 1 again. The parameters used have the characteristic values as shown below.
Parameter Range Typical value
$LL > 12$
$P\rho \leq 1\ 0 < 0.9$
(vigilance parameter)
$b_{ij}\ 0 < b_{ij}(0) < L/(L - 1 + n)\ 1/(1 + n)$(bottom-up weights)
$t_{ji}\ t_{he}(0) = 11$ (top-down weights).

4.4 Skin Cancer Detection Using ART 1 Network

An ART 1 network is developed to provide the user to monitor the similarity of degree among patterns placed on the same cluster. Every kind of training pattern can be introduced many times, as the net is trained. The first time it is introduced, pattern may be placed in one cluster unit and then placed on another cluster once it is added later. The pattern is not returned to a previous cluster by a stable net. Many networks attain constancy via phasing out by the learning rate because several times the same set of training patterns is provided. The capacity of a system to learn a new pattern at

any state equally well is called plasticity. The ART1 nets are equipped for becoming plastic and stable. Pattern recognition is one of the basic tasks that neural nets can be trained to do. Any input vector belongs to a particular class or group is defined in pattern classification problem. For neural net approach, I believe I have a collection of training patterns that are popular for accurate classification. The transfer function describes membership of the class with a 1 response to a membership of a single class in an effective manner. The 0 result indicates that a variable is not one of the classes.

Data representation scheme (Normalization).

The actual data is being viewed as virtual symbols with values varying from 0 to 1.0. Here, the converting and specified data sets into their virtual equal form. Amplification has the function of labeling the desired set of varying form minimum and maximum input into network sizes.

5 Virtual Conversion

Now, the transformation is based on certain ranges specified for each attribute, the sets of data into a binary form. The study demonstrates how bits and range number per attribute are taken. Complete attributes are 10 (1 class category and 13 numeric features). In the analog the 13 numerical attributes are scaled in the range from 0 to 1. The minimum and the maximum value from the input set are chosen first, and the following formula is used to achieve this scaling. i.e. New value = (current value − min value)/(max value − min value). With the following scaling, the new values obtained after truncation will become binary. Grouping is performed on a range basis, the values from the range of 0–0.05 are allocated to "O', and values from range of 0.05–1.0 are allocated to '1'. Such attributes can be fed in for training and testing to the neural network.

6 Simulation and Results of Art1

See Table 3, 4 and 5.

6.1 Using Weka 3 for Classification and Prediction

6.1.1 Preparing a File for Test

I copied the examples of training attribute.

Table 3 Vigilance parameter = 0.4(epoch = 100)

Training vector %	Testing vector %	Efficiency
10	90	75
20	80	71.4286
30	70	66.6667
40	60	60
50	50	50
60	40	50
70	30	33.3333

Table 4 Vigilance parameter = 0.6 (epoch = 100)

Training vector %	Testing vector %	Efficiency
10	90	100
20	80	100
30	70	100
40	60	100
50	50	100
60	40	100
70	30	100

Table 5 Vigilance parameter = 0.8(epoch = 100)

Training vector %	Testing vector %	Efficiency
10	90	100
20	80	100
30	70	100
40	60	100
50	50	100
60	40	100
70	30	100
80	20	100
90	10	100

ARFF file into a new test ARFF file and included in the test file a proper name for the relation, say @relation skin, and test data after the @data statement. 4. Example: skin .arff This could be a single instance

"@relation skin
@specific field feature
@attribute eccentricity real
@attribute major axis real
@attribute minor axis real

@attribute mean real
@attribute st.dev. real
@attribute smoothness real
@attribute third moment real
@attribute uniformity real
@attribute entropy real
@attribute perimeter real
@attribute solidity real
@attribute diameter real
@attribute class {yes, no}
@data"

897, 15, 576, 250, 22, 183, 0.08, 21, 6, 0.6, 1630, 15.0, 110, 4.

6.1.2 Specifying Test Options

I used the Weka Explorer environment and used preprocess mode to load the training file. Tried skin.a rff first and came to classify mode (by clicking on the classify tab) as shown below: defined the test options (by testing the relevant button):

6.1.3 Training Data

Here, input to the leaner is my raw feature values (melanoma. arff), i.e., training data including class value, i.e., continuous. And, the output is a neuron model.

6.1.4 Test Data

In case of test data, input is that neuron model and the test data whose output is the predicted value.

Accuracy of the learner: (for discrete data) = #correct data*100.

#Total predictions (#instances in the test data).

For real valued class, accuracy of the learner is measured by mean-squared error.

M.S.E. $= \surd$ 1/#instances $\Sigma(y-y')^2$.

E. Running a test

I opted a classifier (the default is ZeroR-the majority predictor). Then, selected naive Bayes and set its parameters.

6.2 Classifying New Data

In order to classify a new example, say {333, 61, 89, TRUE,?} I assume for? (Or set it at random), say no, i.e., 333, 61, 89, TRUE, no. I generated a test file after @data and included the above examples. I used the choice supplied test set and loaded test file and ran the classifier as I wanted and looked at the output window of the classifier. I had chosen naive Bayes, correctly classified instances tell that I guessed was correct (according to naive Bayes). I then saw those results from the matrix confusion also: I classified instance as no (b), with real classification (My guess) no (b). I can also save the results of the test to a file and see "Visualize classifier errors". I have added an additional "prediction action" attribute its value is the prediction computed by the algorithm. I mat save this to a file in ARFF format by clicking on the "Save" button. The contents of the paper are as follows:

"@relation skin predicted
@attribute Instance number numeric
@attribute area real
@attribute eccentricity real
@attribute major axis real
@attribute minor axis real
@attribute mean real
@attribute st.dev. real
@attribute smoothness real
@attribute third moment real
@attribute uniformity real
@attribute entropy real
@attribute perimeter real
@attribute solidity real
@attribute diameter real
@attribute predicted play {yes, no}
@attribute class {yes, no}
@data"

0,897, 15, 576, 250, 22, 183, 0.08, 21, 6, 0.6, 1630, 15.0, 110, 4, no, no

As shown above, this file contains both class attributes-the last original and the projected. In this case, their values are the same, so the error is 0 (or 100% accuracy).

6.3 Output of WEKA Classifier

In this paper, I have chosen naive Bayes classifier, which gave 80% correctly classified instances, I also have selected which feature values were giving the better performance. Weka provides a variety of classifiers to choose. I also tested which class was the predicted class on which the classifier gave the correct result. It also

gave the predicted class for an unknown class. Thus, I can compare my test and training data on different classifiers in WEKA.

7 Future Scope

In the future, the classification based on WEKA benchmarking tool will be extended so that classification on different algorithms will be compared. The software that is developed only classifies skin cancer with limited database of image. It can be further extended to detect cancerous growths in all parts of the body, especially the abdominal organs and the brain with large number of image databases. It can also be extended to work with complicated images and also with 3D images. However, due lack of proper resources this couldn't be done but if possible this software can be upgraded for the betterment of mankind.

8 Conclusion

Selected feature values from WEKA can be standard for use by any dermatologists as it gave 80% correctly classified class value. ART 1 Network simulation aims at classification into a given database of melanoma cancer by a better degree of precision based on specific train-test ratios. Therefore, the data is obtained with a reasonable degree of precision for unsupervised input sequence. Additional improvement in the simulation can be obtained by training more patterns to increase the degree of accuracy. The project is simulated using ART 1 learning and testing algorithm for pattern classification of skin cancer images' and approximately 85% accuracy is obtained. ART 1 can be confidently inferred to the best suited description of network patterns as of today. We have worked with the texture properties of the image and also the different boundary parameters of the image or the affected area. The software developed on small scale is very effective for treatment of patients suffering from skin cancer.

References

1. Carpenter GA, Grossberg S (2003) Adaptive resonance theory. In: Arbib MA (ed) The handbook of brain theory and neural networks, 2nd edn, pp 87–90. MIT Press, Cambridge, MA
2. Grossberg S (1987) Competitive learning: From interactive activation to adaptive resonance. Cogn Sci (Publ) 11:23–63
3. Mausam K et al (2019) multi-objective optimization design of die-sinking electric discharge machine (EDM) machining parameter for CNT-reinforced carbon fibre nanocomposite using grey relational analysis. J Braz Soc Mech Sci Eng 41(8):348

4. Goyal M (2018) Shape, size and phonon scattering effect on the thermal conductivity of nanostructures. Pramana 91(6):87
5. Goyal M, Gupta B (2019) Study of shape, size and temperature-dependent elastic properties of nanomaterials. Mod Phys Lett B 33(26):1950310
6. Carpenter GA, Grossberg S (1987) ART 2: self-organization of stable category recognition codes for analog input patterns. Appl Opt 26(23):4919–4930
7. Goyal M, Gupta B (2019) Analysis of shape, size and structure dependent thermodynamic properties of nanowires. High Temp–High Pressures 48
8. Goyal M, Singh M (2020) Size and shape dependence of optical properties of nanostructures. Appl Phys A 126(3):1–8
9. Carpenter GA, Grossberg S, Rosen DB (1991) ART 2-A: an adaptive resonance algorithm for rapid category learning and recognition. Neural Netw (Publ) 4:493–504
10. Digital Image Processing using Matlab (2005) Gonzales, Woods and Eddins
11. Singh PK, Sharma K (2018) Molecular dynamics simulation of glass transition behaviour of polymer based nanocomposites. J Sci Ind Res 77(10):592–595
12. Kumar A, Sharma K, Dixit AR (2020) A review on the mechanical and thermal properties of graphene and graphene-based polymer nanocomposites: understanding of modelling and MD simulation. Mol Simul 46(2):136–154
13. Chanda B, Dutta Majumder D (2005) Digital image processing and analysis
14. Chapman SJ (2005) Matlab Programming for Engineers. Thomson Learning, 3rd edn
15. Lippman RP (1997) Pattern classification using neural network. IEEE Commun Mag, November 1989, IEEE Transaction
16. Kumar A, Sharma K (2018) AR Dixit A review of the mechanical and thermal properties of graphene and its hybrid polymer nanocomposites for structural applications. J Mater Sci 54(8):5992–6026
17. Sharma K, Shukla M (2014) Three-phase carbon fiber amine functionalized carbon nanotubes epoxy composite: processing, characterisation, and multiscale modeling. J Nanomater (2014)

Comparative Study of Fuzzy, PID, and 2DOF-PID Controller for Liquid Level Control System

Gunjan Taneja, Anjali Jain, and Neelam Verma

Abstract The aim of this paper is to design a single tank liquid level nonlinear control system. For this, various control schemes are described. PID, 2DOF-PID and fuzzy controllers are used and are compared on the basis of different parameters like settling time, transient, rise time, and overshoot. By comparing different parameters, it is observed that fuzzy gives the best result in terms of overshoot and settling time. Thus, fuzzy controllers can be considered as best controller for designing such system.

Keywords 2DOF-PID · Fuzzy control scheme · Proportional integral derivative control scheme · Matlab Software and Simulink environment

1 Introduction

Different kinds of controllers are used in industrial processing. Various industrial tasks require to know exact quantity of liquid being present in tanks for their working. Directly working on interior of tank is difficult task, so for this purpose different controllers are required. One such controller that is used for tracking liquid levels is the liquid level controller. Liquid level systems are very complex systems, and traditional control the system. Controlling is done using a device known as Valve whose focus is obtaining the setpoint level and also reach the new level if required. This valve is connected to the controller in the circuit. With the advancement in technology many new controllers have come up but most of them are obtained using PID controller. The PID controller has the simplest algorithm and involves the use of proportional, integral and derivative for its working.

Fuzzy logic is the most commonly used controller. They are used in those cases where there is some vagueness in human ideas and minds. In this work, fuzzy logic controller is also used to track the liquid level along with PID and 2DOF-PID Controller. Dhiman and Kumar [1] discussed controlling water levels using a different

G. Taneja (✉) · A. Jain · N. Verma
Department of Electrical and Electronics Engineering, Amity University Uttar Pradesh, Noida, India
e-mail: gunjan.taneja1998@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_12

process. Control of liquid level using sensor and PID controller is explained by Abdullah et al. [2]. Generation of steam in nuclear plants using different controllers is demonstrated by Tan [3]. Omijeh et al. [4] worked on designing of liquid level system. Automatic control of liquid level is described by Wagoner [5]. Sliding technique for controlling liquid is described by Moradi et al. [6]. The use of a microcontroller for controlling liquid is explained by Reza et al. [7]. V. Galzina focused on maintaining the level in the boiler using the fuzzy method [8]. The twin tank method is described by Taoyan [9]. Wu et al. [10] focused on integrated system of fuzzy logic controller, neural network & DSP system. Safarzadeh et al. involved the use of feedback theory to study water levels in steam generators. Su et al. [11] worked on the programming of the fuzzy logic controller. Z. Zhi worked on the performance of the steam generator water level [12]. Designing of different tanks was done by Chin et al. [13]. V. K. Tayal et al. worked on swarm optimization method or tracking liquid level system [14]. Ye et al. [15] focused on double tank design for liquid level system. M. Yumurtaki et al. worked on arduino based controllers for liquid level system [16]. Adian [17] describes the PI control scheme for different systems. Patil et al. [18] used Labview software for designing liquid levels. Involvement of sensors for tracking liquid levels is described in [19]. Disha focused on working of fuzzy-based simulation models for liquid level system [20]. King et al. [21] worked on different applications of a fuzzy logic system. Wu Gang et al. [22] discussed the working of two tank liquid level system. Meng et al. [23] focused on observed-based model for two-tank system. BAT algorithm based control of liquid level is discussed in [24]. Different methods of implementing liquid level system are focused in [25].

Nonlinearity of the system is there as height is in terms of square root. The level of liquid can be recorded using the Simulink model. The controller compares the liquid level and gives the required output. This paper is organized as follows: Sect. 2 includes the mathematical model of the system. Various schemes are then described for designing the liquid level system. They are followed by the simulation block diagrams, simulation outputs and results in the further parts. Lastly, discussion and conclusion are done for the paper followed by the references used.

2 Mathematical Modeling

In order to design any system, it is required to understand its mathematical model of such system. Mathematics model gives the required equation that can help in determining system parameters. So in this work, mathematical model for single tank liquid level is described in Fig. 1. Here, Q_{in} is the input, Q_{ot} is the output, t_{a} is the time, A_{o} is the area of the tank, and H_{a} is the height of the tank [4].

Equations for the same are defined as follows:

$$Q_{\text{in}}(t_{\text{a}}) - Q_{\text{ot}}(t_{\text{a}}) = A_{\text{o}} \mathrm{d}H_{\text{a}}(t_{\text{a}})/\mathrm{d}t_{\text{a}} \tag{1}$$

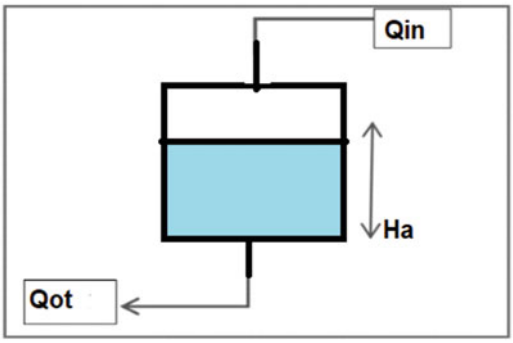


Fig. 1 Liquid level model

The output flow is assumed as

$$Q_{\mathrm{ot}}(t_{\mathrm{a}}) = k\sqrt{H_{\mathrm{a}}}$$

$$Q_{\mathrm{in}}(t_{\mathrm{a}}) - k\sqrt{H_{\mathrm{a}}} = A_{\mathrm{o}}\mathrm{d}H_{\mathrm{a}}(t_{\mathrm{a}})/\mathrm{d}t_{\mathrm{a}} \tag{2}$$

$$Q_{\mathrm{in}}(t_{\mathrm{a}}) = A_{\mathrm{o}}\mathrm{d}H_{\mathrm{a}}(t_{\mathrm{a}})/\mathrm{d}t_{\mathrm{a}} + k\sqrt{H_{\mathrm{a}}} \tag{3}$$

For modeling the system, the transfer function is given by

$$\mathrm{Go(S)} = 0.1533/\mathrm{s\ e} - 0.19\mathrm{s} \tag{4}$$

Following parameters were used to design the water tank in the liquid level control system. Figure 2 describes the values of parameters such as height of the tank in meters, area of the tank in meter square, out pipe cross-section in meter square, and the height of the initial level of liquid in meters.

3 Control Schemes

Various schemes discussed are as follows.

3.1 PID Control Scheme

IT consists of the proportional error signal with derivative along with integral of the error signal.

PID Model for the liquid level system is shown in Fig. 3. The model is made in MATLAB Simulation. Two input sources are given that is control and source flow.

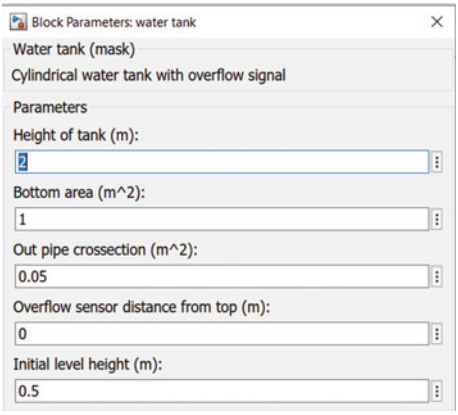


Fig. 2 Parameters used

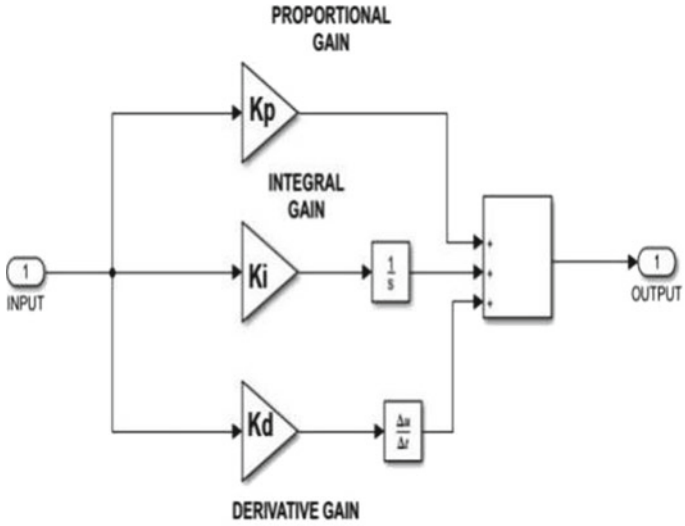


Fig. 3 PID controller

3.2 Fuzzy Design Method

3.2.1 Fuzzy Inference System

Mamdani method is used for this control system. Mamdani method is involved in designing this system as it helps in defining proper output functions and is more flexible in use as compared to other system. Two inputs were considered-LEVEL

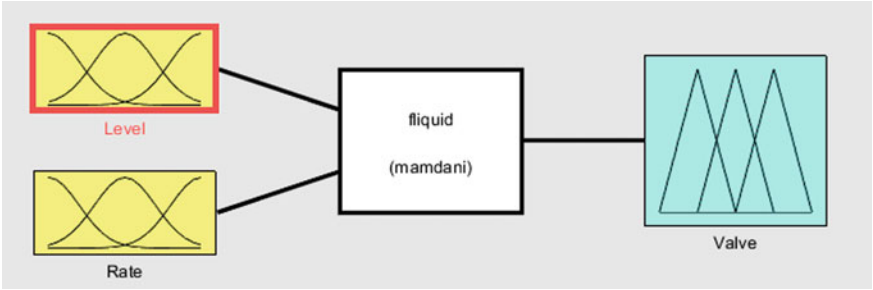


Fig. 4 Mamdani controller

and RATE. According to the rules assigned the controller will work and give output as 'valve' [4] (Fig. 4).

3.2.2 M.F Editor

The Membership Function (M.F) Editor aims to show and allot different membership functions and values for level, range and valve. Triangular membership functions are used in this systems for all three variables (two input variable and one output variable) for reducing complexity and maintaining the uniformity of the system.

Input Variables
Level (−10 to 10)

Figure 5 shows the various membership functions used for describing the first input that is the level of the liquid [−10 to 10]. Table 1 describes the variables, their membership function and range for input variable level.

Rate (−0.5 to 0.5).

Figure 6 shows the various membership functions used for describing the second input that is rate [−0.5 to 0.5]. Table 2 describes the variables, their membership function & range for input variable Rate.

Output Variable
Valve (−5 to 5)

Figure 7 shows the various membership functions used for describing the output that is valve [−5 to 5]. Table 3 describes the variables, their membership function and range for output variable valve.

3.2.3 Rule Editor

Rules for the fuzzy controller are defined considering two inputs, namely, level and rate and one output, namely, valve and their respective variables based on possibility

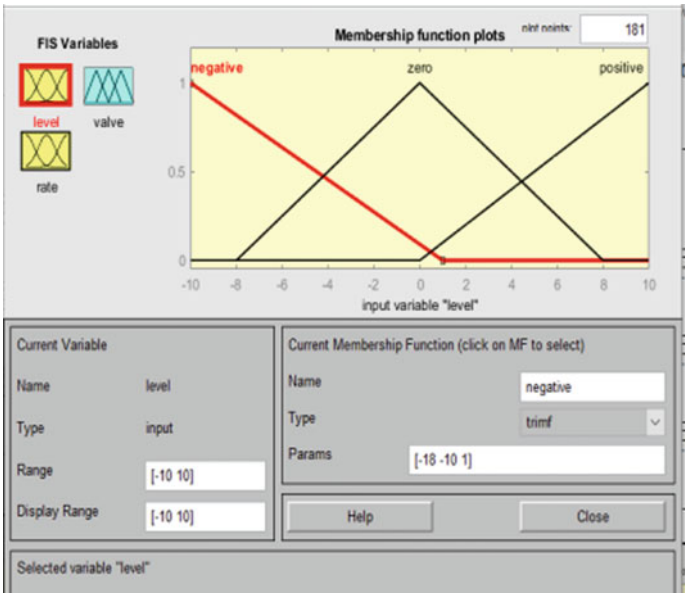


Fig. 5 Liquid Level M.F

Table 1 Liquid Level M.F

Variable used	Membership function	Values
Negative	Triangular	[−9 −5 0]
Zero	Triangular	[−4 0 4]
Positive	Triangular	[0 5 9]

of cases that can be formed. Using different combinations, nine such rules are defined. Figure 8 describes the rule defined that are used to model liquid level control system using the fuzzy toolbox.

3.3 2DOF PID Control Scheme

A PID controller with two degrees of freedom (2DOF) is a controller whose series compensator is a PID element, and the direct compensator is a PD element. It has two inputs and one output. It is used to reduce the effect of disturbance without increasing the value of overshoots in the set point. It is represented by PID2 and gives better performance than the PID controller. The simulation model is shown in Figs. 9 and 15.

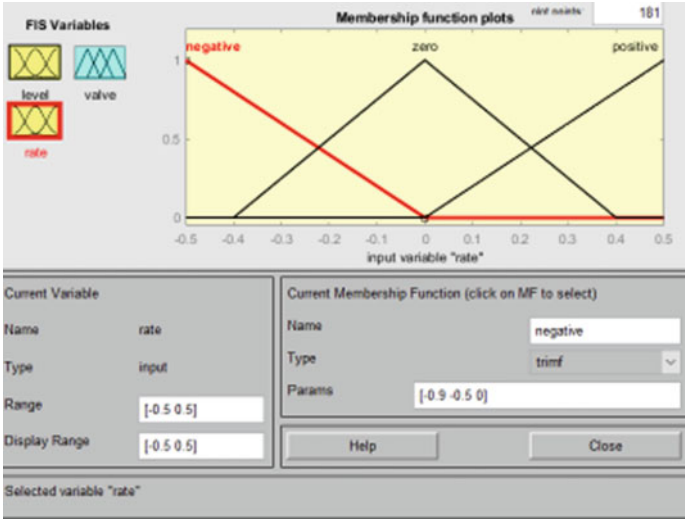


Fig. 6 Rate M.F

Table 2 Rate M.F

Variable used	Membership function	Values
Negative	Triangular	[−18 −10 1]
Zero	Triangular	[−8 0 8]
Positive	Triangular	[0 10 18]

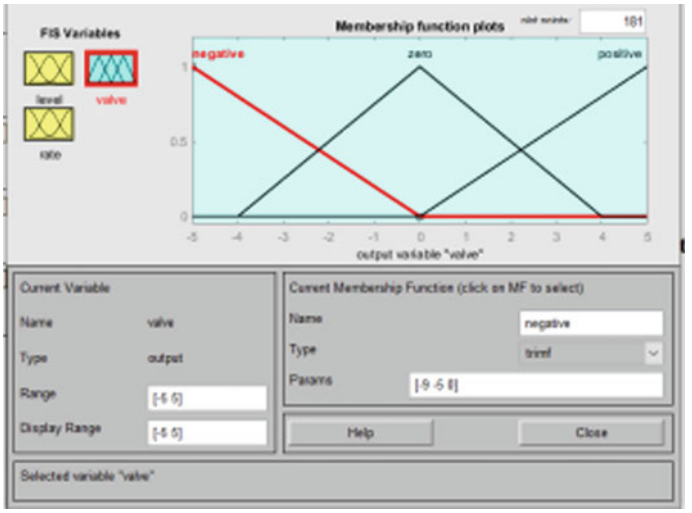


Fig. 7 Valve M.F

Table 3 Valve M.F

Variable used	Membership function	Values
Negative	Triangular	[−0.9 −0.5 0]
Zero	Triangular	[−0.4 0 −0.4]
Positive	Triangular	[0 0.5 0.9]

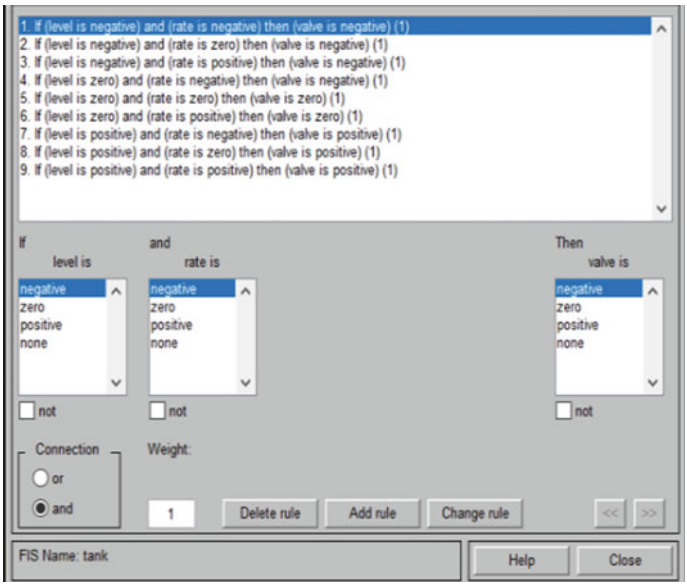


Fig. 8 Editor (rules)

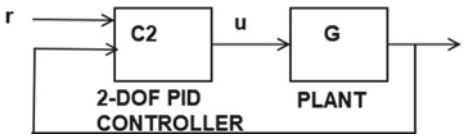


Fig. 9 2 DOF-PID controller

Simulation Model

Figure 10 describes the subsystem of the valve that is used to design the liquid level control system. For PID, 2 DOF PID Controller, fuzzy this subsystem is used.

Figure 11 describes the subsystem of the water tank that is used to design the liquid level control system. For PID, 2 DOF PID Controller, fuzzy this subsystem is used.

Figure 12 gives the model for PID controller where gains are set as $K_p = 3$, $K_i = 2$, and $K_d = 0.2$.

Figure 13 describes the model of the liquid level control system using a PID controller. Here, the subsystem obtained in Figs. 10, 11 and 12 are connected to get

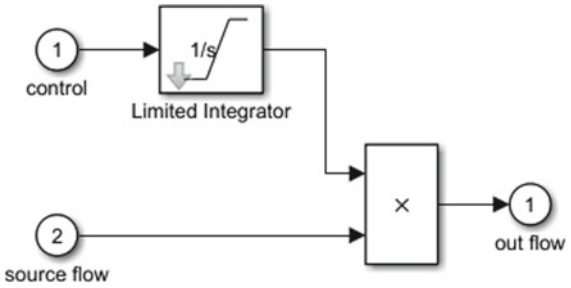


Fig. 10 Valve subsystem

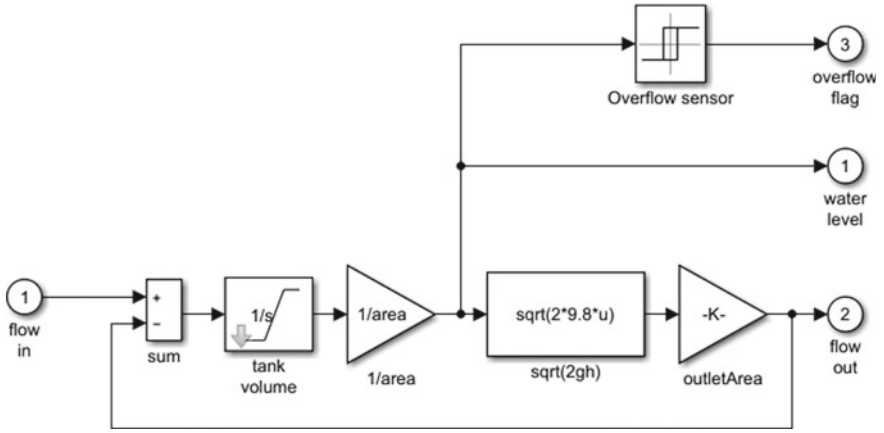


Fig. 11 Water tank subsystem

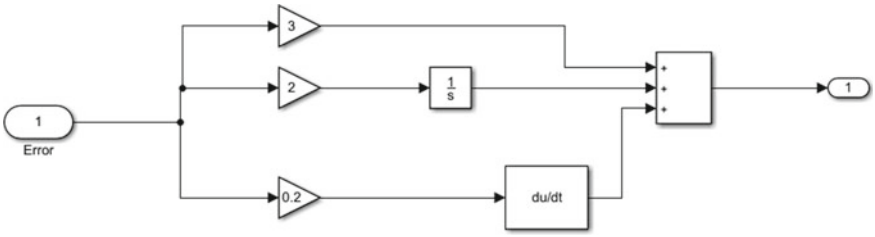


Fig. 12 PID subsystem

the output and with the help of scope, results are observed. Scope 1 gives the value of water level, scope 4 describes the flowout, and 2 describes overflow.

Figure 14 describes the model of liquid level control system using FUZZY controller. Here, the subsystem obtained in Figs. 10 and 11 and rules obtained in Fig. 9 along with above defined membership functions are used to get the output and

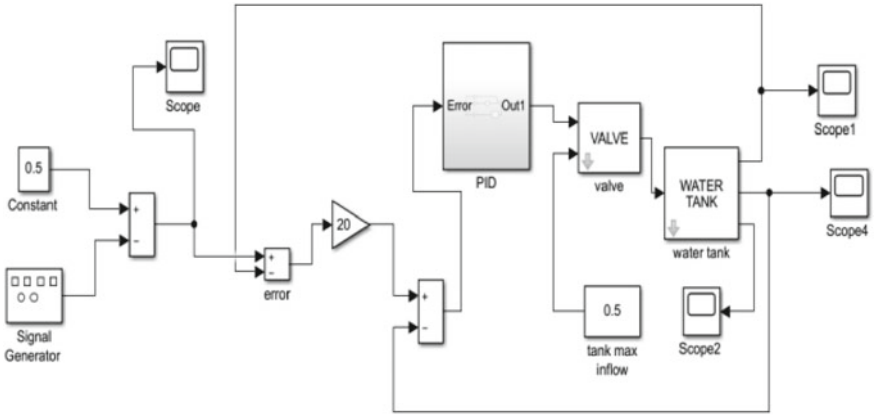


Fig. 13 PID control scheme simulation model

with the help of scope results are observed. Scope 2 gives the value of water level, scope 3 describes the flowout, and 4 describes overflow.

Figure 15 describes the model of the liquid level control system using 2 DOF PID controllers. Here, the subsystem obtained in Figs. 10 and 11 are connected with 2 dof PID controllers to get the output and with the help of scope results are observed. Scope 2 gives the value of water level, scope 3 describes the flowout, and 4 describes overflow.

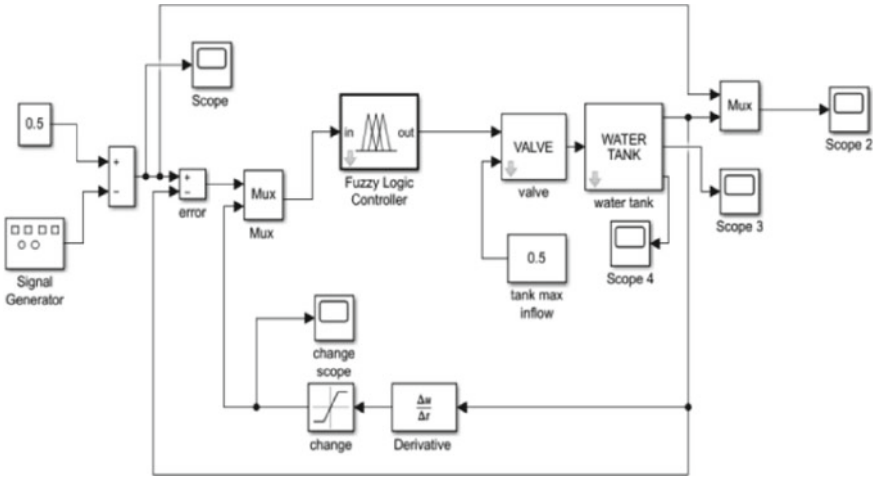


Fig. 14 Fuzzy control scheme simulation model

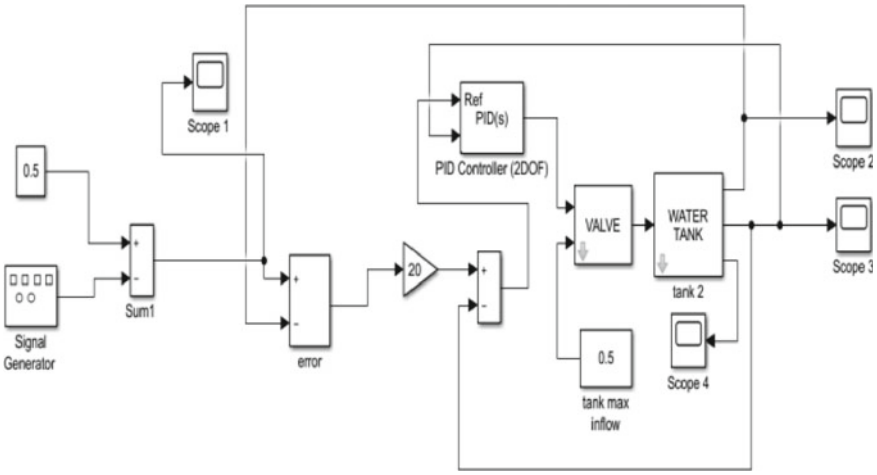


Fig. 15 2DOF-PID control scheme simulation model

4 Simulation Results

Simulation results are obtained from the above models using scope. For different type of controllers, three outputs are obtained for water level, flow out, and overflow. Figures 16, 17, 18 and 19 gives the scope output for PID, fuzzy and 2DOF-PID controller, respectively, along with rules defined. Figure 16a–c describes the output obtained on scope for the water level, flowout, and overflow (which is zero), respectively, for the PID control scheme.

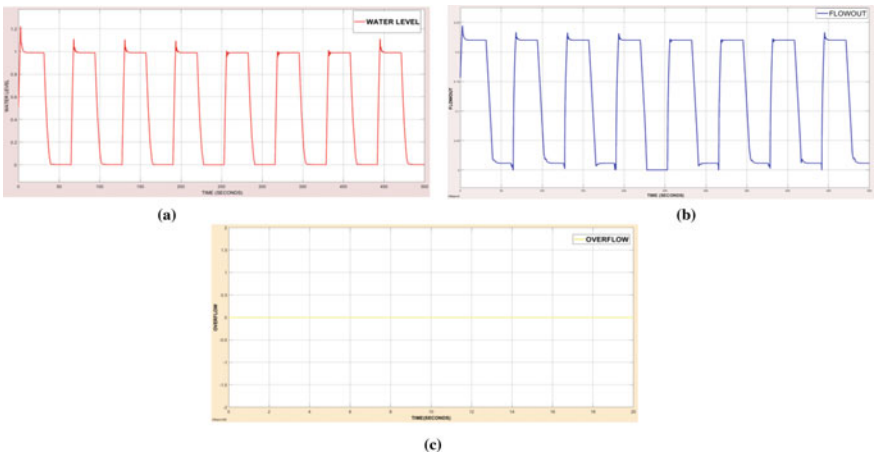


Fig. 16 **a** Water level using PID controller, **b** flowout using PID controller, **c** overflow using PID controller

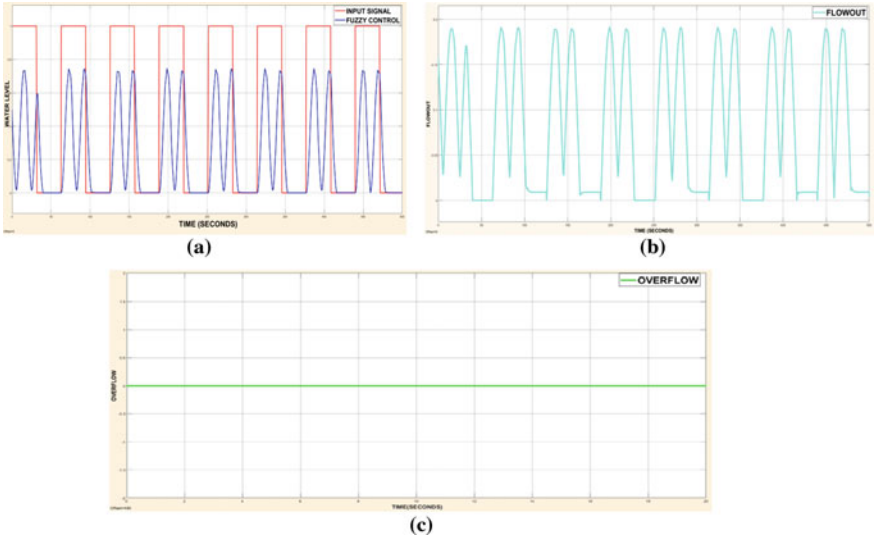


Fig. 17 **a** Water level using fuzzy controller, **b** flowout using fuzzy controller, **c** overflow using fuzzy controller

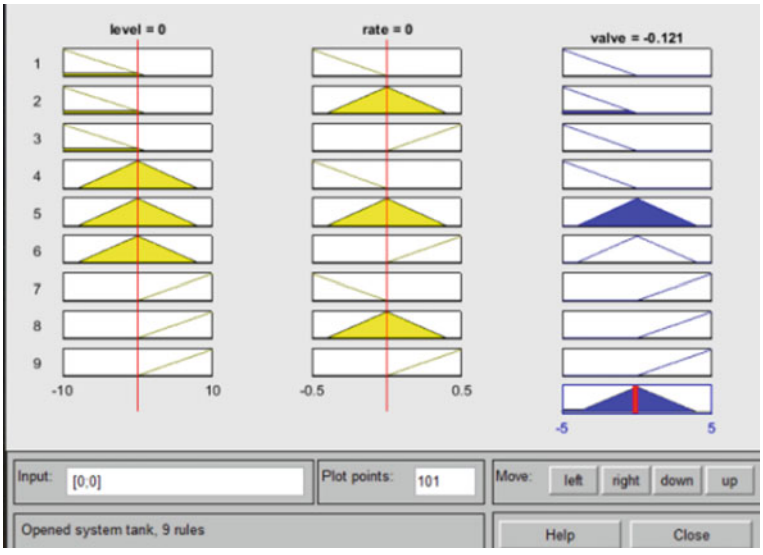


Fig. 18 Rule viewer

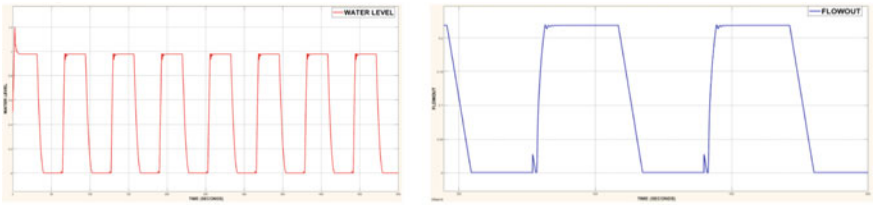


Fig. 19 **a** Water level using 2DOF-PID controller, **b** flowout using 2DOF-PID controller, **c** overflow using 2DOF-PID controller

Table 4 Comparison table of PID, FUZZY, and 2DOF-PID controllers

Parameters	PID controller	Fuzzy logic controller	2DOF PID controller
Settling time(s)	6.9800	6.9800	6.9800
Transient	Present	Not present	Present
Rise time(s)	0.400	0	0.400
Overshoot (%)	100	0	100
Settling time(s)	6.9800	6.9800	6.9800

Figure 17a–c describes the output obtained on the scope for the water level, flowout, and overflow (which is zero), respectively, for the fuzzy control method.

Figure 18 describes the rule obtained using input and output variables.

Figure 19a–c describes the output obtained on scope for the water level, flowout, and overflow (which is zero), respectively, using 2 DOF PID controllers.

From Fig. 16, it can be seen that the PID controller leads to instability in the system because of time delay error. In Fig. 17, it can be seen that the fuzzy control system shows improvement in the system. From Fig. 19, 2 DOF PID control also leads to some instability in the system. Also, overflow occurs in the 2DOF PID system. Table 4 describes the comparative study of PID controller, Fuzzy, 2 DOF PID controllers based on different parameters. Table 4 describes comparison of PID, Fuzzy, and 2DOF PID Controllers.

5 Conclusion

The Simulink model of the liquid level control system is developed and liquid level is controlled using various controllers, viz., PID, 2 DOF PID, and fuzzy logic controllers. It is much easier and simple to implement a fuzzy logic system. Unlike

the PID controller and 2DOF PID controller, the fuzzy logic system has more advantages. With the small rules, one can get better results and responses for liquid level using a fuzzy logic controller. Overshoot is zero on using fuzzy controller. Moreover, rise time is also faster. Thus with fuzzy logic controllers best performance can be achieved in various process industries.

References

1. Dhiman R, Kumar B (2011) Optimization of PID controller for liquid level tank system using intelligent techniques. Can J Electr Electron Eng 2:531–535
2. Abdullah AB (2008) A research project on water level in tank using level sensor and PID controller. University of Malaysia, Pahang
3. Tan W (2011) Water level control for a nuclear steam generator. Nucl Eng Des 241:1873–1880
4. Omijeh BO, Matthew E, Elechi P (2015) Simulated design of water level control system. Com. Eng. Int. Trans. Syst 6:30–40
5. James D (2004) Automatic liquid level controller using a LabVIEW based PC. America Society for Engineering
6. Moradi H, Saffar-Avval M (2012) Sliding mode control of drum water level in an industrial boiler unit with time-varying parameters a comparison with H-∞robust control approach. J Process Control 22:1844–1855
7. Khaled Reza SM (2010) Microcontroller based automated water level sensing and controlling: Design and implementation issues. In: Proceedings World congress on engineering and computer science, San Francisco. USA, vol I, pp 20–22
8. Galzina V, Saric TR (2008) Application of fuzzy logic in boiler control. Tech Gazette 15:15–21
9. Zhao T, Li P, Cao J (2012) Study of interval Type-2 fuzzy controller for twin-tank water level system 20:1102–1106
10. Wu D, Karray F, Song I (2005) Water level control by fuzzy logic and neural networks. In: IEEE conference on control applications, pp 3134–3139
11. D. Su,. K. Ren,. J.: Programmed and simulation of the fuzzy control list in fuzzy control, IEEE/WCICA. (2010) 1935–1940
12. Zhi Z, Hu L (2011) Performance assessment for the water level control system in steam generator of the nuclear power plant. IEEE/CC, pp 5842–5847
13. Chin G (2008) A research project on design and modelling of multiple tank control for fluid circulation system using fuzzy controller. Technical University of Malaysia, Melaka
14. Mukhtar A, Tayal VK, Singh H (2019) PSO optimized PID controller design for the process liquid level control. In: 3rd International conference on recent developments in control, automation & power engineering (RDCAPE), pp 590–593
15. Ye J, Xiaoguang Z, Chuanmao L, Puhao W, Haifeng L (2020) Design of liquid level control system for double tank. IOP Conf Ser: Mater Sci Eng 740(1):012097
16. Yumurtacı M, Verim Ö (2020) Liquid level control with different control methods based on Matlab/Simulink and Arduino for the control systems lesson. Int Adv Res Eng J 4(3):249–254
17. O'Dwyer A (2009) Handbook of a PI and PID controller tuning rules, 3rd edn. National Instruments, DataSheet NI-USB 6008
18. Kumar P, Pathan S, Mashilkar B (2003) Designing PID controller using LabVIEW for controlling fluid level of vessel. Int J Eng Res Appl
19. Georgi (2008) Virtual techniques for liquid level monitoring using differential pressure sensors. Recent 9
20. Disha, Pandey P, Chugh R (2012) Simulation of water level control in a tank using fuzzy logic. IOSR J Electr Electron Eng (IOSRJEEE) 2(3):9–12

21. King, PJ, Mamdani EH (1977) The application of fuzzy control systems to industrial processes. Automatica 13(3):235–242
22. Chen W, Gang W (2006) Modeling of nonlinear two-tank system and model predictive control. J Syst Simul 18(8):2078–2081
23. Meng X, Haisheng Y, Wu H, Tao X (2020) Disturbance observer-based integral backstepping control for a two-tank liquid level system subject to external disturbances. Math Prob Eng 2020 (2020)
24. Thammarat C, Kittisak L, Deacha P, Sarot H, Auttarat N, Supapom S (2018) Application of Bat-inspired algorithm to optimal PIDA controller design for liquid-level system. In: 2018 International electrical engineering congress (iEECON), pp 1–4
25. Alotaibi M, Mohammed B, Waleed A, Fahd A (2019) Implementation of liquid level control system. In: 2019 IEEE international conference on automatic control and intelligent systems (I2CACIS), pp 311–314

IOT-Based Self-nourishing System for Plants

Asha Rani Mishra, Sanjeev Kumar Pippal, Bhavika Rajpal, Arshita Garg, and Sushant Singh

Abstract A proper usage of a self-watering system is very necessary since plants wither and thus it becomes challenging for them to keep their plants healthy and alive. Amount of water to the plants should be also monitored to avoid wastage of large amounts of water. For this reason, an automatic plant watering and soil moisture monitoring system which encourages the plant growth by giving it exactly the right amount of water and in all climatic conditions is required. In this paper, we proposed an IOT based Node MCU microcontroller system to automate the watering process of the plant to help plants grow directly and efficiently without human care or assistance. Proposed system helps to reduce water use and can be used to save time by using the value of moisture content of plants at a given time to decide a particular plant's water need. Desired amount of water is supplied till it reaches a predefined threshold.

Keywords Automatic watering system · Node MCU · Sensors · Relay · Motor · Internet of Things

1 Introduction

Self-watering pots or self-nourishing plant pots are tools that can be used to water the plants from time to time. These pots can be beneficial in places such as balconies and terraces, where the use or use of a water spigot is restricted. It is also used for plants that do not expect to retain water on the leaves. These pots are controlled by the Arduino microcontroller. Self-watering pots are a great idea, largely due to their practicality. They are way better than regular pots and are more reliable and safer. A self-watering pot is, in fact, a source of nourishment for the plant, which constantly eliminates water scarcity and ensures effective growth of plants in the garden [1]. And so, it makes perfect sense to consider investing in a self-nourishing plant pot. Automated self-watering plant systems provide proper quantity of water as

A. R. Mishra (✉) · S. K. Pippal · B. Rajpal · A. Garg · S. Singh
Department of Computer Science and Engineering, GL Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: asha.mishra@glnitm.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_13

per the requirements of species of plants so it will help in keeping plants healthy. An automatic watering plant should be designed to work in both dry and rainy season [2]. Generally, plants need to be watered twice or thrice a day depending upon the climate conditions and altitude.

In this paper, self-nourishing plant pots are designed for both indoor and outdoor use. To make it more real time and efficient it is handled with the help of google assistant. Google assistant provides the real time update about the health of the plant by giving real-time moisture content, temperature, and humidity updates and also checks if there's enough water in the reservoir or not. When the soil dries, water pumping begins. Therefore, microcontrollers are programmed every 4–5 h a day to check soil moisture and save time and keep plants healthy. Soil moisture sensors are controlled by node MCU which monitors water and soil moisture and the system automatically switches on the pump whenever the power is on.

2 Related Work

Devika et al. [3] proposed a prototype of self-watering plant system to address the challenge associated with manual watering system which is much user dependent. The developed prototype waters itself from a large water tank along with the artificial sunlight. The prototype updates status of both current conditions and past along with the reminder for refilling the tank. The system automation provides more assistance to the user.

van Noort [4] No more late watering and always warned in time that the water tank is empty and never too much or too little. At the end of 2015, the prototype of a self-watering pot will be tested. This so-called Parrot Pot is a pot equipped with a 2.5 L water tank that waters the plants via four nozzles. The pot is designed to automatically water based on the amount of available water in the pot, EC, ambient temperature and amount of light. Data is measured every 15 min and sent to the plant database via Bluetooth, and on that basis, water is given or not. The watering is therefore tailored to each plant.

Kumar Sahu and Behera [5] proposes a cost-effective automated smart irrigation model, which controls the water motor and determines direction of water flow using the functionality of soil moisture sensor. Information about flow and direction is provided to the farmers.

Manoj and Udupa [6] proposes a real time response system which ensures sufficient level of water to avoid under irrigation and over-irrigation conditions based on Arduino and soil-based moisture sensors.

Raj et al. [7] discusses research carried out using various pattern recognition techniques in the agricultural sector since there is a need for pre-processing, feature extraction, feature selection and classification process for various applications of the agricultural sector. This helps to analyze how the different pre-processing, feature extraction, feature selection and classification can be applied in gardening.

Tongke [8] in this paper, cloud-based agricultural information using RFID is proposed. Efficient utilization of hardware resources is integrated by using virtualization, dynamic distribution, and balancing load. Radio frequency identification for data collection, wireless environment, automatic control, and IOT cloud-based sensing are used to realize smart agriculture.

Srija and Bala Murali Krishna [9] implements an agricultural automation system based on GSM and WEB. To monitor the status of sensors to optimize the use of water. GPRS technology is used for delivering information at remote locations.

Singh and Saikia [10] proposes an Arduino based irrigation system which considers various environmental parameters like humidity, quantity of water needed and temperature. For this purpose, sensors like flow sensor, temperature sensor and soil moisture sensor were used. Data collected by sensors were linked to a website to give real time interaction. Moreover, users can control irrigation pumps and sprinklers over a distance for producing high quality crop production.

Mediawan et al. [11] proposes a system which automatically remove water and hydroponic nutrients by monitoring the ambient temperature. This system solves the issue of watering, hydroponic and damping which results in fast growth of plants as compared to conventional methods. The system is operational in 208–214 VAC and 15 VDC with Arduino pin voltage of 4.8 VDC.

Jacquline et al. [12] provided a prototype using microcontroller, soil moisture sensors, solenoid valve, relays. If the value of soil moisture falls below the specified value (below 30%), a solenoid valve is opened to pour water. The system stops this when the moisture level reaches 35%.

Prasojo et al. [13] proposes a programmed microcontroller device for controlling watering for edible plants. Need of watering in edible plants depends upon environmental seasons. The device does proper monitoring of moisture content which helps to increase productivity in all seasons. Also, it is suitable in parks, gardens and seeded nurseries. Designed system consists of a microcontroller chip which is programmed to automatic control watering based on soil moisture to detect soil dryness. The designed system used a copper plate sensor acting as an electrode The copper measured soil resistance which is changed to analog form, i.e., voltage. It is passed to Arduino Uno in digital form. The upper limit is decided experimentally based on different soil conditions. Solenoid valves are used to reduce the use of electrical energy instead of pumps.

Astuti Ningtyas et al. [14] proposes IOT cloud-based device for smart Garden by utilizing internet bandwidth for data communication in wireless networks using sensors. This paper discusses IOT-based Arduino to control and monitor all plant related activities.

Bains et al. [15] highlights the need of watering the plants which requires extensive manpower in maintaining greenhouse operation which covers two aspects—when to water and how much to water. In this regard, self-plant watering system can be of great help. Proposed system used Arduino board with ATmega328 Microcontroller which works based on moisture level of plants.

Liu et al. [16] discuss the rise of speedy expansion of both arid and semiarid areas led to the decline in the level of groundwater which in turn makes it impossible for

sufficient watering of plants. To provide water sustainably for good plant life, a self-watering system has been proposed. It is designed to collect water either from the rain water, dew or groundwater to provide reliably to the surface vegetation. Results showed that the proposed system raised the groundwater to a considerable level as compared to sandy ground.

Manoj et al. [17] designed a prototype based on Arduino controllers to improve the scope in the field of agriculture. It senses the humidity level of the plant and accordingly supplies water to the plants using a large water tank. Also, it alerts the user in case of low level in the water tank.

Available state of art includes following limitations were found in the previous models shown in Table 1:

(a) Most of the model was used for single plant use only and uses Arduino.
(b) No concept of vertical plantation or drip system was there.
(c) The previous model wasn't able to give real time updates through Google Assistant.

Table 1 Existing self watering system

Author(s)/researcher(s)	Highlights
Devika et al. [3]	Microcontroller Atmega 328, GSM Module, Motor
van Noort [4]	2.5 L water tank with four nozzles, Bluetooth
Kumar Sahu and Behera [5]	Soil moisture is used to determine both flow and direction of water motor
Manoj and Udupa [6]	Real time, manages over and under irrigation condition, Arduino based
Raj et al. [7]	Pattern recognition techniques like preprocessing, feature extraction
Tongke [8]	Cloud based, RFID
Srija and Bala Murali Krishna [9]	GPRS as information technology
Singh and Saikia [10]	Arduino based, humidity, temperature and level of water as parameters
Mediawan et al. [11]	Remove excess water and hydroponic nutrients
Jacquline et al. [12]	Microcontroller, soil moisture sensors, solenoid valve, relays, threshold for soil moisture 30% (lower), 35% (upper)
Prasojo et al. [13]	Programmed microcontroller
Astuti Ningtyas et al. [14]	IOT cloud-based device, Arduino efficient in wireless networks
Bains et al. [15]	Arduino board with ATmega328 Microcontroller
Liu et al. [16]	Use of rain water and ground water
Manoj et al. [17]	Arduino controllers, humidity sensor and alert to users
Prasojo et al. [13]	Microcontroller chip with copper plated sensor, Arduino Uno

(d) When the water in the reservoir got empty there was no scope to refill it automatically when the plant is used outside the home or on the balcony.

The research aims to design and develop a prototype of a self-watering plant using Node MCU and considers parameters like temperature, humidity, moisture content.

3 Workflow of Self-nourishing System

This research utilized the use of various sensors like soil, humidity and temperature to determine the correct timeline for watering the plants. Figure 1 shows various modules of the proposed prototype whereas Fig. 2 represents the methodology used.

Steps Followed for Self-nourishing System are Given as Follows

(a) The main heart of the self-watering system is Arduino Uno with sensors which manage all the work. Initially, it will measure air temperature. If it is at least 5 °C, then the system will continue, otherwise it will stop, because it's too cold for watering.

(b) After that it will check if there's enough water in the reservoir (a simple bucket) using an ultrasonic sensor. If there is at least 25 cm, the pump will not switch on, as dry running can damage it and will wait for the user to fill up, otherwise it will move to check the soil moisture.

(c) It then checks whether it is also necessary to use a soil moisture sensor. If water is not needed, then apparently nothing happens and the measuring cycle is repeated after a few minutes.

Otherwise, if the soil is too dry, it will turn on the water pump.

(d) The relay will allow the pump to open for 10 s which will provide power to the pump, thus irrigating plants.

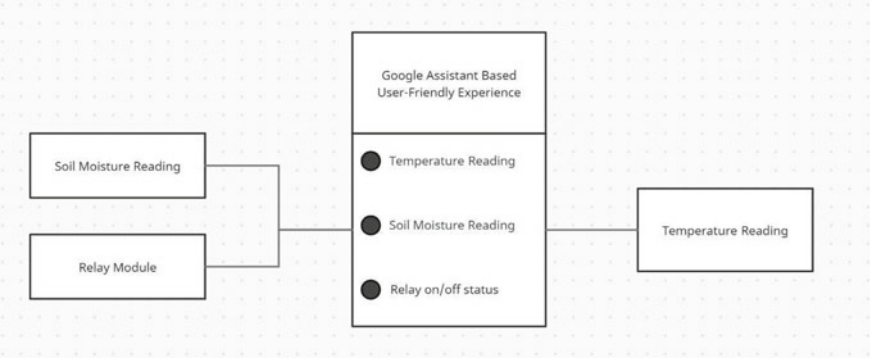


Fig. 1 Modules hierarchy of self-nourishing plant pot

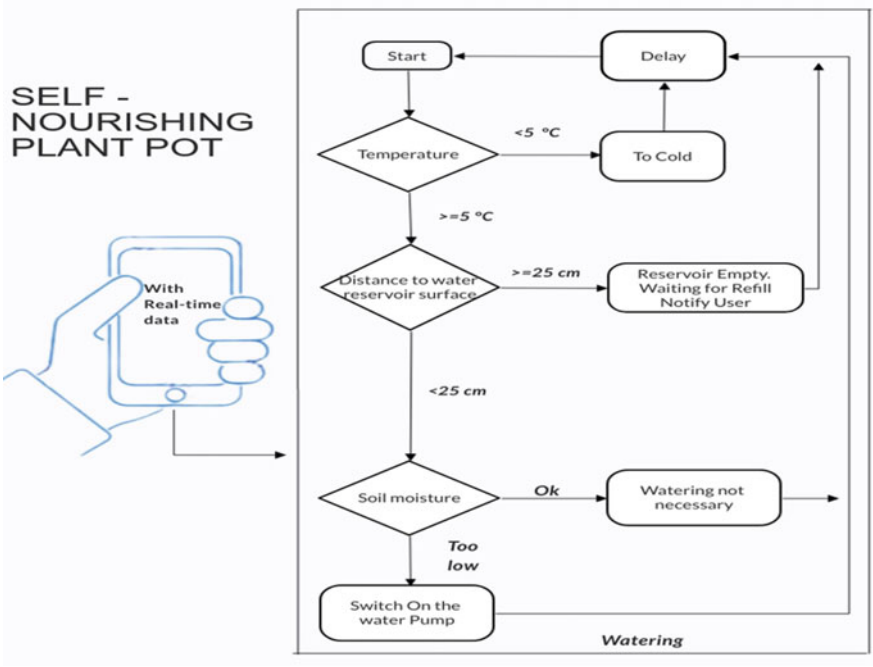


Fig. 2 Methodology of self-nourishing system

(e) The measurement cycle will be repeated after a minute until the soil moisture is above the given threshold value. This occasional irrigation (10 s every minute) has shown that it is better than continuous irrigation, as water requires some time to sink to the bottom of the plant, where the soil moisture sensor is located. With continuous irrigation, there will be too much water in the planter before the sensor can detect it.

The system as the name indicates, 'Self-Nourishing Plant Pot' makes the system more flexible and provides an attractive user interface compared to other automatic watering systems. In this research, we have integrated mobile devices into self-Nourishing systems. A novel architecture for a self-nourishing plant is proposed using relatively new communication technologies shown in Fig. 3. The system consists of mainly five components: a relay module, Node MCU microcontroller, ultrasonic sensor, soil moisture sensor, humidity and temperature sensor, ESP2866 Wi-Fi module, and pump shown in Fig. 4. WI-FI is used as the communication channel between android phone and the Node MCU microcontroller. Node MCU and Arduino IDE are basic components of the system. C++ language is used for its more interactive support in Node MCU and sensor libraries like DHT11 and Capacitive soil moisture. Blynk to store and monitor real time updated data. Google Dialog Flow to build conversational applications and experiences for a company's customers

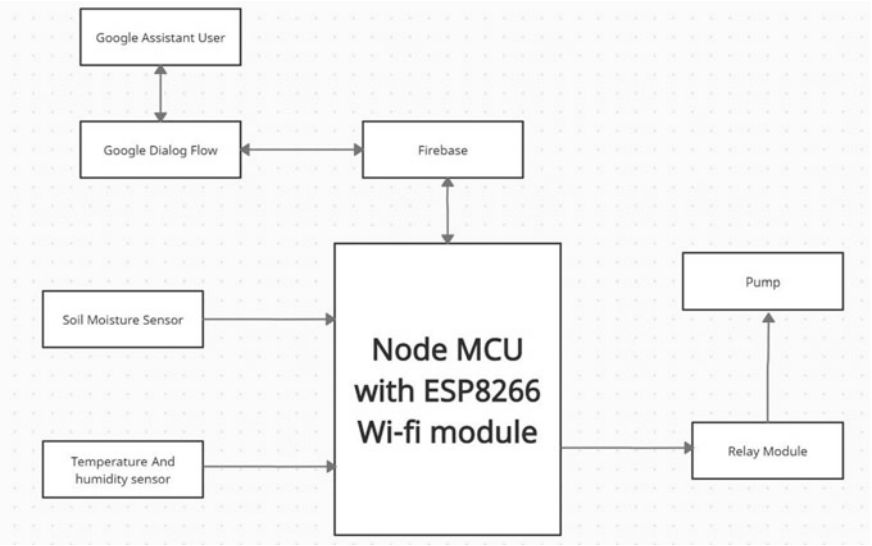


Fig. 3 Proposed architecture

in various languages and on multiple platforms, i.e., it is used for google assistant framework. Table 2 describes various resources required for the proposed prototype.

4 Results and Discussions

The system works once the moisture sensor sends information to the Node MCU about dryness of the soil. Next Arduino will send a command to the relay of the pump to water the plant. Also, all can be irrigated at the same time, using vertical plantation moisture sensors. If the water tank level is low the system will not operate in any respect, even all plant sensors are activated to protect the water pump. Furthermore, this system has been configured in such a manner that in case of a rainy day, it's not going to work, because the temperature sensor will activate, and it will send information to Arduino to stop the water pump and to close all valves too. Further, a real-time time update can be obtained by using the Blynk app. The wire connections from the controllers to the Node MCU have been pretty difficult, due to a single mistake that can damage any electric element. Arduino helped to program the system and upload it in a microcontroller to run the water pump and starting sensors to give best results. Connection of wires from devices to the Arduino was done by means of using the plastic breadboard. For designing the plant, special flexible pipes had been used to facilitate the connection from the water tank to the plant.

Fig. 4 Self nourishing system components

Table 2 Resources required for proposed model

Entity	Attribute	Resources required	Type of resource
Node MCU	Device_id	Node MCU	Microprocessor
	Device_credential	Node MCU	Microprocessor
	SoilM_Reading	Capacitive soil sensor	Sensor
	Temp_Reading	DHT11 sensor	Sensor
Temperature sensor	Device_id	DHT11 sensor	Sensor
	Temp_Reading	DHT11 sensor	Sensor
Soil moisture sensor	Device_id	Capacitive soil sensor	Sensor
	SoilM_Reading	Capacitive soil sensor	Sensor
Relay	Device_id	Relay module	Module
	Pump On/Off	Pump	Part of electric
Self-nourishing pot	Temp_Reading	DHT11 sensor	Sensor
	SoilM_Reading	Capacitive soil sensor	Sensor
	Report	Google assistant	Mobile phone

Proper experiments were carried out to check the functionality of components in the proposed system showed in Fig. 5. Table 3 shows the experimental result carried out. Figure 6 display of values using Blynk App.

Fig. 5 Different stages in testing of self-nourishing plant system

Table 3 Result for sensors test

S. No.	Temperature value (in Celsius)	Status	Soil moisture (in %)	Humidity (in %)	Distance (in cm)	Message/alert
1	34	Ok	27	65	22	Watering needed
2	34	Ok	43	65	22	Watering needed
3	34	Ok	67	65	22	Watering needed
4	32	Ok	75	63	0	Watering not necessary at the moment
5	32	Ok	73	64	24	Watering not necessary at the moment

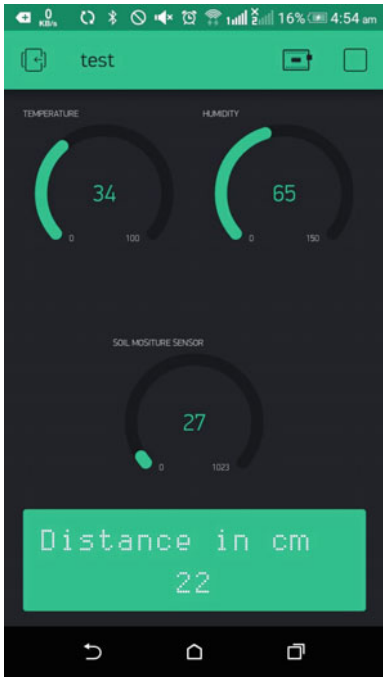


Fig. 6 Display of values using Blynk App

5 Conclusion and Future Scope

In this self-nourishing system, mobile technology is used to design and develop a smart watering system with new communication technologies. It is relatively simpler in nature due to changes in simple related concepts so that it can be fitted with functionality provided by mobile devices for display using less storage. Ease of access, less cost, remote controlled nature and automatic configuration are the main highlights of the proposed system. Moreover, it can be easily extended and is flexible as compared to other automatic watering systems. In future, it can be extended for multiple plants with an objective of minimizing the cost.

References

1. Bhardwaj S, Dhir S, Hooda M (2018) Automatic plant watering system using IoT. In: 2018 Second international conference on green computing and Internet of Things (ICGCIoT), pp 659–663
2. Vimal SP, Kumar NS, Kasiselvanathan M, Gurumoorthy KB (2021) Smart irrigation system in agriculture. J Phys: Conf Ser, p 012028
3. Devika, S. Khamuruddeen S, Khamurunnisa S, Thota J, Shaik K (2014) Arduino based automatic plant watering system. Int J Adv Res Comput Science Softw Eng 4(10)

4. van Noort FR (2015) (Each plant has its own watering–the Parrot Pot https://research.wur.nl/en/publications/wagenin
5. Kumar Sahu C, Behera P (2015) A low-cost smart irrigation control system. In: 2nd International conference on electronics and communication systems, ICECS (2015), pp 1146–1151
6. Manoj HG, Udupa NGS (2015) Application of soil moisture sensor in mixed farming. Int Res J Eng Technol (IRJET) 2(4):311–314
7. Raj MP, Swaminarayan PR, Saini JR, Parmar DK (2015) Applications of pattern recognition algorithms in agriculture: a review. Int J Adv Netw Appl 6:2495
8. Tongke F (2013) Smart agriculture based on cloud computing and IOT. J Convergence Inf Technol (JCIT) 8(2)
9. Srija V, Bala Murali Krishna P (2015) Implementation of agricultural automation system using web and GSM technologies. IJESR. 5(9):1201–1209, ISSN 2277-2685
10. Singh P, Saikia S (2017) Arduino-based smart irrigation using water flow sensor, soil moisture sensor, temperature sensor and ESP8266 WiFi module. In: IEEE Region 10 humanitarian technology conference 2016, R10-HTC 2016 proceedings
11. Mediawan M, Yusro M, Bintoro J (2018) Arduino: automatic watering system in plant house -using Arduino. IOP Conf Ser: Mater Sci Eng 434:012220
12. Jacquline M, Suseno NC, Manaha R (2019) Automatic watering system for plants with IoT monitoring and notification. CogITo Smart J 4(2):316
13. Prasojo I, Maseleno A, Shahu N(2020) Design of automatic watering system based on Arduino. J Robot Control (JRC) 12:59–63
14. Astuti Ningtyas M, Nugraheni M, Suyoto S (2021) Automatic plants watering system for small garden, pp 200–207
15. Bains PS, Jindal RK, Channi HK (2017) Modeling and designing of automatic plant watering system using Arduino. Int J Sci Res Sci Technol (IJSRST) 3(7):676–680
16. Liu Q, Yasufuku N, Omine K (2018) Self-watering system for arid area: a method to combat desertification. Soils Found 58(4):838–852
17. Manoj T, Reddy KM, Sujatha M (2017) Smart farming system using Sms. Indian J Public Health Res Dev 8(4)

Social Distance Alerting Device for COVID-19

Lakshay Bansal, Anant Garg, Ishanvi Prasad, and Ravendra Singh

Abstract We are proposing an IoT-based social distancing device as a preventive measure to COVID-19. It uses NodeMCU in conjunction with ultrasonic sensor temperature sensor, while vibrator buzzer is used for an alarming mechanism. The ultrasonic sensor is used to obtain higher accuracy as it uses LOS principle to measure the distance. The alarm will be raised whenever measured distance is found to be less than six feet. Temperature sensor is used to alert the user to isolate them if their body temperature goes above 102 °F, thereby decreasing the transmission possibility of virus in case he is infected with the virus.

Keywords Social distancing · IoT · NodeMCU · COVID-19

1 Introduction

COVID-19 has impacted the world in an unprecedented manner and has left a global scar. It has changed our lives in an unimaginable manner. COVID-19 which is caused by the new SARS-CoV-2 coronavirus started from the China Wuhan city on December 12, 2019. It rapidly spread throughout the nation and soon, millions of people in China were infected with this virus. It then spread aggressively and spread globally within two month which raise the global alarm, and it has been declared global pandemic. This pandemic which is termed as COVID-19 is caused by severe acute respiratory syndrome coronavirus 2, or SARS-CoV-2. Lockdown are announced to break the chain of coronavirus transmission (like social distancing).

L. Bansal (✉) · A. Garg
Department of Electronics and Communication Engineering, Jamia Millia Islamia, New Delhi, India
e-mail: lakshay.b.jmi@gmail.com

I. Prasad
Indrapuram Public School, Nyaya Khand-I, Ghaziabad, India

R. Singh
Department of Electrical and Electronics Engineering, Galgotias College of Engineering and Technology, Greater Noida, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_14

All publics places like shopping centres, theatres, parks, etc., which are a common source of crowding, the authorities measure the skin temperature of the people by infrared ray-based device before allowing them to enter the premises. Seeing the magnitude of impact, many countries have put into place various counter measures such as physical distancing and quarantine to curb the spread of the virus and prevent it from reactivating. Nevertheless, physical distancing is still a prominent measure to curb the spread of COVID-19 infection.

The number of globally reported cases of COVID-19 are 191 million with the registered death of 41 million people as on 19 July 2021. The count of people that have recovered from the disease is also increasing in which vaccines are playing a supporting role to increase the recovery rate by boosting the immune system. However, the SARS-CoV-2 affects the entire respiratory system, especially the lungs [1]. Hence, the post-COVID recovery process plays an equally important role. Physical distancing which is the best preventive measure to be taken to curb the spread of deadly coronavirus, and it can help the patient remain fit, healthy, and safe.

Social distancing mean to maintain a physical distance between individuals or among crowd in social gathering places or at home. It will help in decline the transmission possibility of SARS-COV-2 as it can transmit more easily by contact then by air transmission. Social distancing was also used to reduce the impact of influenza-like diseases in 2009. In addition to other preventive measures, social distancing can play a more effective role [2]. But social distancing is something new for younger generation who want to go out and have social gathering, want to explore new things. To help them understand the relevance of physical distancing which is necessary to curb the spread of coronavirus, we are proposing a project to alert them whenever they try to or by mistake violate the social distancing guideline.

As wearable technologies devices are gaining the attention of masses, which can be worn by the user and can process and display relevant information in real time about the user. Wearable technology has a variety of applications which grows as the field itself expands. The technologies employed by these latest devices are very sophisticated and state of the art. We have developed a wearable device that have peripheral sensor to detect the distance between the user and other individuals (obstacle) and alert them when the imaginary circular region having radius three feet spaces of one user overlaps with that of another user. It will have temperature sensor to alert the user of device if his/her temperature is above the 102 F (one of the COVID-19 symptom). For alerting the user, we are using devices which simulate the visual, sensational, and auditory attention of the user by LED, vibrator, and buzzer, respectively.

2 Literature Survey

It is possible to minimize the spread of COVID infection by making the following recommendations. Staying at home (home quarantine) and avoiding any direct contact with any healthy (possible asymptomatic patients) or infected person, which

has been called shielding; avoiding nonessential travel; observing social distancing rules like avoiding crowded public places and maintaining at least two metres of distance between each person, especially if they are coughing or sneezing; avoiding shaking hands when greeting others; frequently washing hands for at least 20 s with soap and water or hand sanitizer with at least 60% alcohol, especially after touching common surface areas, using the bathroom, or shaking hands, avoiding touching eyes, nose, and mouth with unwashed hands; and disinfecting surfaces using household sprays or wipes [3]. The actual impact of social distancing is still a debatable topic owing to seasonal changes. While some studies indicate that it is inadequate during autumn and winter seasons, others indicate that it is more efficient during summers [4]. Nevertheless, it plays an important part in curbing the virus and hence must be followed under all circumstances. Social distancing is designed to reduce interactions between people in a broader community in which individuals may be infectious but have not yet been identified hence not yet isolated [5].

The author [6] after analysing the results, it has been found that in a controlled scenario, the peak of the infection curve can be reduced by nearly 50% thus, avoiding unnecessary burden on healthcare systems and considerably reducing the mortality rate. An age-structured SIR model taking into consideration the duration of non-attendance in the workplace, closure of schools, and the lockdown has been analysed to observe the COVID-19 progress in India. The mathematical model takes into consideration two important factors, age, and social distance, as both are important parameters to reduce the effect of COVID-19.

To know the understanding of audience on the issue social distancing, they had conducted the survey in which they find out only 56% of the young adults and teenagers always followed it. They are using ultrasonic sensor in conjunction with mobile application using geolocation feature to alert the user when distance got below 244 cm. And a point-based motivation system to motivate the user to earn more points and points get deducted when user violate social distancing norm [7]. As retail sector is also affected by the coronavirus. They had proposed a smart retail system to real time monitoring of retail outlets, shopping malls (social gathering), shops, etc. A deep learning model to distinguish the people who are wearing mask or not. A wearable device to alert the user whenever detected the violation of social distancing. And a RFID-based shopping cart and an automatic billing system [8].

By using Wi-Fi signals, the author proposes a prototype of a wearable device that calculates the proximity between two users wearing the device. This is done using the strength of the Wi-Fi signals emitted from both the devices. The device sounds an alarm when and if the distance between the two users is less than the threshold distance [9]. The IoT-based system has been used for monitoring indoor and outdoor physical distancing relating to COVID-19. Wi-Fi adaptors catch RSSI signal from wireless devices and use the count of devices for capacity control of interior space and have manual alarm by the supervisor if counted device breached threshold of the maximum allowed capacity. Machine learning model with the data collected from the Wi-Fi-enabled band wear by the user is used to train the model. And use higher precision laser sensors to monitor distance between the user [10].

The author proposed a machine learning-based prediction model for the health of the user, which takes input parameter of user vital like temperature, heartrate, etc., and then predict his medical report [11].

3 Methodology

The possible reasons why people are not maintaining social distancing are lack of motivation and along with knowledge. As most of us love to socialize and exercise outdoors and when we encounter are known ones it hard for us to emotionally as well as physically to not attend them. And most of us think that wearing mask and having sanitized our hand is just enough measure to protect ourselves from SARS virus. In social gatherings, people usually become negligent towards COVID protocols like social distancing. As a result, these places become super spreaders and COVID hotspots. To address this problem, our device acts as an alarm to help people maintain social distancing norms. If a person wearing the device comes at distance less than 6 ft to another person, the device starts vibrating for duration of 5 s. This gives the wearer a soft signal informing him that he is at a distance less than 6 ft to someone. After soft signal, the device waits for 30 s (it is a time given to user maintain the distance as per social gathering norm) before raising an alarm having buzzer sound from the device. This is a hard signal and a strong warning for the wearer to follow social distancing norms (Fig. 1).

Figure 2 demonstrates the application of the device in following social distancing norms. When the imaginary circle which is of radius 3 ft of two people overlap as shown with the pink coloured circle, now made users who are wearing the device have distance between them less than 6 ft which raise the alarm for both the user.

Fig. 1 Social gathering

Fig. 2 Imaginary circle of user overlap

The device starts vibrating for 5 s. After 30 s, still if the distance is not increased beyond 6 ft, then alarm starts ringing from the device as can be seen in the figure. And as of a natural response to alarm, they try to increase their distance by moving away from each other will again maintain the distance between them greater than 6 ft.

Once the distance between the people becomes more than 6 ft the device stops ringing and goes back to the normal state. Also, other people marked by grey circles are at a distance of more than 6 ft from the person closest to them. Hence, the device stays in the normal state, i.e. without any vibration and alarm.

3.1 Flow Chart of the Device Working

The flow chart in Fig. 3 gives the actual and proposed working model of the device. For a new user, the device asks for his Aadhar UID. If the UID provided by the user is valid, the device goes on to connect to the Aarogya Setu app. On the other hand, if the UID is invalid or does not exist then the user is suggested to register for Aadhar UID.

Linking the app to Aadhar will help in uniquely identifying each user so that potential defaulters can be found out. These defaulters can then be put for corrective measures by organizing online/offline session for them, which talks on the importance of social distancing. The device must be worn by a user. If the distance between the user and the next closest person to him is greater than 6 ft, the device remains inactive, and the user can continue as usual. However, if the distance becomes less than 6 ft, then the vibration and alarm mechanism triggers till the appropriate distance is maintained. In this way, the device helps in enforcing social distancing norms. And they are regularly monitor by the device, if the body temperature of the user goes

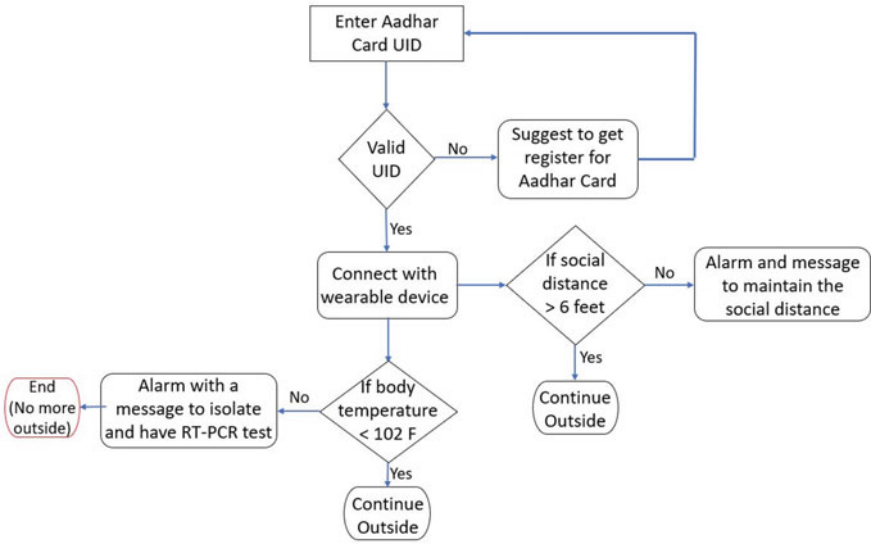


Fig. 3 Flow chart of device working

above 102 °F, then it raise the alarm with a suggestion message to have a RT-PCR test and to isolate himself from the crowded place to decline the transmission possibility of virus to other in case they are found to be COVID positive after having a RT-PCR test. Now, they are not allowed outside anymore.

3.2 Device Block Diagram Overview

As an improvement to this method, we propose to link the device to the person's Aadhar card using the open-source (by government of India) mobile application named Aarogya Setu. Aarogya Setu is an Indian COVID-19 "contact tracing, syndromic mapping, and self-assessment" digital service, primarily a mobile app. It has been developed by the National Informatics Centre (NIC) under the Ministry of Electronics and Information Technology (MeitY). Levering to the open sourcing of this app, we can use it for our benefit. It will use NodeMCU in conjunction with ultrasonic sensor and temperature sensor; while vibrator and buzzer for alarming device, to alert the user to maintain social distance of six foot.

We are using LM35 a temperature sensor which is a precision IC. Analog output, i.e. voltage produced is directly proportional to the temperature of the object in contact [12]. We are using HC SR04 ultrasonic sensor to determine the distance between the user and other people (obstacles) [13]. The ultrasonic wave transmitted by the HC SR04 sensor when encounter any obstacle during its travel in straight line path get reflected at encounter of the obstacle. The time difference between

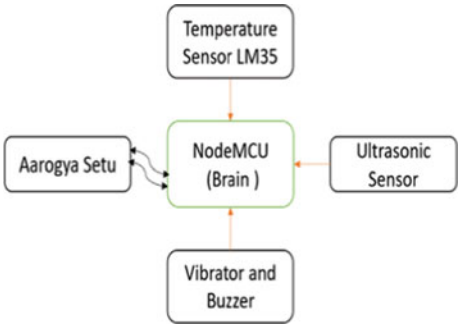


Fig. 4 Block diagram of device

the emission and reception of wave is used to calculate the distance. If the distance measure is found to be less than 6 ft (violation of social distancing guideline), then alarm will be raised.

As seen in Fig. 4, the data from temperature sensor LM35 and ultrasonic sensor is fed to the NodeMCU which is the brain of the device which process the signal and perform necessary actions. It then processes this information to generate results which are then compared with the pre-defined threshold values. If the readings are beyond the threshold values, then the vibrator and buzzer mechanism kicks in. A small electric motor fitted with a weight is used for producing vibrations. When the motor rotates, vibrations are produced due to movement of the weight. Similarly, when disc is vibrated buzzer starts ringing. Leveraging the open-source code of Aarogya Setu, it can be linked with the device using a Wi-Fi module. For the same purpose, NodeMCU is used as it comes with a built in Wi-Fi module which results in cost reduction. On the other hand, using an Arduino Uno is costlier as an external Wi-Fi module is required.

4 Experiment

In Fig. 5, we have shown the experiment set-up of our device using NodeMCU as processing unit of the device. We have used switch button which switch on the device with first push and to stop its operation we must press it again. As shown, we have used green led to show device is working and you are in safe condition. And whenever there is a violation of either social distancing norm or if temperature above the threshold value, the red led will turn ON along with the alarm to notify the user.

LM35 respond for every one-degree rise or fall in Celsius temperature which correspond $+10$ mV or -10 mV respective change in the output voltage produced by the sensor, i.e. have a scale factor is 0.01 V/0C. The analog output from the temperature sensor is converted to a digital by Analog to Digital Converter (ADC) channel. Temperature sensor will give the analog output to the NodeMCU which after processing notify the user by alarming the buzzer if temperature is above 102 °F.

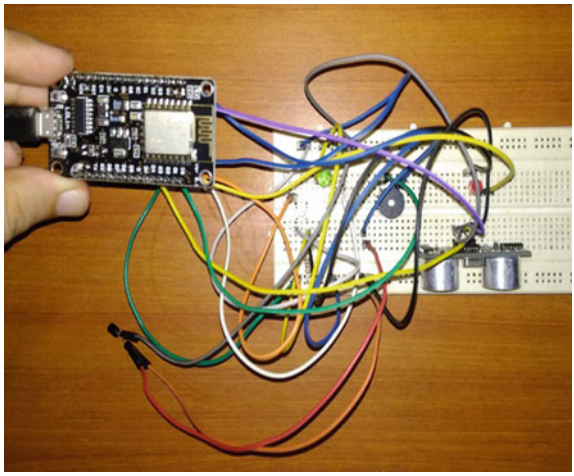


Fig. 5 Device set-up using NodeMCU

$$\begin{aligned} \text{Vout} &= (5/1024) * \text{Analog-Reading} \\ \text{temp (in } ^\circ\text{C)} &= \text{Vout(in V)} * 100 \\ \text{temp (in } ^\circ\text{C)} &= (5/1024) * \text{Analog-Reading} * 100 \\ &= \text{Analog-Reading} * 0.488 \\ \textbf{temp(in F)} &= (\text{temp(in } ^\circ\text{C)} * 1.8) + 32 \end{aligned}$$

While HC SR04, it emits ultrasonic waves that get reflected by the target obstacles. The time between the emission and reception is used to determine the distance. Echo is the basic principle of sensor which uses ultrasonic sound for distance measurement. As, distance is equal to ultrasonic wave speed (340 m/s in air) multiply with time. Hence, it measures the to and fro distance of the object, as it considers the time taken by a wave to travel to and fro from the measuring point which after dividing by two gives us the actual distance.

$$\text{distCm} = (\text{duration} * 0.0340)/2 \tag{1}$$

$$\text{distFoot} = \text{distCm}/30.48 \tag{2}$$

In Fig. 6, we send user's temperature and its social distance to a Blynk App. Where he/she can keep track of their body temperature and can see from the data how many times they had broken the COVID protocol of social distancing.

Figure 7 shows the error message on serial monitor along with the suggestion to what to do next if the social distance between the user wearing the device and the nearest person to him got less than 6 ft between them.

In Fig. 8, error message along with suggestion is displayed on serial monitor. Alarm is raised as body temperature indicate user having high probability of being

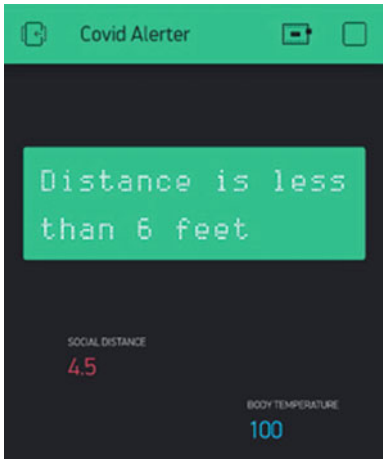


Fig. 6 User log on Blynk App

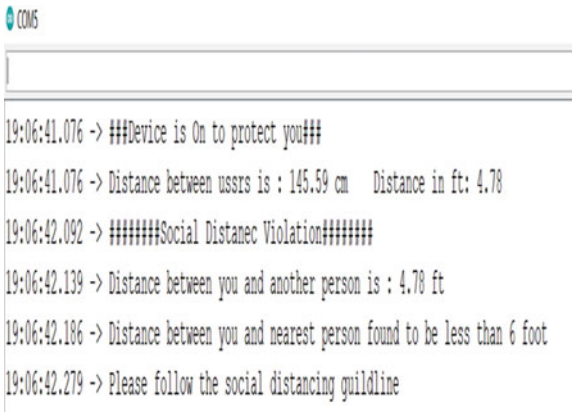


Fig. 7 Distance is less than 6 ft hence result in suggestion

COVID positive, and red led is ON to notify the user to isolate himself from the public place and with a suggestion message to have RT-PCR test if the high fever persist for long time.

4.1 Application

As shown in Fig. 9, the device also helps in warning against potential COVID positive patients. It is often seen that before entering any shop or restaurant, the temperature of the person is measured by a staff member. However, if there are long queues before entering the store, it is a high probability that the person might have affected others. To avoid this, our device continuously measures a person's temperature in real time.

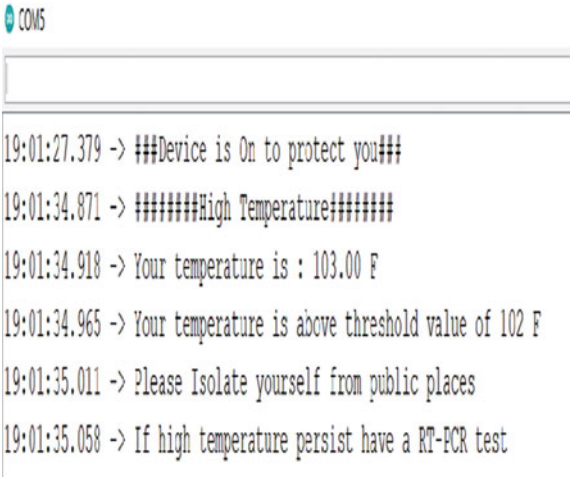


Fig. 8 Temperature above 102 °F recommend user RT-PCR test

If the temperature goes beyond 102 Fahrenheit, an alarm is sounded by the device with a message for the person to conduct RT-PCR test. Otherwise, the device remains in normal state.

Fig. 9 When entering premises of any building

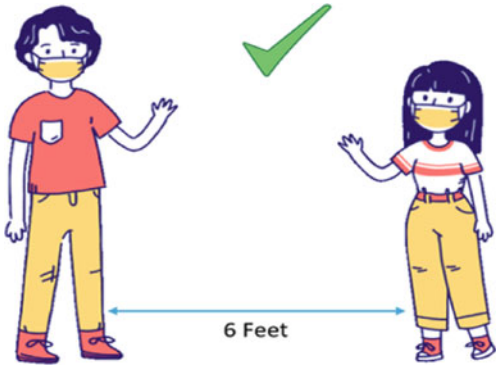


Fig. 10 Social distancing of 6 ft is recommended

5 Conclusion

To keep everyone protected from COVID-19, we are promoting to follow social distancing guideline at the public places along with other effective methods like using face masks, regular hand washing, and sanitisation (Fig. 10).

As we human are not alert all the time, to keep track on the whether we are following the social gathering guideline. We had made a wearable device which will notify any violation of distance at the public place or if they have a symptom of fever by measuring the skin temperature of the user wearing the device. We can raise the alert via LED for visual attention, vibrator for sensational attention, and buzzer for auditory attention. When it comes to health care or any other field, smart technology is the cheapest and the reliable solution for the remote location where capital investment and affordability is a challenge.

6 Future Scope

Presently, we are using ultrasonic sensor which can detect the people or object in front of it but in future, we can use MaxBotix ultrasonic sensors which can hunt out on four sides. As fitness bands have become more common due to their compact size and easy to use; they can be used to measure parameters such as pulse, temperature, and oxygen level. Remote health care has advanced, and its importance has increased tremendously which can be achieved by hosting the data (his/her vitals) of user on website like ThingSpeak, etc. which can then be visualized by the doctor on the real-time basis. This ensures safety to both the doctor and the patient as in current time where social distancing is a key to suppress the spread of coronavirus.

However, it continues to remain a novel concept and people are not yet fully aware of its specifications and benefits. Hence, a point-based system where they will be rewarded with points on daily basis if they haven't violated the social distancing norm

during a day, which can in negotiation with e-commerce site and local outlet can be used to have discount for those users and can be an effective method to promote social distancing.

References

1. COVID GA, Group P-ACS (2020) Post-COVID-19 global health strategies: the need for an interdisciplinary approach. Aging Clin Exp Res (2020)
2. Uzicanin A, Ahmed F, Zviedrite N (2018) Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. BMC Public Health 18(1)
3. Melika Lotfi MR (2020) COVID-19: transmission, prevention, and potential therapeutic opportunities. Clin Chim Acta
4. Kissler S, Tedijanto C, Lipsitch M, Grad YH (2020) Social distancing strategies for curbing the covid-19 epidemic. medRxiv
5. Wilder-Smith A, Freedman D (2019) Isolation, quarantine, social distancing and community containment: pivotal role for oldstyle public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med
6. Singh R, Adhikari R (2020) Age-structured impact of social distancing on the COVID-19 epidemic in India. arXiv
7. Sekhani B, Mahajan M (2021) Social distance encourager and motivation system. Int J Innov Eng Res Technol 8(05):101–106
8. Thilagavathi S, Nivethitha KS, Preeti P, Vikram DT (2021) IoT based smart retail system with social distancing for COVID 19 outbreak. J Phys: Conf Ser, 1917(1):012030
9. Cunha AO, Loureiro JV, Guimaraes RL (2021) Design and development of a wearable device for monitoring social distance using received signal strength indicator. In: Proceedings of the Brazilian symposium on multimedia and the web, ser. WebMedia'20. Association for Computing Machinery, New York, USA, pp 57–60
10. Yacchirema D, Chura A (2021) Safemobility: An iot-based system for safer mobility using machine learning in the age of COVID-19. Procedia Comput Sci 184:524–531
11. R Kanmani, SM (2021). Prediction of covid disease using advanced iot and support vector machine. In: 7th International conference on advanced computing and communication systems, ICACCS, pp 546–550
12. Mansor H, Shukor MHA, Meskam SS, Rusli NQAM, Zamery NS (2013) Body temperature measurement for remote health monitoring system. In: 2013 IEEE International conference on smart instrumentation, measurement and applications (ICSIMA), (2013), pp 1–5
13. Singh NA, Borschbach M (2017) Effect of external factors on accuracy of distance measurement using ultrasonic sensors. In: 2017 International conference on signals and systems (ICSigSys), (2017), pp 266–271

BWDS-Based Spatially Dabble Couple Horizontal-Vertical 4-Port PIPO DRA with Poly-Pattern Diversity

Ramkrishna and Rajveer S. Yaduvanshi

Abstract This speaks to data transmission analyst surface (BWDS) given 4-port various information numerous yield (PIPO) tube-shaped dielectric resonator reception apparatus (cDRA) framework which upholds poly-directional example variety trademark. The utilization of BWD and consecutive game plans of cDRA is a portion of the highlights. The cDRA is arranged upon a stratum through a mutual conviction plane. The upper side has 2 oppositely defied cDRA that stimulated in port-1 and 2 utilizing co-planar waveguide (CPW) dealt with conformal strip lines. Comparative cDRA direction additionally exists at absolute lower side and energized in microstrip lines took care of conformal strip lines through port-3 and port-4 separately. To improve detachments and association, there are equivalent stage tendency BWDS which are put at various side of the stratum. Appropriately, for arranged radiation, plans are gained from 4-port which ensure multidirectional plan assortment and properties of proposed radio wire. The proposed method offers very 9.8 dB and 74% improvement in seclusion and connection esteems individually. Thus, in noticing the deliberate outcome, it is attested that the proposed structure is needful for the WLAN path (5.20–5.40 GHz) applications.

Keywords Poly-input-different yield (PIPO) · Isolation · Multi-port · BWDS · WLAN

1 Introduction

The fast advancement in remote correspondence and innovation which requests unrivalled rate and dependability inside the restricted range. To fulfil the requests, PIPO innovation displays up as an arising innovation that is used in a few remote principles in the present day. This innovation utilizes numerous receiving wires inside the

Ramkrishna (✉)
Department of Electronics and Communication Engineering, AIACTR, GGSIPU, Delhi, India
e-mail: yadavramkrishna@gmail.com

R. S. Yaduvanshi
Department of Electronics and Communication Engineering, NSUT, Delhi, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_15

transmitter and beneficiary end to improve the connection unwavering quality. The information rate is without the need for extra force or range [1].

Lately, DRA specialists are occupied with planning PIPO receiving wire for microwave and millimetre wave recurrence groups. Due to the fact that, they require a different preferred positioning than a small strip fix receiving wire, such as having a larger data transmission capacity, a higher radiation productivity, a higher addition, a greater ability to exert force, and so on [2]. A 4-port poly input poly output (PIPO) DRA reception apparatus has display up in writing [3–5], and the greater part of those radio wires utilizes a few segregated improvement procedures like symmetrical mode age [3], the use of stole away ground structure [4], metallic strip or parasitic parts [5] or blend of each one method [6]. The work fundamentally in focus to port coupling upgrade between the receiving wire component. The greater part of those examination articles utilizes the S-boundary-based articulation [7] for envelope coefficient of relationship (ECC) figuring. However, inside the instance of a DRA, it is realized that it emanates because of the excitation of a specific model, and each DR mode incorporates a special field plan. Normal coupling between 2 DRAs is especially in the z1 coupling and that we can say that a lot of the common coupling between DR-based PIPO gathering mechanical assembly is a direct result of radiation layout covering. The shared coupling between the components of the reception device is usually reduced at that point in the planned radiation chart if it is spatially separated. However, the writing shows that S-limit based enunciation is incorrect for ECC tallies [8]. Thus, ECC computation field-based verbalization gives certified characteristics because it accounts the radiation chart of various gathering mechanical assembly parts which is clear inside the principal equation of the ECC [8].

Given the wonders, a couple of exploration articles are accessible in writing where radiation graph is isolated spatially to upgrade shared coupling furthermore as ECC [9–11]. It demonstrates that the de-relating radiation diagram of radio wire segments consistently improves the imprisonment and ECC respect.

With the points of view, 4-port PIPO radio wire proposed during this paper where gathering mechanical assembly emanates are separated spatially recognizes high detachment and low relationship. To get this, 2-stage slant BWDSs are used and put on the 2 sides of the stratum as the superstrate. Inside given accepting wire, 2cDRA is situated at top of the stratum and another 2 are set at its base. With this exceptional procedure, the radio wire accomplishes every 20 dB of seclusion and fantastic ECC esteems (under 0.1).

2 Arrangement and Analysis of Proposed Antenna

To acknowledge the inclined pillar during a corresponding course, 2-stage slope BWDS which goes about as a halfway reflecting surface (PRS) is utilized. A Fabry-saucy (FP) pit radio wire is demonstrated in putting the PRS at a specific good way from the receiving wire component [12]. The PRS might be a mix of several unit cell, and math of a unit is display in Fig. 1.

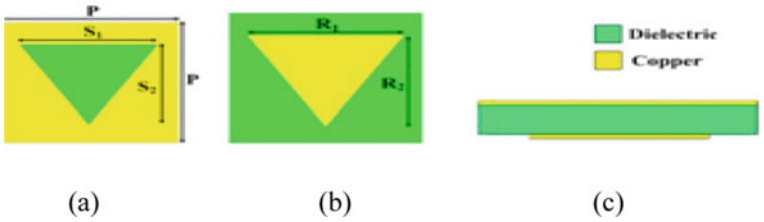


Fig. 1 Arrangement of unit cell **a** upper side view, **b** base view, **c** side view ($P = 15.75$ mm, $S_1 = 12.12$ mm, $S_2 = 10.5$ mm, R_1 and R_2 are variable)

Plan and recreation of the proposed unit are cultivated in utilizing Ansoff HBWDS in considering appropriate limit conditions. The proposed number of unit cell might be integrated structure because it includes capacitive and inductive grids. Total base side and upper side contain three-sided patches and three-sided hole independently that are termed as inductive and capacitive cross section [13]. These 2 networks are imprinted in a Fr for stratum ($\varepsilon_{\text{sub}} = 4.4$) having 0.16 cm thickness. PRS includes a couple of unit cell, and therefore, column is slanted and may be varying the reflection time of the PRS [13]. Consistent stage variety is regularly accomplished in 2 different ways, for the model in changing the structure of the capacitive organization or in changing the components of the inductive system [14]. Inductive lattice is different for changing S_1 and S_2 values; correspondingly, the capacitive matrix shifts with changes in R_1 and R_2. Here, the capacitive network differs to understand the persistent stage variety while having an immovable inductive matrix. Figure 2 displays the degree and stage assortment of the proposed unit of various R_1 and R_2 values in the middle of 5.15–5.35 GHz. It is frequently seen from Fig. 2 that the move in thunderous recurrence is made thanks change in R_1 and R_2 values. The unit has a distinctive size and stage qualities for different R_1 and R_2 value. The unit of the PRS picked in such a manner that there will be consistent stage assortment. With this reference, PRS-1 is suggested that unit of most outrageous reflection size (cell-1) is found at 1 side, and consequently, different cell and progressively diminishing reflection greatness are put on contrary side. Figure 3a displays the most noteworthy and base perspectives on PRS-1. It uncovers most elevated 3-sided opening (inductive matrix) which is the same for all cell yet absolute down 3-sided fix (capacitive network) changes to acknowledge constant greatness and stage variety. PRS1 comprises 2×4 unit cell.

In the proposed receiving wire structure, the reception apparatus bar is inclined in another manner. To understand this, we have to make symmetric PRS, altogether that symmetric beams are going to be formed in another way. We have to deal with the symmetricity to acknowledge equivalent branch power proportion, in any case, the PIPO execution will be corrupted. To comprehend this marvel, given PRS is made with blend of 2 PRS-1 as displayed up in Fig. 3b. Inside given PRS, centre unit (cell-1) is viewed as a standard cell. The component of R_1 and R_2 changes ceaselessly from a standard unit (cell-1) to left bearing or right heading. By Figs. 2 and 3b, it is consistently observed that unit (cell-1) the most extraordinary reflection degree regard as set at focal point of given PRS. The opposite unit cell with gradually

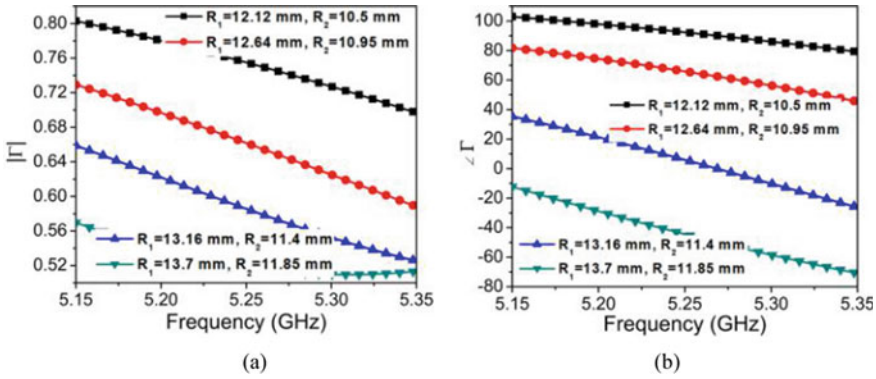


Fig. 2 Magnitude and stage reaction of the proposed unit cell **a** extent reaction; **b** stage reaction

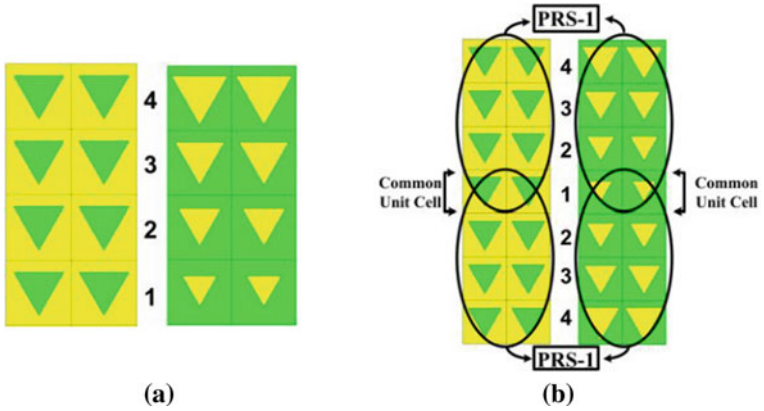


Fig. 3 **a** Top view and base view of the proposed PRS-1, **b** design of proposed PRS

decreasing reflection degree is put towards sting of PRS. This proposes unit cell-4 which is put at the sting of PRS in both the bearing because it is the base reflection size esteem. The evenness is accomplished inside the proposed PRS. To make the inclined shaft design, the size and stage appropriation are changed in grading the 3-sided fix. The contrast between 3-sided fix measurements of neighbouring areas is chosen in the degree consistent (ΔR). The size of the 3-sided patches is framed utilizing [14]:

$$(R_2)_{cell-i} = (R_2) + i * \Delta R_2$$

$$(R_1)_{cell-i} = (R_1) + i * \Delta R_1$$

Table 1 Different unit cell utilized in PRS

Unit Cell No.	R_2 (mm)	R_1 (mm)	Magnitude	Phase (°)
1	10.5	12.12	0.75	92.21
2	10.95	12.64	0.65	63.74
3	11.4	13.16	0.58	5.64
4	11.85	13.7	0.51	−44.70

where R_1 and R_2 are the 3-sided fix measurements for references of unit (cell-1) and that we need to address that the unit no. For given PRS, degree constant is picked as 0.45 and 0.52 for ΔR_2 and ΔR_1, independently.

Table 1 displays enormity and stage assessments of the different unit cell of PRS at the 5.25 GHz. Figure 2 and Table 1 reveal size and stage esteems declines as size of 3-sided fix increments.

The plan detail of 4-port PIPO reception apparatus cavity framework has appeared in Fig. 4. 2 tube-shaped dielectric resonators name DR-1 and several (εcDRA = 9.8) are put of most noteworthy side of FR4 stratum (εr, sub = 4.4) appeared in Fig. 4. That cDRA is put inverse to each other. Port-1 and several comprises CPW taken care of based conformal strip lines. DR-1 and 2 or 3 are energized with the help of port-1 and a few, separately. The suggested 2 × 7 cell-based PRS discussed in the past portion arranged on this 2cDRA and goes probably as a superstrate. Space (T) among stratum and PRS area unit 0.26 λ at 5.25 GHz. The projected PRS is placed in such however that the unit that has a most outrageous reflection extent is put in the middle of DR-1 and 2 or 3. Accordingly, receiving the wire shaft is inclined in another way [15]. The between component separation between DR-1 and several ought to be kept up in such manner that each cDRA is cheerful broadside way. The between component dispersing (S) between the most noteworthy cDRA is kept in 14 mm which satisfactory to 0.24 λ at 5.25 GHz. This is frequently base separation at which most elevated cDRA is energized broadside way. Bar has digressed from the broadside heading if space is decreased, and high-shared coupling is gotten between the radio wire components. Another 2 oppositely went up against cDRA to be explicit DR-3 and 4 is set in unquestionably base side of stratum. They are stimulated through a micro-strip lines-based conformal strip lines employing port-3 and 4 separately. Port-3 and 4 are utilized symmetrically concerning port-1 and several. Therefore, symmetrical modes are energized among the upper and base DRAs. Essentially, another given PRS is put at without a doubt base side of stratum which goes probably as a superstrate for DR-3 and 4 as appeared in Fig. 4. Absolute lower side of PRS is put symmetrically regarding top PRS. The between component separation between base cDRA is kept the same because of the main 1.

Calculation for bandwidth for the first circuit

$$V_s = nV_i = nV_m \sin wt$$

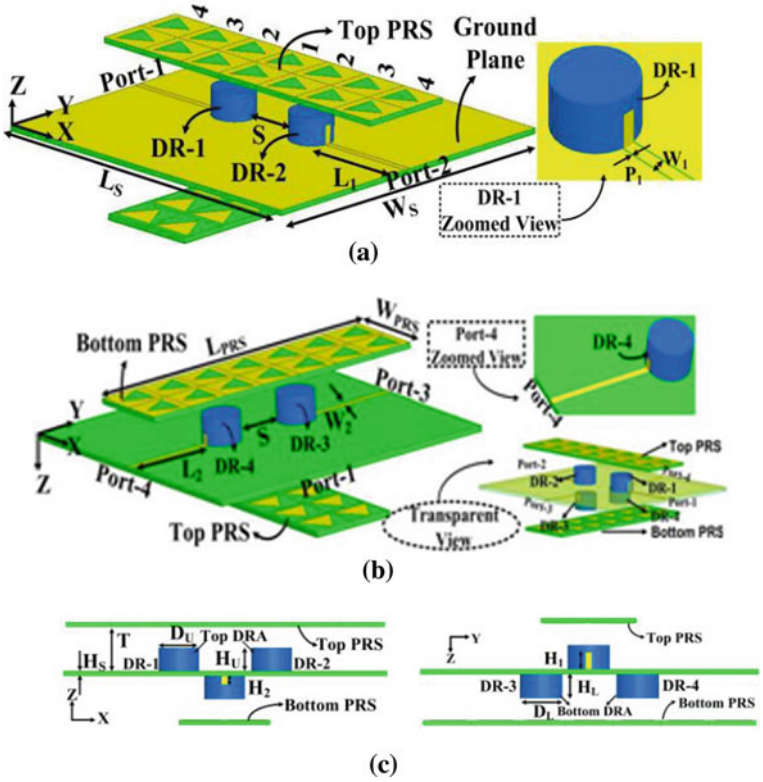


Fig. 4 Configuration of given PIPO radio wire; **a** upper read (3D), **b** lower read (3D), **c** prospect (different measurements are: $L_S = W_S = 112$, $L_1 = 15$, $H_S = H_{PRS} = 1.6$, $W_1 = 2$, $P_1 = 0.5$, $H_1 = 5.5$, $S = 14$, $L_2 = 15$, $W_2 = 2$, $H_2 = 3.1$, $L_{PRS} = 110$, $W_{PRS} = 31.5$, $D_U = D_L = 14$, $H_U = H_L = 8$, $T = 15$ mm, all are in mm)

$$v_s(t) = \frac{i_s(t)}{y_s},\ v_s(t) = \frac{l_{in}}{n}\frac{1}{\frac{1}{R_s} + \bar{j}\left(\omega C s_{\pi_s}^{-1}\right)}.$$

$$v_s(t) = \frac{I_m \sin \omega t}{n\left[\frac{1}{R_s} + \bar{j}\left(\mathrm{wC}_s - \frac{1}{\omega L_s}\right)\right]};\ \frac{I_m}{n} < 0.$$

$$|v_s| = \frac{\frac{I_m}{n}}{\sqrt{\left[\frac{1}{R_s^2} + \left(\mathrm{wC}_s - \frac{1}{\omega L_s}\right)^2\right]}}$$

At $\omega c_s - \frac{1}{wL_s} = 0$ then $|v_s|$ max.

For $\frac{|v_s|_{\max}}{\sqrt{2}}$, two frequency is present, that is called cut-off frequency and ω_1 is called lower cut-off frequency and ω_2 is called high cut-off frequency.

$$\mathrm{wc}_s - \frac{1}{\mathrm{wL}_s} = -\frac{1}{R_s} \tag{1}$$

$$\mathrm{wc}_s + \frac{1}{\mathrm{wL}_s} = \frac{1}{R_s} \tag{2}$$

For lower cut-off frequency, put $\omega = \omega_1$ in Eq. 1

$$w_1 c_s - \frac{1}{w_1 L_s} = -\frac{1}{R_s}$$

$$\omega_1^2 L_s C_s + w_1 L_s G_s - 1 = 0$$

By solving the above equation

$$w_1 = -\frac{G_s}{2c_s} \pm \sqrt{\left(\frac{G_s}{2c_s}\right)^2 + \frac{1}{L_s C_s}}$$

Similarly, for higher cut-off frequency put $\omega = \omega_2$ for Eq. 2 then

$$\omega_2^2 L_s C_s - w_1 L_s G_s - 1 = 0$$

$$w_2 = -\frac{G_s}{2c_s} \pm \sqrt{\left(\frac{G_s}{2c_s}\right)^2 + \frac{1}{L_s C_s}}$$

Summarize Eqs. 1 and 2

$$w_1 = -\frac{G_s}{2c_s} + \sqrt{\left(\frac{G_s}{2c_s}\right)^2 + \frac{1}{L_s C_s}}$$

$$w_2 = \frac{G_s}{2c_s} - \sqrt{\left(\frac{G_s}{2c_s}\right)^2 + \frac{1}{L_s C_s}}$$

$$\text{For bandwidth (BW)} = w_2 - w_1 = \frac{G_s}{2c_s} \pm \sqrt{\left(\frac{G_s}{2c_s}\right)^2 + \frac{1}{L_s C_s}} + \frac{G}{2C_s} - \sqrt{\left(\frac{G_s}{2c_s}\right)^2 + \frac{1}{L_s C_s}}$$

$$\mathrm{BW} = \frac{G_s}{C_s} = \frac{1}{R_s C_s} = \text{bandwidth}$$

$$\omega_r = \frac{1}{\sqrt{L_s c_s}}$$

$$\frac{G_s}{C_s} = \frac{1}{R_s C_s} = \text{bandwidth} = \text{bandwidth} = \frac{1}{\tau_s}$$

$$\text{Quality factor}\,(Q_s) = R_s\sqrt{\frac{c_s}{\text{Ls}}}$$

$$Q_s = R_s w_s C_s$$

bandwidth $=\frac{1}{\tau_s}$ noticed. On the contrary hand, low confinement esteems are seen between port-(1, 2) and port-(3, 4) due to the high front of radiation shaft. With extension of PRS, radiation light outflow segment is slanted from the main side course. Figure 7 displays three-dimension polar plot of the proposed getting wire in various port. It is oftentimes displaying that DR-one and DR-two is slanted towards -350 and $+\,350$ independently in XZ-plane. On the opposite DR-three and four are slanted to -1450 and $+\,1450$ independently in YZ-plane. Consequently, each cDRA of the proposed radio wire transmits a few ways which affirm the multidirectional qualities of the proposed receiving wire. Subsequently, the detachment between port-(1, 2) and port-(3, 4) is improved through 10–12 DB. Figure 5 revealing that division level between port-(1, 2) and port-(3, 4) is 12 dB without PRS. However, in wake of adding PRS, the division level between the port is best than 20 DB. Consequently, proposed PRS clearly displays that adequacy of those techniques. In any case, we noticed the presence of some side-flap level inside the PRS case.

On the contrary hand, DR-3 and 4 are energized inverse main sideway ($\theta = 1800$ and $\phi = 00$). Consequently, their radiation emanates are disconnected spatially also as in symmetrical mode which is created in the middle of them. Therefore, high separation esteems are taken care of.

The dispersing boundary of proposed 4-port PIPO reception apparatus with and without PRS is confirmed up in Fig. 5. It well-known displays that without PRS, information move limit of the suggested radio wire is 7.5% (5.1–5.5 GHz). Yet, data transfer capacity is decreased to 3.5% (5.15–5.35 GHz) which is the point at which the PRS is to be put.

DR-(1 and 2) and DR-(3 and 4) are set backwards to 1 another and in this way, a comparative model is happy inside the cDRA because of comparable taking care of the instrument, for example, HE11δ mode. Thus, the solid common coupling is produced in the middle of port-(1and 2) and port-(3and 4). The restriction responses between the port has display up in Fig. 5. Without PRS, low detachment regards (25 DB) are seen between port-(1, 3), (1, 4), (2, 3) and (2, 4). Improvement in a division between the port is much of the time clarified in seeing the three-dimension polar plot of the gathering contraption parts. Figure 6 indicates 3D polar plot without PRS. It exhibits that DR-1 and a couple are stimulated main sideway ($0 = 00$ and ϕ).

Above this diagram of multi-input to multi-output, 4 DRA connected in this manner and 1 DRA simply behaves like a parallel RLC circuit.

$$(Q_s) = R_s\sqrt{\frac{c_s}{Ls}}$$

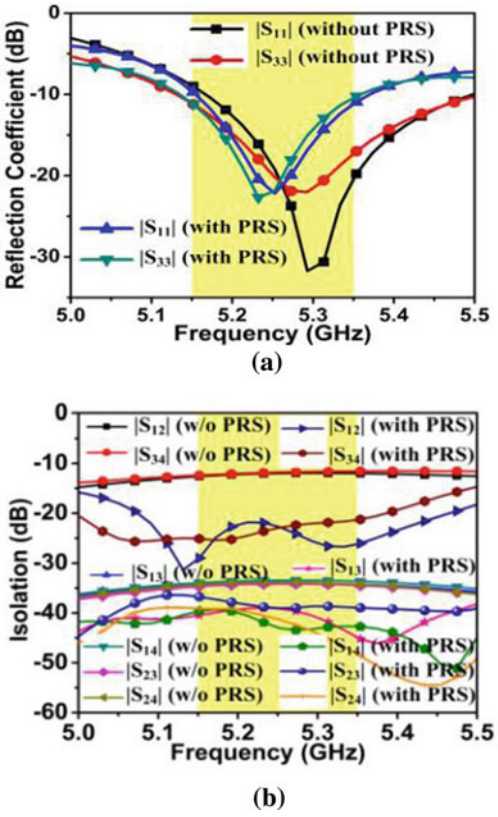


Fig. 5 Scattering parameters of the proposed antenna; **a** reflection coefficient, **b** isolation between the ports

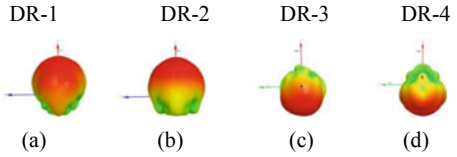


Fig. 6 Three-dimension polar plots of antenna without PRS at 5.3 GHz; **a** port-1, **b** port-2, **c** port-3, **d** port-4

$$\text{bandwidth} = \frac{1}{R_s C_s} \;\&\; w_s = \frac{1}{\sqrt{L_s c_s}}$$

The dispersing boundaries of proposed 4-port PIPO reception apparatus with and without PRS are confirmed up in Fig. 5. It well-known displays that without PRS, information move limit of the suggested radio wire is 7.5% (5.1–5.5 GHz). Yet, data transfer capacity is decreased to 3 0.5% (5.15–5.35 GHz) which is the point at which the PRS is to be put.

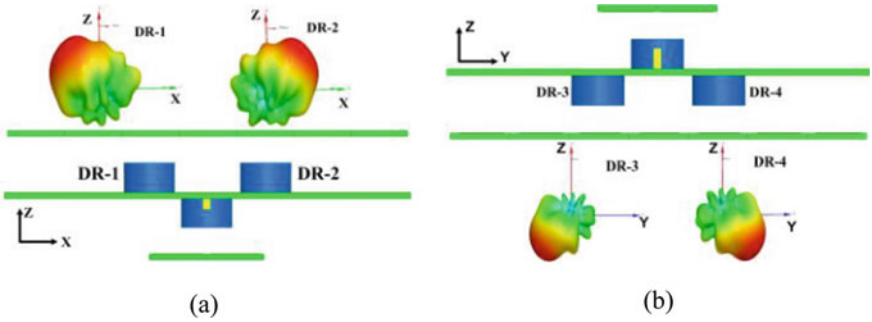


Fig. 7 Three-dimension polar plot of the reception apparatus at 5.25 GHz with PRS; **a** port-1, and 2, **b** port-3 and 4

3 Experimental Results

To oblige stratum of PRS in both sides and multistrand materials which is foam (εr < 1.05) is fabricated in a 4-port MIMO antenna is display in Fig. 8.

The estimated dispersing boundaries of reception apparatus are display in Fig. 9. Figure 9a uncovers given shape covers the recurrence range among 5.15 GHz to 5.42 GHz for all of the port. Figure 9b it additionally displays the deliberate disengagement reaction of the given reception apparatus and various port. It displays > 20 dB separation of all the port.

The re-enacted and estimated radiation graph of the proposed different port structure as in Fig. 10, it uncovers graph of radiation of each DR-components which are isolated spatially. The bar peaks of port-1 and port -2 are inclined to −350 and + 350 individually in XZ-plane. On other hand, most extreme pillar bearing of port-3 and port-4 is at −1450 and + 1450 separately in YZ-plane.

This marvel is also clarified in Fig. 7 in noticing the 3D polar plot of the antenna. The envelope coefficient of relationship (ECC) bend of given MIMO gathering contraption with and without PRS has appeared in Fig. 11 between port-(1 and 2) and (3 also 4). This MIMO execution boundary is determined in utilizing complex far-field design examined in [16]. It uncovers that with and without. PRS ECC esteem is

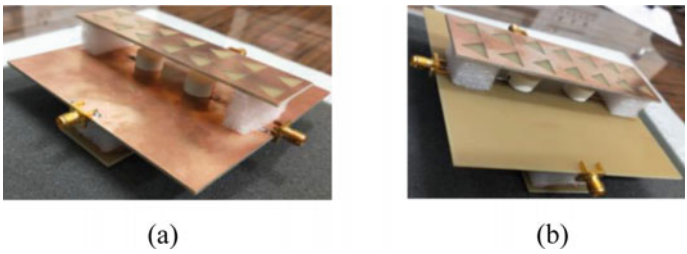


Fig. 8 Fabricated antenna structure; **a** top view; **b** bottom view

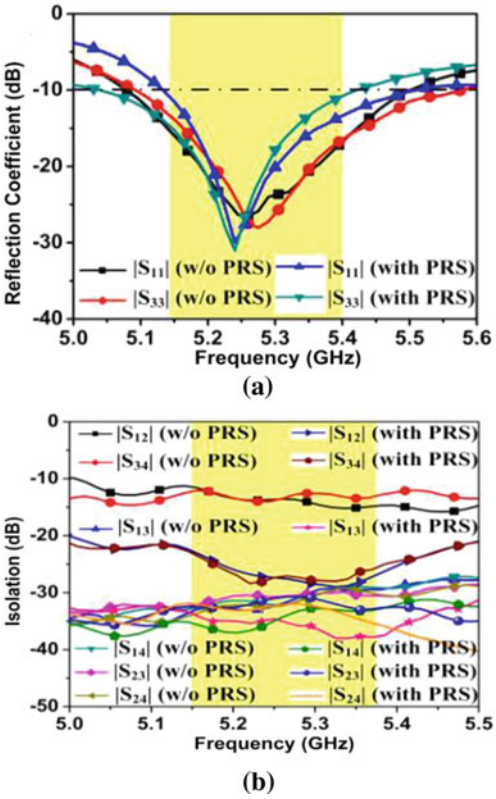


Fig. 9 Measuring dispersing boundary of given reception apparatus; **a** Reflexion coefficient (S11), **b** disengagement variety

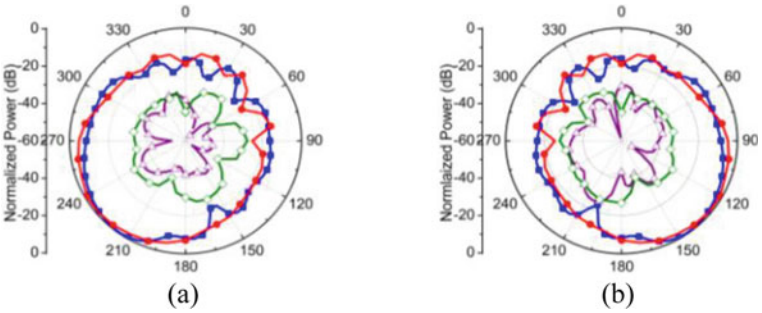


Fig. 10 Examples of radiation of reception apparatus; **a** Port-1 (X-Plane), **b** port-2 (X-Plane)

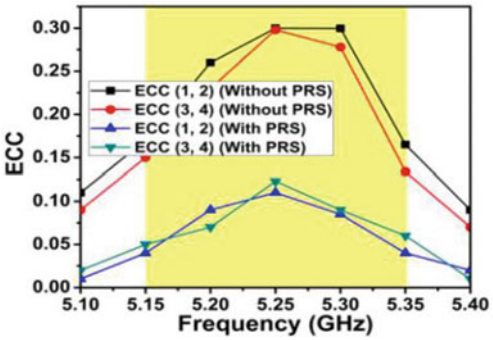


Fig. 11 ECC of proposed antenna

Table 2 Comparison of performance of proposed antenna with another antenna

Reference	Covered bandwidth (GHz)	Isolation (dB)	No. of port	Isolation improvement technique	ECC level
[9]	5.6–5.9	17	8	DR placement	‹0.2
[1]	5.6–5.9	20	4	Back to back DR placement	‹0.15
[1]	29.5–31	30	2	Special feeding technique	‹0.15
This work	5.15–5.35	22	4	Use of PRS	‹0.11

0.31 and 0.10 individually at 5.25 GHz. Thus, this PRS gives every 75% improvement in ECC esteems. Proposed strategies are utilized not exclusively to upgrade disconnection; however, it additionally improves ECC esteems between the receiving wire components. Addition and radiation profitability of given gathering contraption is 4.9 dB, 89% and 7.2 dB, 81% without and with PRS exclusively. Table 2 displays exhibition correlation of gave structure other DRA-based MIMO reception apparatus. It is uncovered that proposed radio wire gives an improved disengagement levelled with great ECC esteem.

4 Conclusions

The decoupled 4-port MIMO is in a spatially gathering contraption which is an application in WLAN. Utilization of consecutive PRS helps in the progress of 10 dB of seclusion and 75% ECC is accomplished. The introduced structure can be helpful for WLAN applications (5.15–5.35 GHz).

References

1. Zhou Q, Dai H (2006) Joint antenna selection & link adaptation for MIMO systems. IEEE Trans Veh Technol 55:243–255
2. Petosa A (2007) Dielectric resonator reception apparatus. Artech House, H&Book, Norwood, MA, USA
3. Nasir J, Jamaluddin MH, Khalily M, Kamarudin MR, Ullah I, Selvaraju R (2015) A diminished size double port MIMO DRA with high segregation for 4G applications. Int J RF Microw Comp Support Eng 25:495–501
4. Nasir J, Jamaluddin MH, Khalily M, Kamarudin MR, Ullah I (2016) Plan of a MIMO dielectric resonator antenna for 4G applications. Wirel Pers Commun 88:525–536
5. Sharma A, Das G, Gangwar RK (2016) "Double spellbound triple-b& cross breed MIMO round & hollow dielectric resonator reception apparatus for LTE2500/WLAN/WiMAX applications. Int J RF Microw Comp Supported Eng 26:763–772
6. Khan AA, Jamaluddin MH, Aqeel S, Nasir J, Kazim JR, Owais O (2017) A dual-B & MIMO dielectric resonator antenna for WiMAX/WLAN applications. IET Microw Antennas Propag 11:113–120
7. Whiten S, Romeu J, Corbella I (2003) Careful portrayal of reception apparatus framework variety execution from input boundary depiction. IET Electron Lett 39:705–707
8. Sharawi MS (2017) Current abuses & future possibilities for printed different info, numerous yield reception apparatus frameworks. IEEE Antennas Propag Mag 59:162–170
9. Das G, Sharma A, Gangwar RK, Sharawi MS (2018) Minimized consecutive DRA-based 4-port MIMO receiving wire framework with bi-directional variety. IET Electron Lett 54:884–886
10. Das G, Sahu NK, Sharma A, Gangwar RK, Sharawi MS (2019) Dielectric resonator-based 4-comp1nt eight-port MIMO reception apparatus with multi-directional example variety. IET Microw Antennas Propag 13:16–22
11. Sharawi MS, Podilchak SK, Hussain MT, Antar YMM (2016) Dielectric resonator-based MIMO radio wire framework empowering millimetre-wave cell ph1s. IET Microw Antennas Propag 11:287–293
12. Ratni B, Merzouk WA, de Lustrac A, Villers S, Piau GP, Burokur SN (2017) Plan of phase-modulated Metasurfaces for Beam Steering in Fabry–Perot Cavity Antennas. IEEE Antennas Wirel Propag Lett 16:1401–1404
13. Nakano H, Mitsui S, Yamauchi J (2014) Inclinesd beam high gain antenna system composed of a patch antenna & periodically displayed loops. IEEE Trans Radio Wires Propag 62:2917–2925
14. Qin F, Gao S, Mao C, Wei G, Xu J, Li J (2015) Low-profile high-gain inclinesd bar Fabry-Perot radio wire. In: Proceedings of ninth European conference on receiving wires propagation (EuCAP), pp 1–5
15. Hassan T, Khan MU, Attia H, Sharawi MS (2018) A BWDS based correlation reduction technique for MIMO antennas. IEEE Trans Receiving Wires Propag 66:4900–4905
16. Sharawi MS (2014) Printed MIMO antenna engineering. Artech House, Norwood, MA, USA

Design of U-Shaped MIMO Antenna for Wi-Fi Application

Om Prakash and Prabina Pattanayak

Abstract The capacity of a wireless system can be improved by using several antennas on both the sending and receiving sides. Multiple-Input–Multiple-Output is a system that uses multiple antennas to meet the requirement for larger data rates. A four-element compact Multiple-Input-Multiple-Output (MIMO) antenna is suggested in this research. The antenna is made up of four U-shaped patch components and has a maximum gain of 7.8 dB at 5 GHz. The proposed antennas are configured on a software platform using HFSS simulators. To address the security issue of Wi-Fi application, the suggested MIMO antenna has 10 dB impedance and a restricted bandwidth of 50 MHz. The isolation between the ports of adjacent antenna is more than 15 dB.

Keywords MIMO · HFSS · Wi-Fi application · Isolation · Return loss

1 Introduction

MIMO systems, which are capable of transmitting and receiving signals of the same power level in parallel channels, are employed to meet the growing demands of high data rates in current communication systems. This is due to the fact that MIMO systems can enhance channel capacity on both the sending and receiving sides without adding bandwidth or transmission power. The potential for MIMO systems to improve wireless mobile communications reliability and channel capacity has piqued interest [1–3]. Multiple-Input–Multiple-Output (MIMO) technology is gaining popularity as a way to improve the performance of wireless communication systems [4, 5]. It improves line-of-sight (LOS) and non-line-of-sight (NLOS) propagation dependability by boosting channel capacity, enhancing coverage, and increasing reliability

O. Prakash (✉) · P. Pattanayak
Department of Electronics and Communication Engineering, National Institute of Technology, Silchar, Assam, India
e-mail: om21_rs@ece.nits.ac.in

P. Pattanayak
e-mail: prabina@ece.nits.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_16

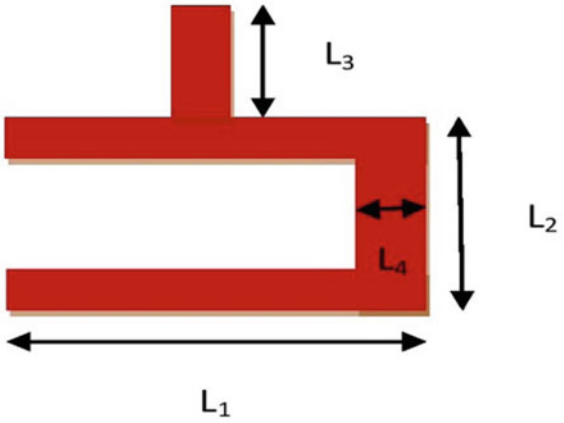


Fig. 1 Dimension of U-shaped patch

[2, 3]. Mutual coupling, reflections, envelope correlation coefficient, diversity gain, and efficiency are all factors to consider while building MIMO antennas [6, 7]. It is worth noting that the type of substrate utilized has a significant impact on the aforementioned MIMO antenna parameters. The most popular substrates for MIMO antennas are the inexpensive FR4 and less lossy stiff laminates from Rogers and Taconic.

2 U-Shaped MIMO Antenna Design

The proposed U-shaped MIMO antenna for 5 GHz shown in Fig. 2 and Fig. 1 shows dimension of U-shape structure, where $L1 = 19$ mm, $L2 = 4$ mm, $L3 = 4$ mm and $L4 = 2$ mm. The selected substrate FR4 material (dielectric constant $(\varepsilon_r) = 4.4$ and loss tangent $(\delta = 0.02)$) has a thickness of 1.6 mm. This U-shaped MIMO antenna is divided into four parts; all the four parts have similar U-shaped antenna with dimension 19×4 mm^2. Ground and substrate dimension is same, i.e., 76×58 mm^2. The thickness of patch and ground is taken 0.05 mm and 0.035 mm, respectively. The antenna is mounted on the dielectric material's top surface (FR4). Copper conductors are used for the conductive patches and the ground.

3 Results and Discussion

The simulation result for S11 are presented in Fig. 3 for 5 GHz. Antenna for 5.GHz confirms -10 dB impedance bandwidth of 50 MHz. The peak gain and the efficiency for at 5 GHz antenna are obtained as 7.8 dBi and 42.12%, respectively. Poor efficiency may be due to the use of 1.6 mm FR4 material as antenna substrate. The isolation

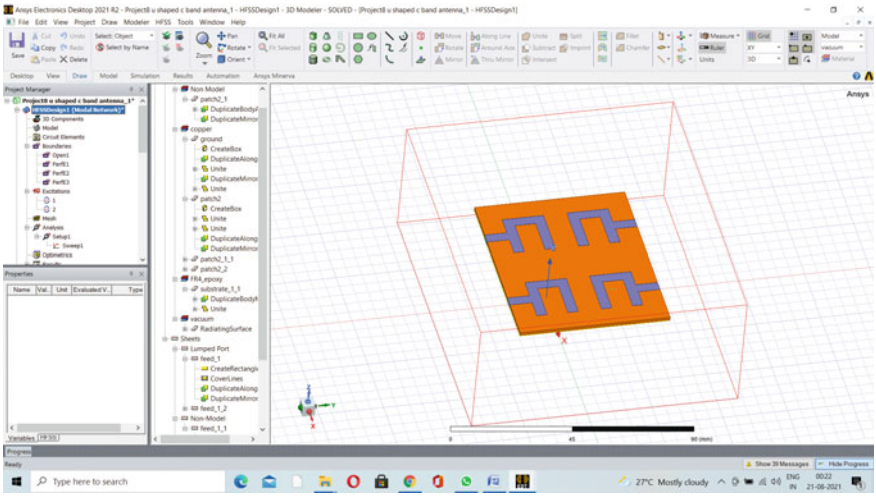


Fig. 2 Graphical design of U-shaped MIMO antenna

between the ports of adjacent antenna is more than 15 dB. To address the security issue in Wi-Fi application, the proposed antenna has narrow bandwidth of 50 MHz.

The VSWR of simulated U-shaped MIMO antenna is 2.4 at a resonating frequency of 5 GHz as shown in Fig. 4.

The simulated E-plane and H-plane radiation patterns for 5 GHz antenna as shown in Fig. 5 ensure satisfactory coverage of surrounding region.

3D gain and radiation pattern is shown in Figs. 6 and 7, respectively. The maximum gain obtained for the proposed antenna is 7.8 dBi.

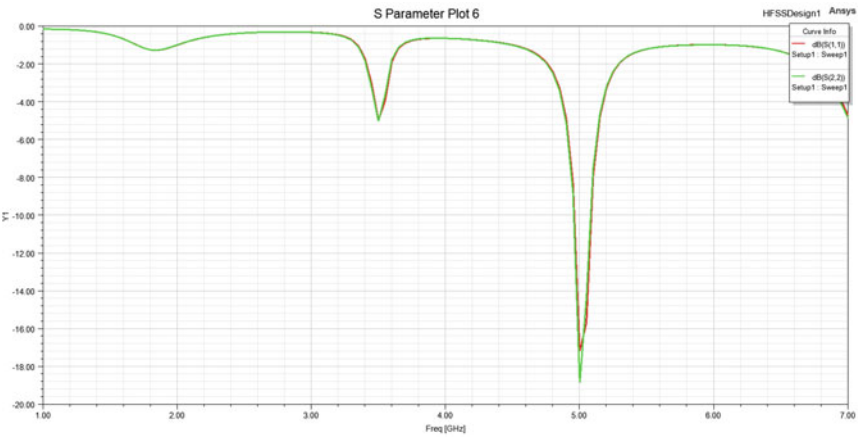


Fig. 3 Return loss at 5 GHz

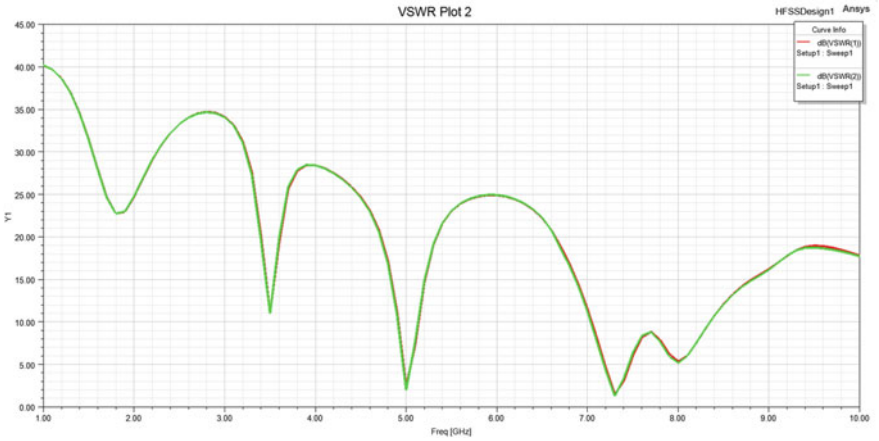


Fig. 4 VSWR U-shaped MIMO antenna at 5 GHz

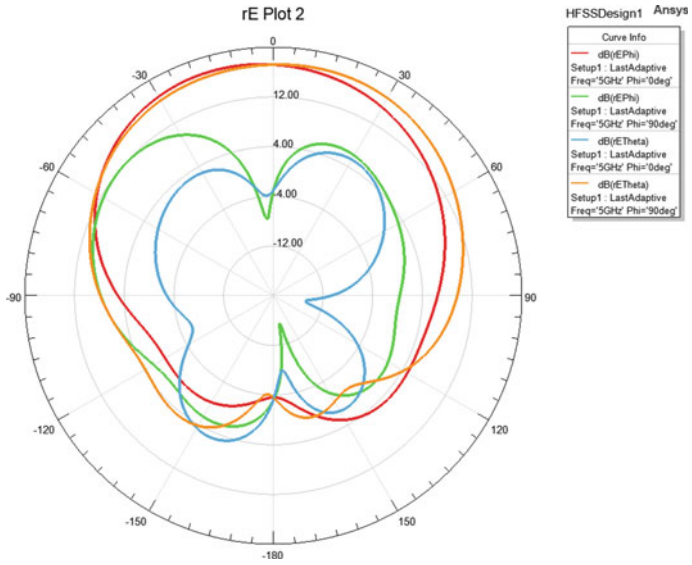


Fig. 5 E- and H-plane patterns for 5 GHz

4 Conclusion

This study looks at a 5 GHz Wi-Fi application with a U-shaped MIMO antenna. The performance of a U-shaped MIMO antenna made of FR4 substrates was discussed. At 5 GHz, a complete description of antenna geometry was published. The results reveal that the suggested U-shaped MIMO antenna provides adequate simulation

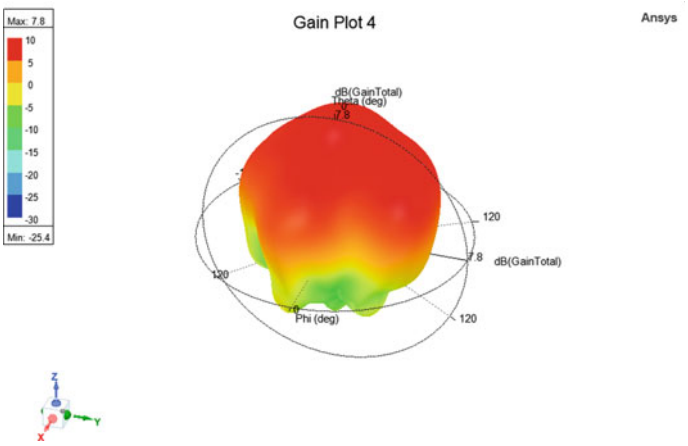


Fig. 6 3D gain pattern at 5 GHz

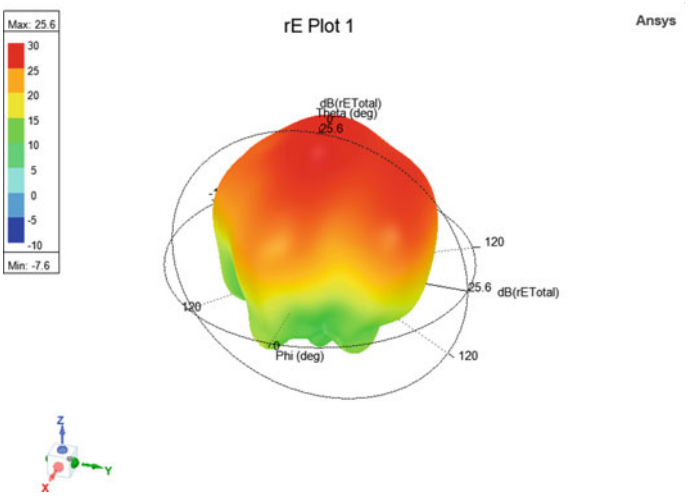


Fig. 7 3D radiation pattern at 5 GHz

and efficiency for the intended frequency bands. The physical construction of this antenna can be coupled to embedded electronics, allowing it to be used in a number of wireless communication applications.

References

1. Huang J, Dong G, Cai J, Li H, Liu G (2021) A quad-port dual-band MIMO antenna array for 5G smartphone applications. Electronics 10:542–551
2. Ghosh S, Tran T-N, Le-Ngoc T (2014) Dual-layer EBG-based miniaturized multi-element antenna for MIMO systems. IEEE Trans Antennas Propag 62(8):3985–3997
3. Zhang S, Zetterberg P, He S (2010) Printed MIMO antenna system of four closely-spaced elements with large bandwidth and high isolation. Electron Lett 46(15):1052–1063
4. Gautam B, Verma P, Singha A, Islam H, Prakash O, Das S (2019) Design of multiple collar stay antennas for wireless wearable compact devices. Microw Opt Technol Lett 62(2):743–749
5. Verma P, Prakash O (2020) C-shaped radiator structure antenna for wireless communication. In: Bera R, Pradhan PC, Liu CM, Dhar S, Sur SN (eds) Advances in communication, devices and networking. ICCDN. Lecture notes in electrical engineering, vol 662, pp 473–477. Springer, Singapore. https://doi.org/10.1007/978-981-15-4932-8_53
6. Xiang Z, Quan X, Li R (2012) A dual-broadband MIMO antenna system for GSM/UMTS/LTE and WLAN handsets. IEEE Antennas Wireless Propag Lett 11:551–554
7. Allen B, Malik WQ, Smith PJ, Edwards DJ (2006) Demystifying MIMO. Commun Eng 4(6):38–42

Comparative Study of Parametric Disturbances Effect on Cart-Inverted Pendulum System Stabilization

Neelam Verma and Sudarshan K. Valluru

Abstract The main motivation behind the present work was to validate the impact of pendulum mass, cart mass, and length of pendulum on stabilization and swing-up of cart-inverted pendulum. Inverted pendulum system is a classic benchmark example of underactuated electro-mechanical systems, where the degree of freedom is less than the number of inputs available. PID controller was designed for a set of parameter values of the system. MATLAB simulation was carried out for different values of pendulum mass (0.2–0.25 kg), pendulum length (0.3–0.34 m), and cart mass (2.3–2.45 kg). The performance of considered cart- inverted pendulum system is demonstrated with simulation results for a different set of results, and a comparative analysis was made to comment on their effects on swing-up and stabilization.

Keywords Underactuated mechanical systems · Cart-inverted pendulum · Degree of freedom · PID controller · Stabilization · Swing-up

1 Introduction

Cart Inverted-Pendulum is one of the typical benchmark examples of underactuated systems, as it has inherent unstable nonlinear behaviors with minimum phase complexity [1]. Cart inverted pendulum system provides many challenges while designing optimal and robust controllers. The main reason behind its extensive research is that many upcoming engineering systems can approximately be modelled as these systems. The examples are rocket guidance system, robot system, wheel balancing motion systems, gantry crane system, biomechanics vehicle based on bipedal walking, etc. [2]. Underactuated systems belong to the categories of systems, which are minimum on passive joint (i.e., not control by actuators). And because of this reason, they cannot have control in all degree of motion. Dynamic response characteristics of underactuated systems differ from fully actuated system. And this difference arises the demand for dedicated controller designs, as systems designed

N. Verma · S. K. Valluru (✉)
Department of Electrical Engineering, Center for Control of Dynamical Systems and Computation, Delhi Technological University, Delhi, India
e-mail: sudarshan_valluru@dce.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_17

for fully actuated one, cannot be used for underactuated systems [3]. Therefore, with all these reasons, the problem of real-time tracking control and stabilization of inverted pendulum servomechanism becomes, one of highly significant dynamic control problem. It also requires fast and precise performance [4–7]. To address different challenge for stabilizing inverted pendulum precisely, many techniques are used in literature, for example, PID, nonlinear PID, fractional PID, hybrid control approach, LQR, energy control strategy, particle swarm optimization (PSO), fuzzy logic, neural network, sliding-mode control, etc. [8–10].

Nonlinear unstable inverted pendulum system exhibits some real-time problem while implemented in industry applications, for examples, behavior changing with changing operational conditions, external disturbance inputs and physical restrictions on some variables [11]. In literature, these systems are further classified into three categories. In first case, rotating platform carries a pendulum with a one degree of freedom (DOF) [6]. In second case, double and triple inverted pendulum having more than one degree of freedom, are being actuated by external force on a cart. And in third case, pendulum has two DOF mounted on a robot arranged with decoupled or weak coupling links [12].

System dynamics of the nonlinear plant based on inverted pendulum, can cause instability is its behavior if the sudden load mass change occurs. This load mass could have constant changes or time-dependent changes (changes vary with time) [13]. Mass of the cart can also be changed by adding additional load on cart, even after the system becomes stable. These cart mass changes could also vary with time [14]. There are two more dominating factors, pendulum length and rail length, which can modify swing-up accuracy and stability of cart-inverted pendulum [15]. The work in this paper, analysed the variation of cart inverted pendulum system parameters such as, pendulum mass, pendulum length, and cart mass which can alter the system dynamics. The remainder of the paper organized as follows: Sect. 2 gives model of an Inverted Pendulum on a Cart; Sect. 3 analyses the Simulations and Results and finally Sect. 4 as Conclusion.

2 Model of an Inverted Pendulum on a Cart

The Mathematical modeling of DC motor driven cart-inverted pendulum system is derived as below, with the help of its free body diagram [16]. Mass-less pendulum rod and friction-less hinge are assumed for derivation. The pendulum mass and cart mass are denoted by m and M, respectively [17].

System is operating under the influence of two forces, externally applied $u(t)$ and gravitational force (g). In Fig. 1, cart position is represented by $x(t)$ and pendulum angular displacement is represented by $\emptyset t$.

The basic parameters of considered system are given below in the Table 1.

Total externally applied force will be balanced by combined efforts of cart-mass acceleration force and equivalent x-direction pendulum acceleration force, which can be shown as below.

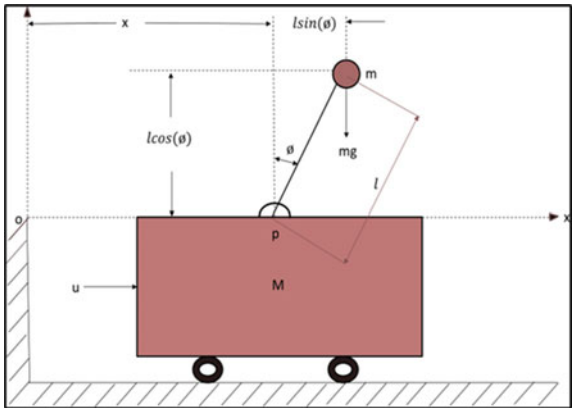


Fig. 1 Cart-inverted pendulum system

Table 1 Parameters of the inverted pendulum system

Notation	Parameter	Value	Units
M	Cart's mass	2.3	Kg
m	Mass of pendulum	0.2	Kg
L	Pendulum's length	0.3	m
B	Coefficient of friction	0.00005	Ns/m
I	Pole's Motion of inertia	0.0099	$Kg.m^2$
g	Gravity	9.8	m/s^2
u	Force on cart externally	1	N
x	Cart Displacement	+ 0.4 m to −0.4 m	m
$\dot{x}$	Linear velocity of the cart	–	m/s
Ø	Angular Displacement	π	rad
$\dot{Ø}$	Angular Velocity	–	Rad/sec

$$M\frac{d^2}{dt^2}x + m\frac{d^2}{dt^2}x_m = u \tag{1}$$

where equivalent displacement for pendulum mass in x and y direction will be

$$x_m = l\sin Ø + x \text{ and } y_m = l\cos Ø \tag{2}$$

By substituting values of Eq. (2) in Eq. (1), combined result will be

$$M\frac{d^2}{dt^2}x + m\frac{d^2}{dt^2}(l\sin Ø + x) = u \tag{3}$$

further simplification will lead us to

$$(M+m)\ddot{x} - (\mathrm{ml})\sin\emptyset\dot{\emptyset}^2 + \left(\mathrm{ml}\cos\emptyset\ddot{\emptyset}\right) = u \tag{4}$$

Now, we can find out the torque balance equations of the system by considering, resultant product of y component of a force and pendulum length, torque on the pendulum mass because of acceleration and torque because of gravitational force. The resultant equations will be

$$(u_x\cos\emptyset l) - \left(u_y\sin\emptyset l\right) = (mg\sin\emptyset l) \tag{5}$$

and these external force components are as follows

$$u_x = m\frac{d^2}{\mathrm{d}t^2}x_m = m\left[\ddot{x} - l\sin\emptyset\dot{\emptyset}^2 + l\cos\emptyset\ddot{\emptyset}\right] \tag{6}$$

$$u_y = m\frac{d^2}{\mathrm{d}t^2}y_m = -m\left[l\cos\emptyset\dot{\emptyset}^2 + l\sin\emptyset\ddot{\emptyset}\right] \tag{7}$$

substituting Eq. (6) and (7) in main torque Eq. (5), we have

$$m\ddot{x}\cos(\emptyset) + ml\ddot{\emptyset} = mgsin\emptyset \tag{8}$$

Equation (4) and (8) are defining the nonlinear behavior of the system for cart-inverted pendulum system.

By solving the above system in s-domain, following system transfer function will be achieved

$$T_{\mathrm{cart}} = \frac{X}{U} = \frac{0.132\left(s^2 + 0.0378s - 12.98\right)}{0.339\left(s^3 + 0.058s^2 - 13.27s - 0.253\right)} \tag{9}$$

$$T_{\mathrm{pendulum}} = \frac{\emptyset}{U} = \frac{0.874s^2}{0.339\left(s^3 + 0.058s^2 - 13.27s - 0.253\right)} \tag{10}$$

3 Simulations and Results

The Matlab-Simulink model used for dynamics analysis for the cart-inverted pendulum, are developed and shown in Fig. 2. The inverted pendulum has two states: stable pendant position and unstable upright position. The process of stabilizing pendulum at upright position will be done in two-phase. Initially, pendulum needs to swing-up from stable to unstable position and then stabilize there with least possible

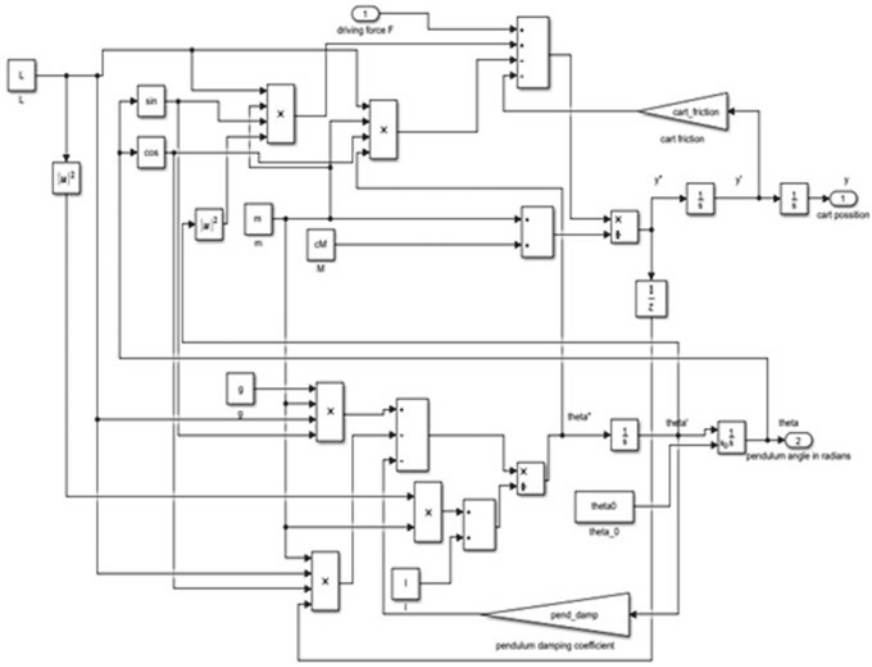


Fig. 2 Simulink model for cart-inverted pendulum system

oscillation. As requirement for both phases will be different, we need well-designed switching strategy in place [18, 19].

Two classical PID controllers are used for swing-up and stabilization phase. Swing-up controller will have real time pendulum angle and cart displacement as an input, and control voltage and switching signal as an output. Stabilization controller will compare cart position and pendulum angle with their respective reference values, in order to generate control signal [20].

The PID controller equations are given:

$$y_{p=}k_{\mathrm{pp}}e_{\emptyset}(t) + k_{\mathrm{ip}}\int e_{\emptyset}(t) + k_{\mathrm{dp}}\frac{\mathrm{d}}{\mathrm{d}t}e_{\emptyset}(t) \tag{11}$$

$$y_{\emptyset} = k_{p\emptyset}e_{\emptyset}(t) + k_{i\emptyset}\int e_{\emptyset}(t) + k_{d\emptyset}\frac{d}{dt}e_{\emptyset}(t) \tag{12}$$

where $e_{\emptyset}(t)$ is angle error, y_p and $y_{\emptyset}$ are position and angle output, k_{pp} and $k_{p\emptyset}$ are proportional coefficients for the position and angle, k_{ip} and $k_{i\emptyset}$ are integral coefficients for the position and angle, k_{dp} and $k_{d\emptyset}$ are derivative coefficients for the position and angle.

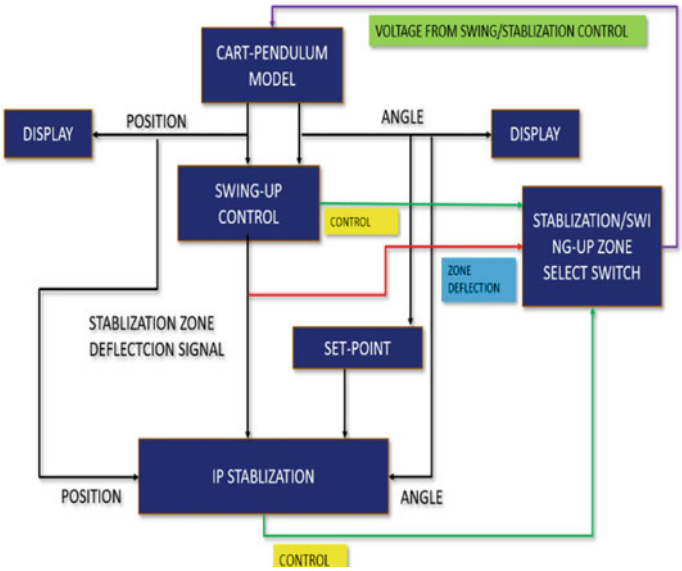


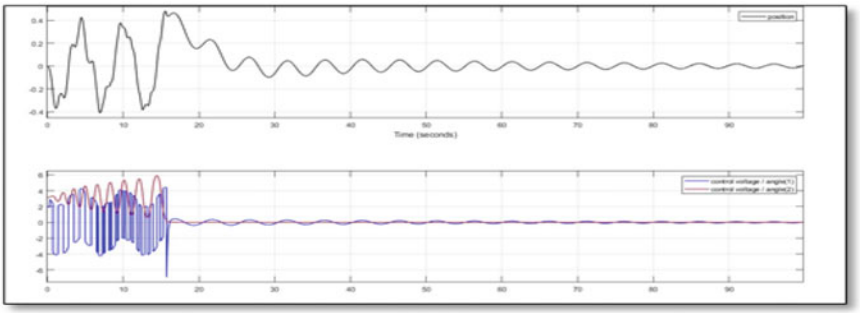
Fig. 3 Process flowchart for cart-inverted pendulum system

The controller tuning becomes more tedious, because dynamics of pendulum angle and cart displacement are inter-dependent. Any alteration in one controller parameter will modify, position of cart, and pendulum angle both [21].

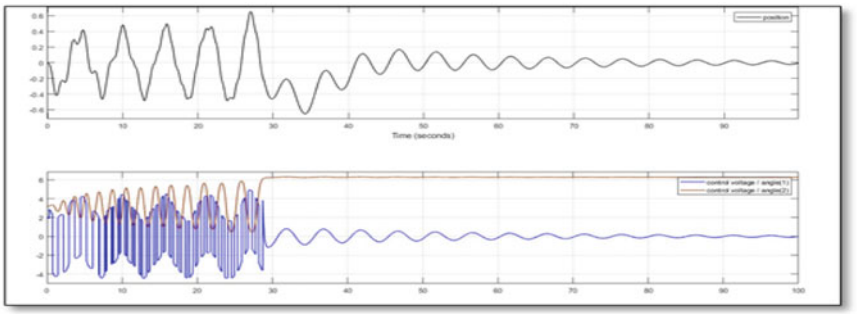
All system and controller design parameters are kept constant for complete simulation work. Figure 3 shows the process flow for implementation of controlling and stabilization.

3.1 Variable Pendulum Mass (M)

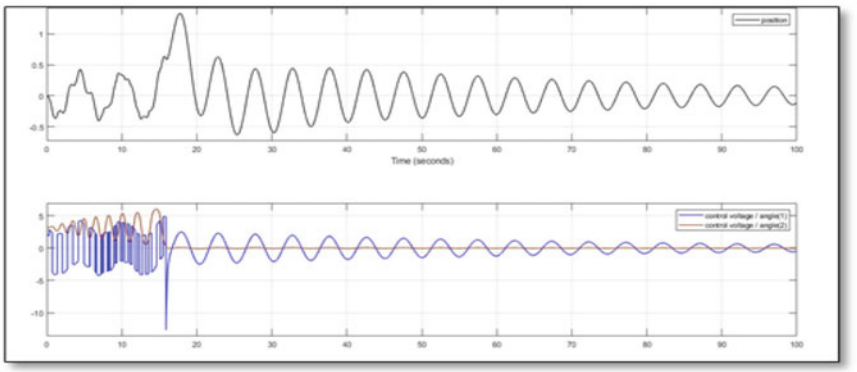
The pendulum mass (m) plays a key role in inverted pendulum stabilization, as we can see from Eq. (8). System is considered here without disturbance for simulation work. MATLAB simulation is carried out for the different value of pendulum mass (m), and results are shown in Fig. 4. Initially, oscillations are increased in both cart motion and pendulum angle with the decrement of pendulum mass, as shown in Fig. 4a, b, but finally reaching equilibrium state with less time and error as compared to default value response. On the other hand, when pendulum mass is increased from 0.2 kg to 0.225 kg, oscillations significantly increases, peak overshoot increases and degradation of stability occurs, as shown in Fig. 4c, d, e, and f. Figure 4g shown unstable system response with 0.25 kg pendulum mass.



(a) m=0.2kg

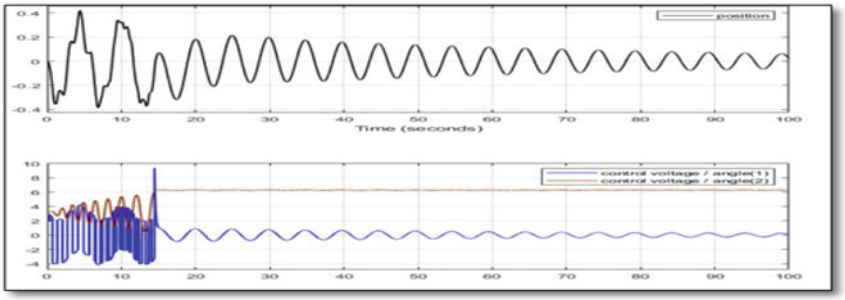


(b) m=0.15kg

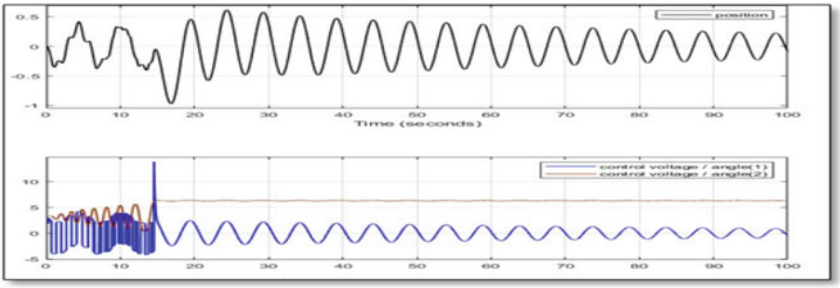


(c) m=0.205kg

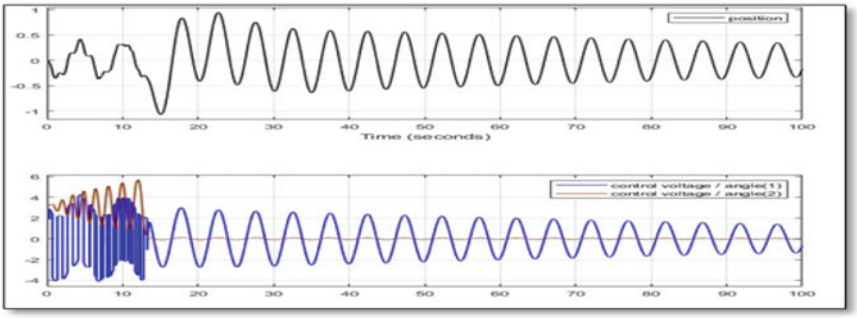
Fig. 4 Simulation results with variable pendulum mass values: **a** $m = 0.2$ kg, **b** $m = 0.15$ kg, **c** $m = 0.205$ kg, **d** $m = 0.\ 215$ kg, **e** $m = 0.22$ kg, **f** $m = 0.225$ kg, **g** $m = 0.25$ kg



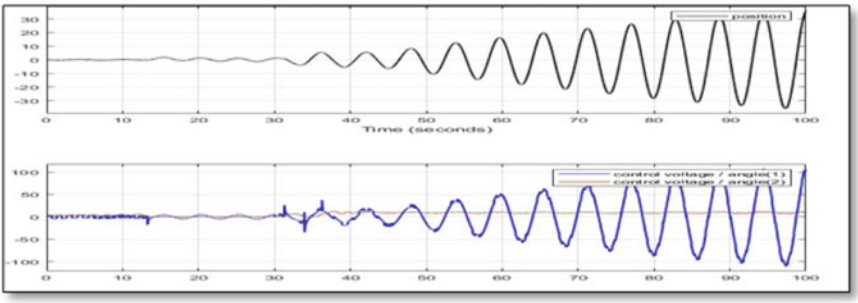
(d) m=0.215kg



(e) m=0.22kg

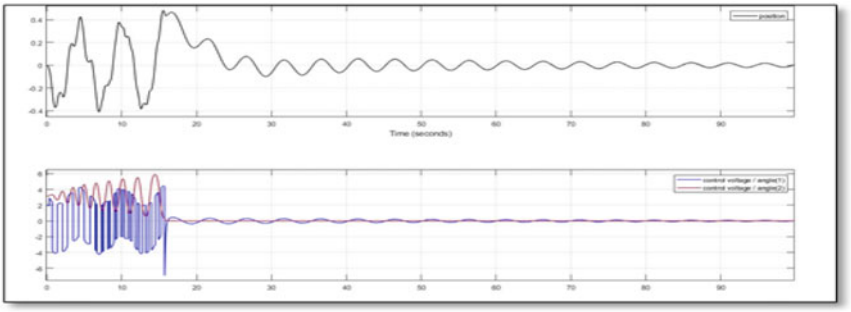


(f) m=0.225kg

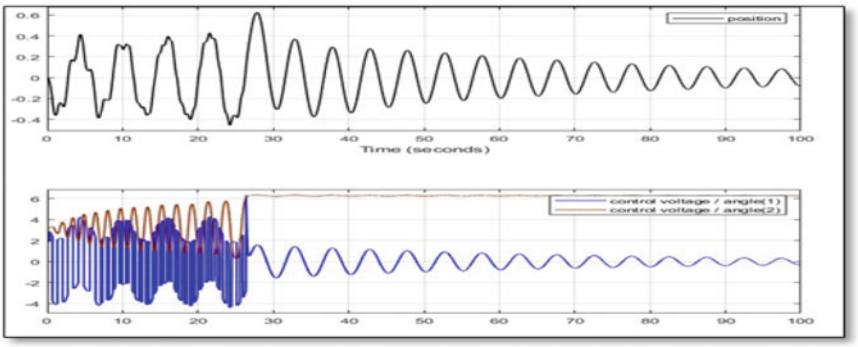


(g) m=0.25kg

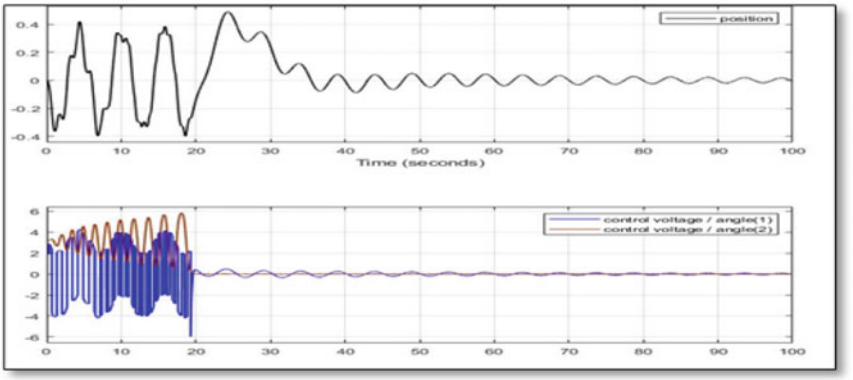
Fig. 4 (continued)



(a) L=0.3m

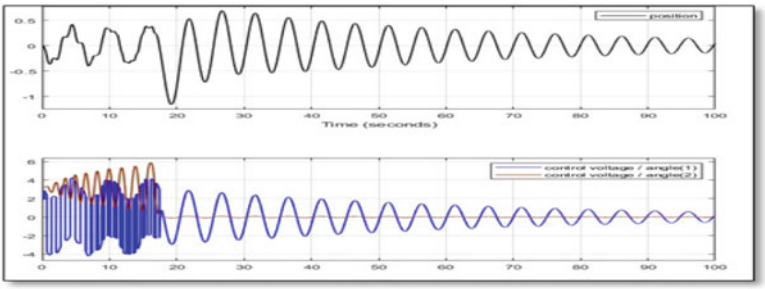


(b) L=0.25m

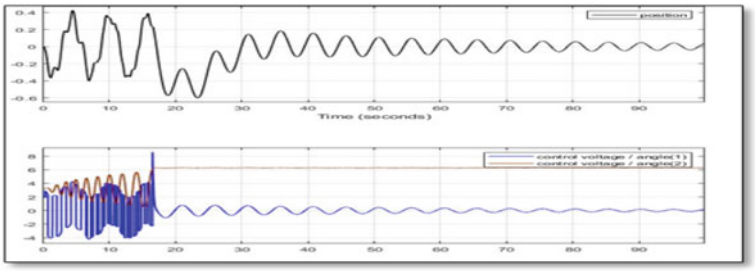


(c) L=0.27m

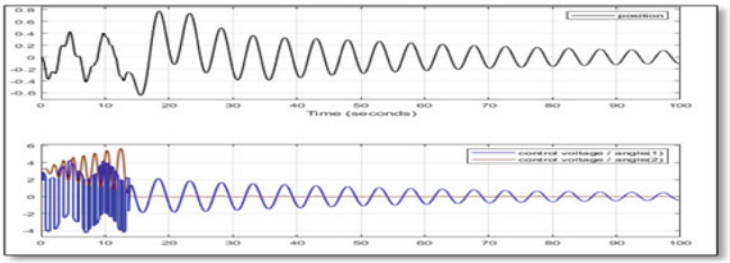
Fig. 5 Simulation results with variable pendulum length values



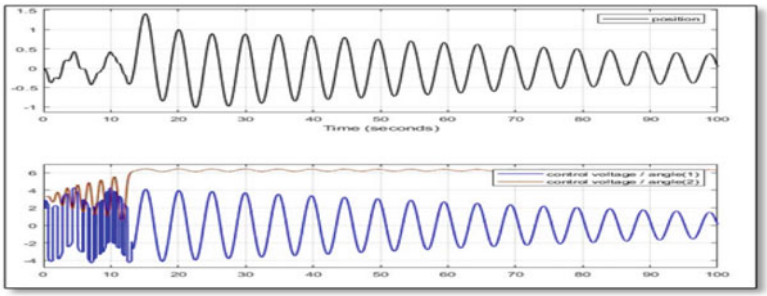
(d) L=0.28m



(e) L=0.29m



(f) L=0.32m



(g) L=0.34m

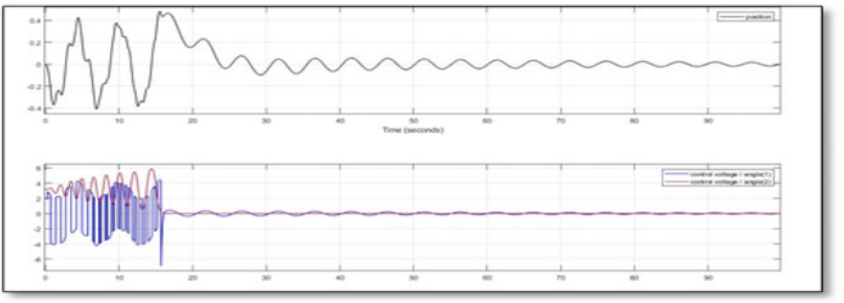
Fig. 5 (continued)

3.2 Variable Pendulum Length (l)

Torque force and equivalent displacement of pendulum mass are depending upon the pendulum length (l), as shown in Eq. (2) and (8). When the system, in Fig. 2, is simulated for different value of pendulum length (l), the responses are shown in Fig. 5. Oscillation's magnitude and period increased very prominently in both the cases of increase and decrease but showing more dominance in case of increasing the length.

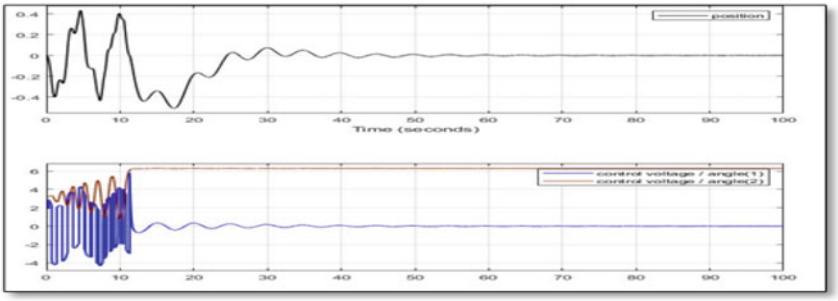
3.3 Variable Cart Mass (M)

Varying cart mass can be done in two ways: as system parameter or as disturbance parameter (giving addition weight during the operation) [22]. This changed cart mass, will also affect the requirement of external input force to swing and hold the pendulum in upright position, as shown in Eq. (1). From the simulation results given in Fig. 6, it can be observed that, even a slight variation in cart mass (M) will cause serious damage to the stability of the pendulum [23]. As a result of cart mass decrement, the amplitude of oscillations increases and then after continues to move toward a stable position. Increase in cart mass value results in unstable system, as shown in Fig. 6e and f.

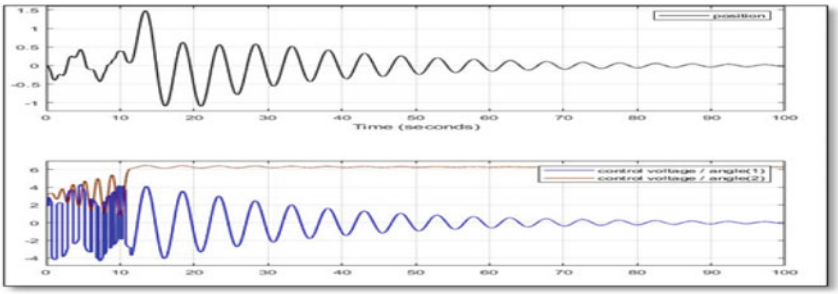


(a) M=2.3Kg

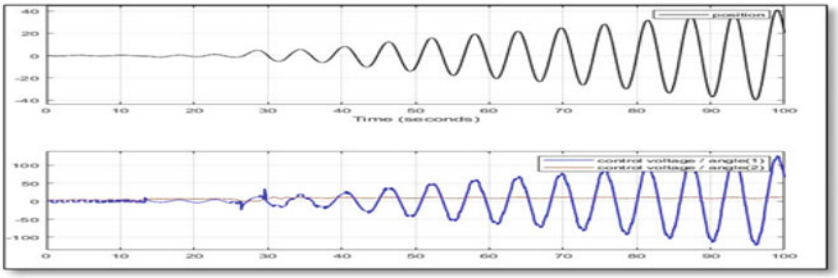
Fig. 6 Simulation results with variable cart- mass values: **a** $M = 2.3$ kg, **b** $M = 2.22$ kg, **c** $M = 2.25$ kg, **d** $M = 2.4$ kg, **e** $M = 2.45$ kg



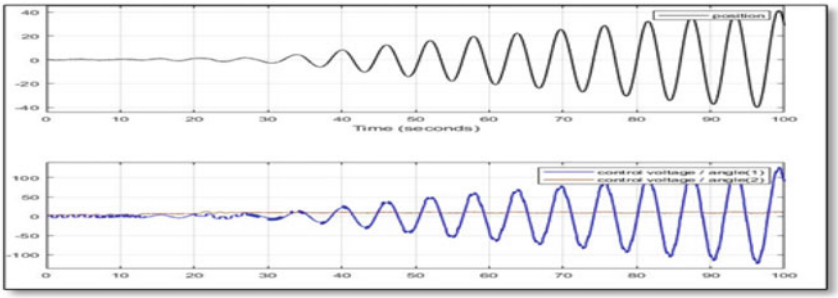
(b) M=2.22Kg



(c) M=2.25Kg



(d) M=2.4Kg



(e) M=2.45Kg

Fig. 6 (continued)

4 Conclusion

A simulation is carried out for different values of three dominating design parameters, pendulum mass, pendulum length, and cart mass to show their effects on stabilization and swing-up. By simulating the cart-inverted pendulum system for different value of pendulum mass, it has been observed that the reduction of mass will increase the oscillation. Still, the response continues to move towards a stable position, but a significant increase in the mass value will bring system towards instability. Similarly, system response for different values of pendulum length is also obtained, and it shows prominent effects of both, increasing and decreasing value on stabilization. A similar process is carried out for a set of cart mass values, which can also be considered as real-time disturbance by introducing additional weights on cart during the process. And, this variation results in unstable condition for the system, more prominently on increasing values. An analysis can be carried out for other parameters like friction on the rail track, the initial angle of pendulum, rail-length restriction, etc.

References

1. Davison EJ (1990) Benchmark problems for the control system design: report of the IFAC Theory Committee. Schlossplatz 12, A-2361 Laxenburg, Austria
2. El-Hawwary MI, Elshafei A-L, Emara HM, Abdel Fattah HA (2006) Adaptive fuzzy control of the inverted pendulum problem. IEEE Trans Control Syst Technol 14(6):1135–1144
3. Mahindrakar AD, Rao S, Banavar RN (2006) Point-to-point control of a 2R planar horizontal underactuated manipulator. Mech Mach Theory 41(7):838–844
4. Wai R-J, Kuo M-A, Lee J-D (2007) Design of cascade direct adaptive fuzzy control for two-axis inverted-pendulum servomechanism. In: IECON 2007–33rd Annual conference of the IEEE industrial electronics society. IEEE, pp 898–903
5. Wang L-X (1996) Stable adaptive fuzzy controllers with application to inverted pendulum tracking. IEEE Trans Syst Man Cybern Part B (Cybern) 26(5):677–691
6. Eltohamy KG, Kuo C-Y (1997) Real time stabilisation of a triple link inverted pendulum using single control input. IEE Proc Control Theory Appl 144(5):498–504
7. Kajiwara H, Apkarian P, Gahinet P (1999) LPV techniques for control of an inverted pendulum. IEEE Control Syst Mag 19(1):44–54
8. Valluru SK, Singh M, Singh M (2018) Application of linear quadratic methods to stabilize cart inverted pendulum systems. In: 2018 2nd IEEE International conference on power electronics, intelligent control and energy systems (ICPEICES). IEEE, pp 1027–1031
9. Valluru SK, Singh M, Mayank S, Khattar V (2018) Experimental validation of PID and LQR control techniques for stabilization of cart inverted pendulum system. In: 3rd IEEE International conference on recent trends in electronics, information and communication technology, pp 708–712
10. Valluru SK, Singh M (2018) Performance investigations of APSO tuned linear and nonlinear PID controllers for a nonlinear dynamical system. J Electr Syst Inf Technol 5(3):442–452
11. How JP (2015) Benchmarks. IEEE Control Syst Mag 6–7
12. Wai R-J, Chang L-J (2006) Adaptive stabilizing and tracking control for a nonlinear inverted-pendulum system via sliding-mode technique. IEEE Trans Industr Electron 53(2):674–692
13. Pangaribuan T, Nasruddin MN, Marlianto E, Sigiro M (2017) The influences of load mass changing on inverted pendulum stability based on simulation study. IOP Conf Ser Mater Sci Eng 237(1):012005

14. Vasudevan H, Dollar AM, Morrell JB (2015) Design for control of wheeled inverted pendulum platforms. J Mech Robot 7(4)
15. Muskinja N, Tovornik B (2006) Swinging up and stabilization of a real inverted pendulum. IEEE Trans Industr Electron 53(2):631–639
16. Prasad LB, Tyagi B, Om Gupta H (2012) Modelling and simulation for optimal control of nonlinear inverted pendulum dynamical system using PID controller and LQR. In: 2012 Sixth Asia modelling symposium. IEEE, pp 138–143
17. Ray G, Das SK, Tyagi B (2007) Stabilization of inverted pendulum via fuzzy control. J Inst Eng India. Part El Electrical Engineering Division 88
18. Chatterjee DD, Patra A, Joglekar HK (2002) Swing-up and stabilization of a cart–pendulum system under restricted cart track length. Syst Control Lett 47(4):355–364
19. Aracil F, Gordillo J (2004) The inverted pendulum: a benchmark in nonlinear control. IEEE World Autom Congr 16:468–482
20. Valluru SK, Singh M (2017) Stabilization of nonlinear inverted pendulum system using MOGA and APSO tuned nonlinear PID controller. Taylor Francis Cogent Eng 4(1):1–15
21. Singhal NK, Swarup A (2018) Application and comparison of few control methods to inverted pendulum dynamics. In: 2018 4th International conference on computing communication and automation (ICCCA). IEEE, pp 1–6
22. Sahnehsaraei MA, Mahmoodabadi MJ (2021) Approximate feedback linearization based optimal robust control for an inverted pendulum system with time-varying uncertainties. Int J Dyn Control 9(1):160–172
23. Mondal R, Dey J (2020) Performance analysis and implementation of fractional order 2-DOF control on cart-inverted pendulum system. IEEE Trans Ind Appl 56(6):7055–7066

Efficient Design of QCA ADD-SUB Subsystem for Future Microprocessors

Ashish Mohan, Shilpee Patil, and V. K. Pandeyr

Abstract In the fast growing world the life of people is revolving around electronic devices where the battery life and speed of edge devices needs to be continuously extended. Quantum-dot Cellular Automata (QCA) is one of the most attractive technologies for extremely low power computing at nanoscale. It is a transistor less computation paradigm that also addresses the problems of device density and interconnection. Being a prominent subsystem within as microprocessors and application specific integrated circuits, an efficient QCA-based design of Adder Subtractor is presented in this paper for future microprocessors. Simulations are performed on QCA designer tool and the results are also compared with previous works which also proves the efficacy of presented work.

Keywords QCA · Adder design · Low power

1 Introduction

The idea to work particularly in the field can be started releasing advanced world where our lives are connected to the smart gadgets in our hand and all the electronic devices that surrounds us. On beginning with our research in this field we came across a law which is kind of responsible for the whole upgradation process, i.e., the Moore's Law. So, from here our journey to update the microprocessors began. In recent years, the utilization of CMOS technology is restricted by high consumption, low speed, and density beyond 10 nm and now the problems arising due to Denard Scaling [1]. To overcome these problems, a variety of researchers are ascertained to seek out the answer for this classical CMOS technology. Quantum and Reversible logic circuits with the upcoming nanotechnologies like DNA computing, Nuclear

A. Mohan (✉) · S. Patil · V. K. Pandeyr
Department of Electronics and Communication Engineering, NIET, Greater Noida, India
e-mail: erashishmohan@gmail.com

S. Patil
e-mail: shilpee.patil@niet.co.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_18

Magnetic Resonance (NMR), Quantum Cellular Automata (QCA) are trending for high-speed application [2]. Several design methodologies has been proposed for reversible circuits and testing of these circuits is also done rigorously in the past decade [3, 4].

Data Path Elements (DPE) are the processing circuits within the microprocessor that has the responsibility to compute all the mathematical and logical operations. Adder circuit is one of the significant subsystem which has to be utilized in nearly all other mathematical calculations, logical operations and ALU within the microprocessor [5]. In this regard, an efficient QCA-based design of adder-subtractor circuit (ADD-SUB) is presented in this paper. The circuit can be adopted in designing for improving high speed and size of microprocessors.

2 Logical Premitives of QCA

In this section, the background of QCA technology and how this new concept will take the great advantage of a physical effect called the Coulomb force is described. The information propagation through QCA cell by clicking is also elaborated including the essential elements and gates used for QCA circuits. To understand the working of a basic four dot QCA cell, the motion of an electron in infinite potential well has to be studied. The walls of the potential well prevents the electron to tunnel between adjacent dots. Electrons in an infinite potential well exist as a wave function that gives us the probability of finding an electron within that potential well. This probability is proportional to the solution to the Schrodinger wave equation for a free electron [6]. The nonlinear behavior of a cell is used to polarize its neighbors cells as the role of voltage gain in traditional logic circuits. The logic is called as restoring logic which is proven experimentally. It assists in restoring input voltage levels float on the wire and gives a level for the immunity to changes in cell position and placement.

By positioning QCA cells, an inverter (or a NOT gate) as shown in Fig. 1 and a majority gate shown in Fig. 2 can be implemented. On applying input '0', we get output of '1' which amounts to flipping of the input which depicts the realization of the circuit of an inverter. Similarly in the majority gate, if '001' is applied the logic '0' will be turned at the output. For logic input '011', the output will be logic value '1' and same as for other logic 2^n inputs. It proves the working of a majority function, i.e., if A, B and C are inputs, then the output can be given by $M(A, B, C) = AB + BC + AC$.

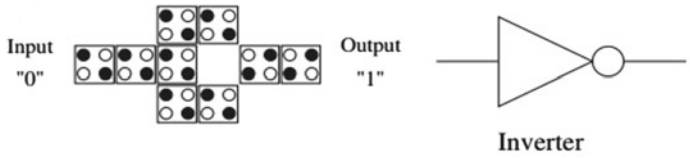


Fig. 1 QCA inverter

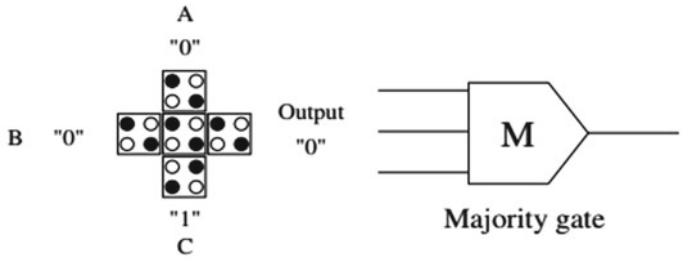


Fig. 2 QCA majority gate

By changing any one input to a constant logic value, the majority gate can be used to implement basic Boolean function, like AND and OR. If one of the input (C) is fixed to 0, the majority gate work as an AND gate, i.e., $M(A, B, 0) = AB$. Similarly, by fixing one of the inputs to logic 1, the majority gate will work as an OR gate ($M(A, B, 1) = A + B$). It can be analyzed that a majority gate is advantageous in producing AND and OR function, since three inputs enable implementation of several other arithmetic functions like carry in a full-adder with one majority gate only. Moreover, a majority gate requires same number of QCA cells as in the case of implementing an AND gate and OR gate. These QCA gates can also be compared CMOS technology, much larger number of transistors are required to implement a three input majority gate.

Note that, the majority gate in combination with a NOT gate is designated as a *logical primitives' gates* in QCA designing which can be utilized to implement universal Boolean function NAND and NOR. Furthermore, a rich properties for majority gates can be observed that provides low-complexity QCA design of several arithmetic operations that influence us to use this gate in our ADD/SUB design.

3 Performance Measures

It has been revised from the literature, the proposed design and methodologies using QCA cells are justified by means of certain performance parameters. On the basis of these parameters; the quality and complexity can be evaluated [7]. These performance measures are written as follows:

Clocking: In QCA since there is no flow of current, hence to control the flow of information and to control the timing of QCA circuit basically QCA cell switching and relaxation, a clock signal is necessary to control the states of QCA cells from input to output. Not only does clocking control the information flow, but it also supplies the real power in QCA [8].

Wire Crossing: In QCA, there are following types of wire crossing: Direct Single Layer Wire Crossing where two types of QCA Cell (90° and 45°) are required. The

first wire has only direct cells (non-rotating) and the second wire has only rotating cells. Multiplayer Wire Crossing, where two separate layers are used for two wires of the same cell type.

Power Dissipation: The electronic circuits and devices need a certain amount of power to operate which should be kept as low as possible. The power dissipation of a gate is the power required by the gate for its operation and is expressed in milliwatts (mW). It represents the power supplied to the gate from power supply; rather than the power delivered from other gates. It is proportional to the heat produced by the system. Excessive heat dissipation can enhance operating temperature which is also one of the reasons of power dissipation. The designer must have these perspective during its implementation cycle.

Logical Completeness: A 3 input majority gate is considered a programmable AND and OR function gate. When they combined with the NOT operation, these form a logically complete function. Both of these gates have been validated on metal island dot cells.

Number of Cells, Area and Latency: In QCA circuits, the delay increases because of the increased cell counts and wire connections which further the complexity increases [9]. The wiring channels for the input and output synchronization must be minimized in addition, since channels include to the circuit area. For an efficient design in QCA, it is necessary to minimize the size with new optimized layouts. The delay is also correlated to the number of clock zones utilized to function a circuit that results in terms of latency of a QCA circuit.

4 Proposed ADD-SUB Circuit

A Half adder has been taken as the logical primitives of the circuit as shown in Fig. 3. It has been designed over QCA-based architecture. A half adder performs the addition of two boolean variables and gives two output, one is the sum and the other is the carry. Let A and B be two input variables. Then the sum and carry is respectively given by $S = A \oplus B$ and $C = AB$, respectively. The logic input outputs is also depicted in Table 1.

The proposed design using majority and inverter gates can be seen in Fig. 4. The design is better in the performance measures than the single-layer design and utilizes only $0.11\,\mu m^2$ area with 90 cells, which is highly efficient. The proposed design does addition and subtraction both simultaneously and offers the user a choice to pick between these actions. There are 2 inputs A and B, the XOR gate in the middle provides the SUM or DIFFERNCE output whereas the CARRY output is reserved for the adder and similarly the BORROW is reserved for the subtractor.

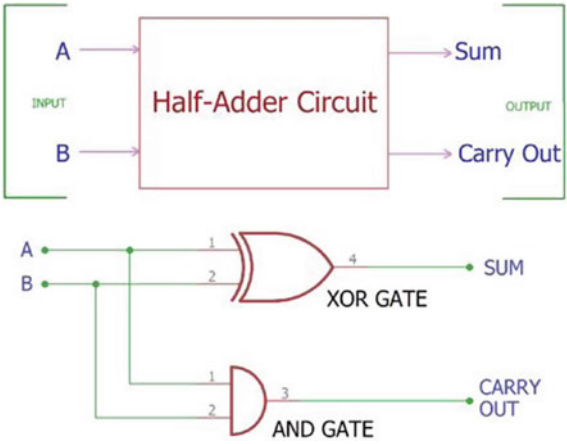


Fig. 3 Logic primitives of a half adder

Table 1 Input output data flow

Input			Output		
A	B	Permutation	C	S	Permutation
0	0	0	0	0	0
0	1	1	0	1	1
1	0	2	0	1	1
1	1	3	1	0	2

5 Simulation Results and Comparisons

We have performed the following calculations with the help of QCADesigner tool whose layout can be seen in Fig. 4. Multiple cells connected together form a wire which transports the electrical signal from point A to point B. Each cell contains 4 quantum dots, these dots along with the polarization define the purpose of a cell. The CARRY output is a basic 'AND' gate which computes A AND B and the BORROW output is calculated by A' AND B. The SUM/DIFF output is calculated with the help of a basic 'XOR' gate, i.e., A'B + AB'. At the SUM/DIFF output two wires carrying the signal A'B and AB' meet where the Sum function is carried out to give the XOR output. Three clock zones (Green, Violet, Cyan) are used in our proposed design, with each zone there is an output delay of 0.25 and therefore a total output delay of 0.75 can be observed from the simulation waveform in Fig. 5.

An efficient design based on adders/subtractor is presented. Comparisons with three prior designs [10–12] are also given in Table 2. Latency can be calculated with a simple formula Latency = Area × $(\text{Latency})^2$.

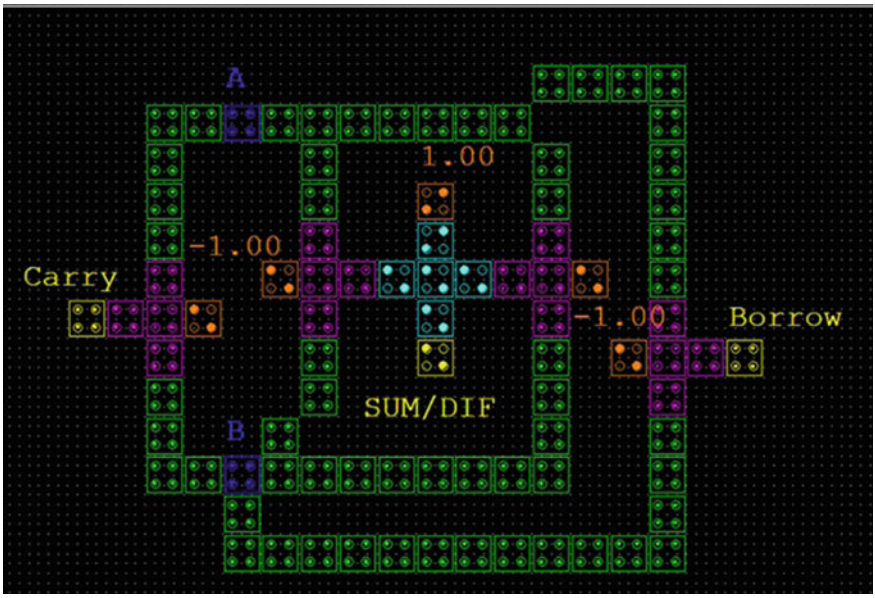


Fig. 4 Proposed QCA ADD-SUB circuit

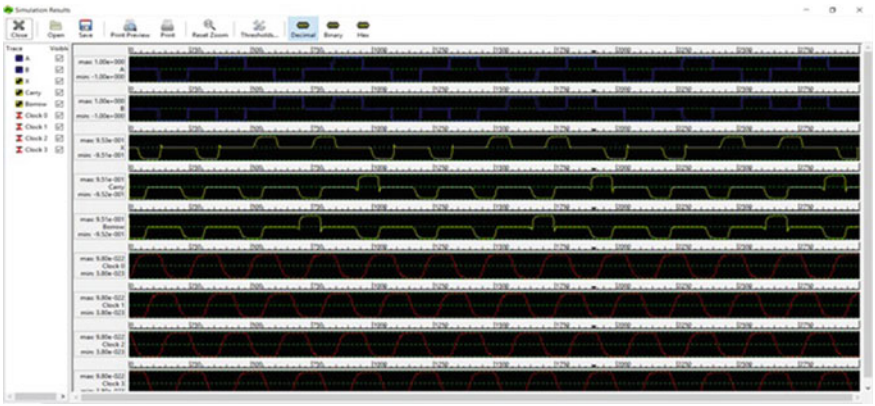


Fig. 5 Input output simulation waveforms

Table 2 Simulation results

Design	Cells	Area (μm^2)	Latency	Cost ($)
Design 1 [10]	–	0.688	4	11.008
Design 2 [11]	–	0.43	3	3.87
Design 3 [12]	126	0.23	3	2.07
Proposed Circuit	90	0.11	3	0.99

6 Conclusion

Computer algorithms for basic QCA data path elements have been described at the initial state of this paper. We have considered for logic, clocking, hierarchical design, low power dissipation, and memory for the designing. An efficient adder and subtractor ADD-SUB circuit is proposed that has proved better than most of its previous works in the area. It has been obtained using simulations over the most popular QCA designer tool that the proposed design utilizes fewer cells to realize the given logic and avoids the concept of inclusion of additional propagation delays and complexity.

References

1. Mack C (2015) The multiple lives of Moore's law. IEEE Spectrum 52(4):31
2. Gaur HM, Sasamal TN, Singh AK, Mohan A, Pradhan DK (2020) Reversible logic: an introduction. In: Design and testing of reversible logic. Springer, Singapore, pp 3–18
3. Gaur HM, Singh AK (2016) Design of reversible circuits with high testability. Electron Lett 52(13):1102–1104
4. Gaur HM, Singh AK, Ghanekar U (2015) A review on online testability for reversible logic. Procedia Comput Sci 1(70):384–391
5. Gaur HM, Singh AK, Ghanekar U (2019) Design of reversible arithmetic logic unit with built-in testability. IEEE Des. Test. 36(5):54–61
6. Sasamal TN, Gaur HM, Singh AK, Mohan A (2020) Novel approaches for designing reversible counters. In: Design and testing of reversible logic. Springer, Singapore, pp 37–48
7. Sasamal TN, Singh AK, Mohan A (2016) An optimal design of full adder based on 5-input majority gate in coplanar quantum-dot cellular automata. Optik 127(20):8576–8591
8. Walus K, Jullien GA (2006) Design tools for an emerging SoC technology: quantum-dot cellular automata. Proc. IEEE 94(6):1225–1244
9. Sasamal TN, Singh AK, Mohan A (2018) Design of cost-efficient QCA reversible circuits via clock-zone-based crossover. Int. J. Theoret. Phys. 57(10):3127–3140
10. Kummamuru RK, Timler J, Toth G, Lent CS, Ramasubramaniam R, Orlov AO, Bernstein GH, Snider GL (2002) Power gain in a quantum-dot cellular automata latch. Appl. Phys. Lett. 81(7):1332–1334
11. Pudi V, Sridharan K (2013) Efficient design of Baugh-Wooley multiplier in quantum-dot cellular automata. In: 2013 13th IEEE International conference on nanotechnology (IEEE-NANO 2013). IEEE, pp 702–706
12. Jaiswal R, Sasamal TN (2017) Efficient design of exclusive-OR gate using 5-input majority gate in QCA. IOP Conf Ser Mater Sci Eng 225(1):012143

Predication of Diabetes Through Machine Learning Technique

Arun Kumar Singh, Ashish Tripathi, Pushpa Choudhary, Prem Chand Vashist, and Arjun Singh

Abstract There are couples of man-made intelligence frameworks that are used to perform farsighted examination over tremendous data in various fields. Farsighted assessment in human administrations is a troublesome task regardless can help experts with choosing colossal data instructed advantageous decisions about patient's prosperity and treatment. The research paper discuss the perceptive assessment in therapeutic administrations; six unmistakable artificial intelligence counts are used in this investigation work. For break down reason, a dataset of a patient's restorative verification is procured and six assorted man-made intelligence estimations are applied on the dataset. Execution and precision of the applied computations are discussed and took a gander at. Assessment of the different man-made intelligence techniques used in this examination reveals which count is most suitable for a desire for diabetes. The paper intends to support masters and experts in early desire for diabetes using man-made intelligence frameworks.

Keywords Dataset analytics · Predictive analytics · Machine learning · Diabetes · Artificial intelligence · Support vector machine · Random Forest

1 Introduction

Inadequate insulin production in humans results in diabetes. Insulin is created by the β cell of human hormones, and which allows glucose to be used as fuel by muscles, body fat cells, and the liver.

In the case of diabetes, a complex group of disorders arises, with insulin being deficient or absent, with the final result being hyperglycemia. Extensive insulin therapy to maintain near-normal blood glucose levels minimizes chronic problems [1], but may raise the risk of symptom hypoglycemia and possibly fatal extreme hypoglycemia.

The primary obstacle to concentrated diabetes management is hypoglycemia [2], a long-life issue for the level of adaptation for diabetic patients who lack the risk of

A. K. Singh (✉) · A. Tripathi · P. Choudhary · P. C. Vashist · A. Singh
Department of Information Technology, G.L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: singh_arun7@yahoo.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_19

hypoglycemia to maintain glycemic control. Thus, hypoglycemia has been identified as the primary inhibition of severe diabetes supervision [2].

In the case of large blood sugar (LBG) fluctuations over time of struggle for control of tight glycemic [3]. Such a process usually occurs mainly for external reasons, including the amount and timing of insulin, physical activity, eating time, etc., so we can say that in the case of diabetes, the fluctuations of BG The role of gives quantitative results.

A composite dynamic system, which is subject to other influences such as external as well as internal [4, 5].

One metabolic disorder of diabetes is diabetes mellitus, which increases blood sugar levels over a long period. The root cause of diabetes is a decrease in blood sugar, this means that sugar production is not sufficient, in the literary survey; it has been observed that due to the recurring detection of diabetic patients there is an increase in many cases of mellitus.

Therefore, a fundamental solution is to detect symptoms early through early diagnosis. Critical parameters are always necessary and recommended condition for detection of diabetes, over the last few decades, a lot of technology has been reported to, which today helps in the early diagnosis of diabetes, created model doctors and hospitals clinical helps in decisions and planned treatment process for patients [6].

1.1 Machine Learning

Machine learning, a branch of Artificial Intelligence that interacts with algorithms, mathematical analysis, and statistics, was introduced by Arthur Samuel in the 1950s. Machines were primarily programmed to obey specific instructions and did not have the opportunity to make these decisions.

In the digital age, information is something that is available in abundance. The main constituent predictive modeling or maintaining a piece of memories from such a previous encounter cannot be best solved using traditional analysis techniques (Fig. 1).

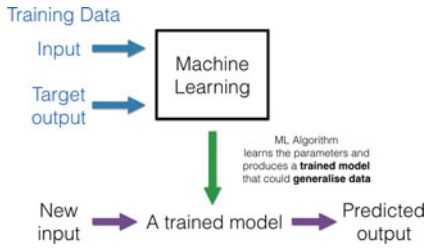


Fig. 1 The learning process of supervised ML

1.2 Types of Machine Learning

Machine learning has its origins in analytics, which is already known as the art of data collected from various sources. Methods like regression analysis and Bayesian statistics, both of which are over two centuries old, are still at the core of machine learning nowadays. There are a number of ways to express this concept, however, the three most common are supervised learning, un-supervised learning, and re-enforcement learning. The field of machine learning (ML) is very often subdivided into subareas based on the types of challenges being addressed. The following is a rough classification system, as seen in the diagram (Fig. 2).

2 Types of Diabetes

Type-1 Diabetes: A genuine, unending ailment happens oftentimes occurs in youths and adults. Here, pancreas thoroughly stops the production of insulin. The individual struck by Type 1 is thoroughly subject to insulin from external medications to control the sugar levels in the body. The (Diabetes Control and complexities trail (DCCT) 2 helped the person through the once-over arrangements with being removed after to keep from the side effects, extraordinary troubles on various organs and live longer better life through the guidelines and sustenance penchants. A dietary philosophy was found through these guidelines.

Type-2 Diabetics: It is a class of endless; non-insulin subordinate pain routinely happens in grown-ups. There are several substances of the events of sort 2 that are acquired and metabolic parts, family heritage, physical dormancy overweight, largeness, heartbreaking eating standard, and smoking affinities develop the danger of diabetes.

Predication of Diabetics: It is a phase before sort 2 diabetes, where glucose measurement of the individual has been higher than typical yet not to the elements

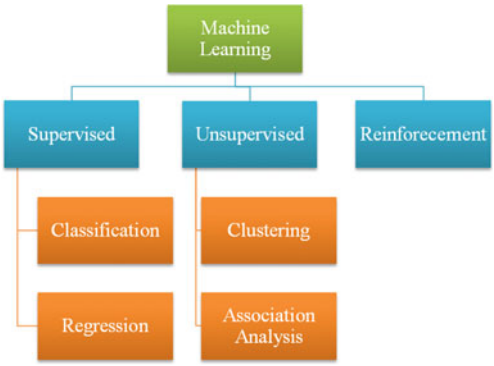


Fig. 2 Classification of machine learning

of sort 2. A man with a pre-diabetes condition has more odds of getting type 2 under express conditions and measures.

Gestational of Diabetics: It is a fundamental grouping affected for women in the midst of pregnancy an assortment of hormones in the midst of pregnancy and extended insulin substance can incite the high blood glucose level. The recently imagined children have the chance of making diabetes.

3 Literary Survey

Diabetes a non-transmittable illness is prompting long haul difficulties and genuine medical issues. A report from the World Health Organization tends to diabetes and its entanglements that sway on individuals physically, monetarily, and financially over the families.

The review says about 1.2 million passings because of the uncontrolled phase of wellbeing lead to death. About 2.2 million passings happened because of the hazard variables of diabetes like cardiovascular and different illnesses [7].

A significant assortment of writing has been distributed on anticipating frequency of type II diabetes inside five to ten years depending on malady history and clinical estimations. Most of these examinations utilized logistic relapse, which takes into account simpler elucidation of impact sizes yet may neglect to catch progressively complex nonlinear connections and communications among highlights.

As a lifestyle all over the world, diabetes is a type of persistent disease which an indirect effect on human health in a critical way. There are approximate 382 million people have diabetic across the world, as per the report of the International Diabetes Federation. But as the scenario of the cases, it may be approximate 592 million double by 2035 [8]. Basically simple diabetics may occur due to increases at the initial stage of blood glucose.

As a group of metabolic disorders known as diabetes mellitus (DM), which mostly arise due to irregular insulin discharge or action, insulin deficiency causes high blood sugar (hyperglycemia), fat and protein and Impaired metabolic levels can increase, which is an important endocrine disorder affecting more than 200 million people in the world.

Different registering systems are applied in the Human Services area. The focal point of the writing review here is on the utilization of machine learning in the social insurance space [9]. So as to make a shrewd learning Medicinal services framework, there are uncertain logical information challenges.

Through machine learning the connection between information designs is comprehended and extra incentive from the immense human services information is revealed [10]. There are different research patterns and difficulties all through the information lifecycle while actualizing machine learning and is very much depicted in diabetes space difficulties are in improving exploration stages.

The truth mining is another methodology, for example, utilizing huge information to contemplate the patient's conduct via mobile phone and sensors that help to

improve social insurance quality. Separating valuable data from the electronic well-being record is reviewed. A keen configuration utilizing machine learning is proposed which is an online application that gives a proficient stage to the improvement of complex appraisal of wellbeing alongside checking method [11].

As of now in the medicinal services space by actualizing a machine learning calculation and utilizing factual therapeutic information intriguing examples with regards to illness information are found [12, 13]. For improved findings and guesses, AI can be utilized successfully is exhibited with two contextual analyses for malady determination.

The Support Vector Machine, Random Forest Algorithms, and Logistic Regression are used in the Support Vector Machine, Random Forest Algorithms, and Logistic Regression, which is the best selection of incredibly expressive network leveraging information-digging strategies for illness arrangement [14, 15].

The advantages of employing SVM are well-explained below.

- SVM searches the "best" margins that divides the categories, minimizing the probability of discrepancy, but logistic regression doesn't seem to; rather, it might have many decision boundaries having varying weight vectors that seem to be close to the ideal location.
- SVM performs effectively for unstructured and semi-structured text and pictures, whereas logistic regression operates with independent factors that have previously been established.
- SVM relies on geometrical characteristics of inputs, whereas logistic regression relies on statistical methodologies.
- SVM has a lower risk of overfitting, whereas logistic regression is more prone to overfitting.

Users may use logistic regression or support vector machine based on the quantity of training dataset (data)/features that have. Like an illustration, consider the following:

1. Utilize logistic regression or SVM with such a linear kernel if n is high (1–10,000) and m is modest (10–1000).
2. When n is small (1–10,000) and m is intermediate (10–10,000), apply SVM with a kernel (Gaussian, polynomial, etc.).
3. While *n* is small (1–10,000), and m is huge (50,000–1,000,000+), then: Initially, manually add additional features before using logistic regression or SVM with a linear kernel.

n denotes the number of characteristics, where *m* denotes the number of training instances.

In general, it is best to start with logistic regression to observe how the model performs; if that fails, you may attempt SVM without a kernel [16].

4 Methodology

Data mining is a significant and tremendous technology that is starting at now being used in the business for performing data evaluation and getting information into the data. Data mining uses a wide range of data mining techniques, for instance, electronic thinking, man-made thinking, and veritable appraisal. In this assessment, the man-made information structure is used for ailment measure [12] (Fig. 3).

PC based information gives a pool of devices and structures, using these instruments and procedures terrible data can be changed over into some monstrous, significant information by PCs. There are extensively two sorts of Machine Algorithm that are at present being used.

4.1 Parameters Uses in Dataset of Diabetes

In this paper we are using datasets from National Institute of Diabetes and kidney diseases; prediction of diabetic at early stage without diagnosis process is the main motive of this research paper, dataset contains dependent variables such as target value/output value and independent variables medical features/predictor, here variable is basically denoted by class, variable, or Label (Class). Independent variables or features include blood sugar, plasma glucose, HbA1c, age, and so on (Fig. 4).

Dataset Features:

- Age: This represents age of the patient
- Blood Sugar Fast: This is the blood sugar in fasting (before meal)
- Blood Sugar Meal: This is the blood sugar in 90 min after meal
- Plasma Random: Plasma glucose test at random time
- Plasma Fasting: Plasma glucose test fasting
- HBA1C: HbA1c refers to glycated hemoglobin
- Dataset Label
- Class: Represents whether patient has diabetes or not (0 or 1)

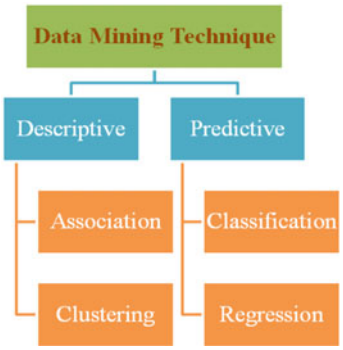


Fig. 3 Data mining techniques

Fig. 4 Parametric description of data set

Column_name	Type	Description
age	integer	This is age of patients
bs_fast	decimal	This is the Blood Sugar in fasting(Before meals).mmol/lL
bs_pp	decimal	This is Blood Sugar in 90 minutes after meals.mmol/L
plasma_r	decimal	Plasma glucose test Random taken at any time.mmol/L
plasma_f	decimal	Plasma glucose test fasting usually taken in the morning because it should be taken after atleast 8 hours.mmol/L
hba1c	integer	mmol/mol
type	string	diabetes type 1, diabetes type 2
class	boolean	0,1

4.2 Proposed Model

Initially, the diabetes dataset is taken which is produced initially from the Kidney Diseases and National Institute of Diabetes and Digestive. The dataset includes several features related to diabetes such as blood sugar, plasma glucose, HbA1c, Age, and so on. It also contains a label known as a class (0 or 1) indicating the presence or absence of diabetes. As there are many missing values in the dataset, they all need to be taken care of which is done in data pre-processing phase. Imputer function is used for replacing the missing value by using mean strategy. In data pre-processing scaling of the dataset is also done to improve the accuracy of classification algorithms applied. After preprocessing three classification algorithms, i.e., SVM, random forest, and logistic regression are applied one by one to train the model. Accuracy of all the algorithms was calculated so that we can compare them and choose the best one in this case to predict the occurrence or nonexistence of diabetes. After getting the best algorithm, an application is built using that algorithm to predict the occurrence or nonexistence of diabetes.

The proposed model for prediction of diabetics are review and explain in the below figure through the model diagram, the figure basically shown the steps of research accomplish in building the model (Fig. 5).

4.3 Brief Description of the Algorithms

Supervised machine learning is accomplished by classification, and support vector machine (SVM) is a form of supervised learning model that specializes in classification, in our proposal we have two types of training sample data set of SVM, which find out the hyperplane at initial stage for finest peak margin between the two classes. In the better view of results, hyperplane doesn't need to be closed to any data points that belong to other classes, The hyperplane must be selected in such a way that it stays away from the data points of each class while the points that are close to the margin of the classifier are support vectors [17].

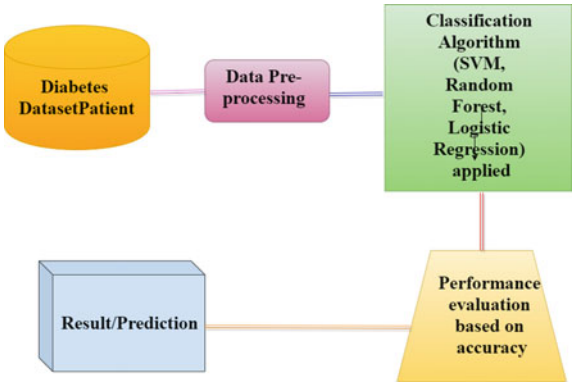


Fig. 5 Workflow model for prediction of diabetics using ML

Random Forest: Another algorithm random forest also comes under the supervised classification algorithm, here the name forest reflect lots of trees, so commonly the beauty of forest is more and more numbers of trees which make this algorithm to robust. With the huge integration of trees, the random forest algorithm gives an elevated accurate result. It can be used in regression and classification both problems and in case of classification, random forest from individual trees cumulative a class vote and after that predicts the result with the mass vote [7].

Logistic Regression: To solve classical problems, the necessary and acclaimed algorithms are logistic regression; here, the name logistic regression refers to fairly early techniques for linear regression. As the name suggests "logistic", this technique uses logistic techniques for classification. These techniques generally focus for prediction on the association between dependent and independent variables, where independent variable refers to actual output in another case dependent variable includes labels of variables, for actual value, these probabilities can be transformed into a binary value, basically, for all of these tasks are defined in mathematically by Sigmoid function. The Sigmoid function is in shape of the S curve, which mapped all those values between zero (0) and one (1), but in reality, it never gets the same value and these values can be classified through the threshold classifier. (http://www.csie.ntu.edu.tw/~cjlin/liblinear/).

4.4 Stepwise Process of Proposed Methodology

Step 1: For the prediction of diabetes disease, we will process the diabetic input dataset using spyder in the early stage process.
Step 2: Split the data set into two parts test and training datasets and this ratio can be 70:30.
Step 3: Now the appropriate ML algorithm has to be chosen between logistic regression algorithm, random forest algorithm, and support vector machine.

Step 4: Further training, the set may be used to develop a classifier model using machine learning techniques.
Step 5: Next, validate the classifier model using the test set for the aforementioned machine learning methods.
Step 6: Conduct a comparison analysis of the experimental performance results collected for all classifiers to determine which one produced the right result.
Step 7: After analyzing the data using multiple metrics, the highest performing algorithm is selected.

The proposal consists of a spyder and classifier model, which are based on the successful execution of each step, after that analysis of the experiment results for the final conclusion.

5 Results and Discussions

The proposal consists of a classifier model and spyder, which, after the analysis of the results, experiments to a final and correct conclusion, play a critical role in the successful execution of the next stage.

5.1 Confusion Matrix

For analyzing the performance of the classification algorithm uses a matrix named confusion matrix in the case of machine learning. It is in tabular form, and the column corresponds to predicted class, where row characterizes actual class (Fig. 6).

In Fig. 7, Logistic Regression Confusion Matrix the Value of True Negative = 36, False Negative = 0, False Positive = 70, True Positive = 197.

$$\begin{aligned}\text{Accuracy} &= (\text{True Positive} + \text{True Positive})/\text{Total No. of Instance} \\ &= (36 + 70)/303 = 0.7690\end{aligned}$$

- Predicted Class: the label of the class after making the classifier generally shows the Predicted Class.
- True Positive: Actual positive values are mainly the number of occurrences predicted in positive.
- False Positive: Actual negative values are mainly the number of occurrences predicted in negative.
- True Negative: Actual negative values are mainly the number of occurrences predicted in positive

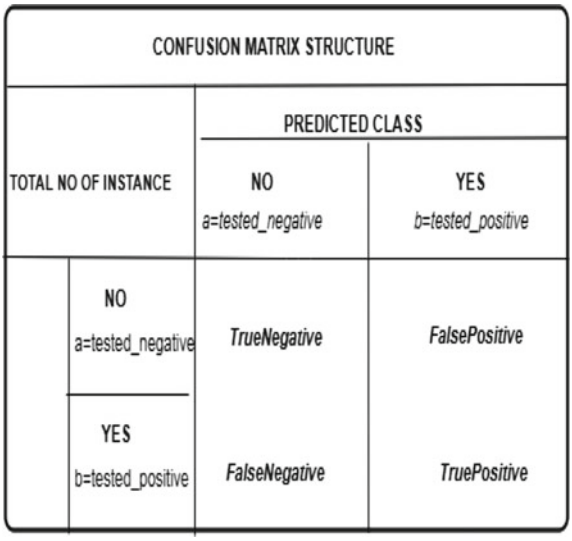


Fig. 6 General confusion matrix structure

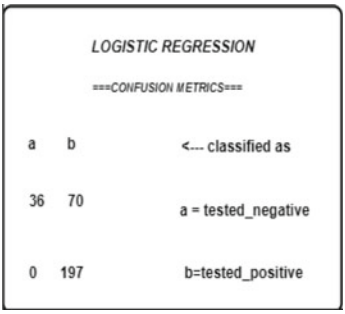


Fig. 7 Logistic regression confusion matrix

- False Negative: Actual positive values are mainly the number of occurrences predicted in negative.
- Total no. of Instances: The final result is the summation of all the occurrences which have been classified by the classifier.

In the given Fig. 8, according to confusion matrix of random forest is the Value of "True Negative" = 106, where False Negative = 0, False Positive = 197, and True Positive = 0.

$$\begin{aligned}\text{Accuracy} &= (\text{True Positive} + \text{True Positive})/\text{Total No. of Instance} \\ &= (106 + 0)/303 = 0.3494\end{aligned}$$

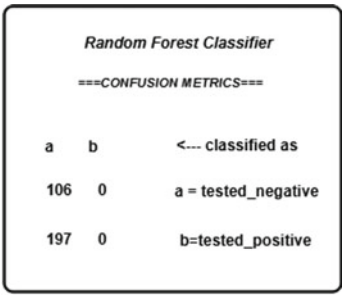


Fig. 8 Random forest confusion matrix

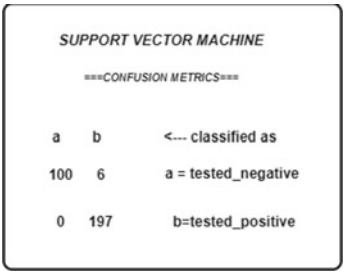


Fig. 9 Support Vector Machine Confusion Matrix

As seen in Fig. 9, the True Negative Value is determined using the Random Forest Confusion Matrix = 100, False Negative = 0, False Positive = 6, True Positive = 197.

$$\begin{aligned}\text{Accuracy} &= (\text{True Positive} + \text{True Positive})/\text{Total No. of Instance} \\ &= (100 + 197)/303 = 0.9801\end{aligned}$$

In general, the results of all experiments are discussed in below section, experiment used Spyder framework.

The accuracy of classification is remarkable, with the maximum accuracy of SVM approximation at 98.01%, where in case of Logistic Regression, accuracy is approximate 76.90% which are improved with respect to Random Forest (Random Forest Accuracy value: 34.94%). if we analysis time constraint then Support Vector Machine (SVM) takes more time than Logistic Regression, where Random Forest have maximum (highest) time. As per the theory of Classification Accuracy, evaluate represent in Fig. 10.

The SVM has the greatest compute value of 98.01% and the random forest has the lowest at 34.90%, where the precision of the support vector machine (SVM) is the highest/maximum at 98.01% and the random forest has the lowest precision value at 34.90%.

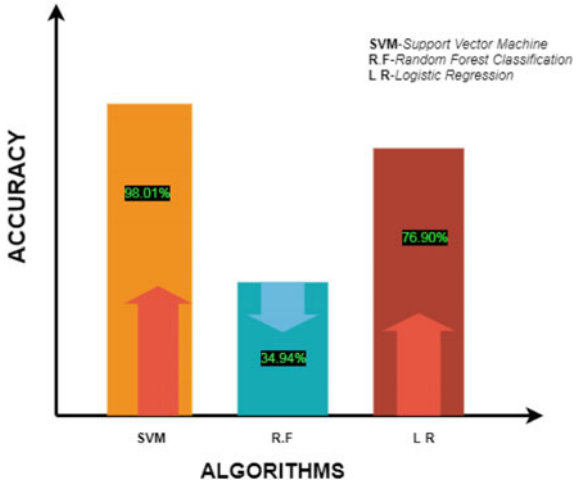


Fig. 10 Comparison of diabetics accuracy

6 Conclusion and Future Scope

Here, three learning techniques are explored to predict T2DM. On the basis of data extracted from patient medical records, we utilized three classification techniques to estimate the probability or absent of a type II diabetes diagnosis. Among some of the three methods used for learning, the SVM model, the random forest, and the final logistic regression, whichever will perform best will be used for the development of the application to predict diabetes. The data set is highly diverse and contained significant amounts of invalid entries. Preprocessing is the key to improve the accuracy.

Android applications can be made for users so that they can predict whether he/she has diabetes or not at any time. Web application can be deployed on the cloud so that anyone can access the web address and make predictions. This model can be developed more accurately by using the high performance machine learning technique which will be developed in the future.

References

1. Reichard P, Pihl M (1994) Mortality and treatment side-effects during long-term intensified conventional insulin treatment in the Stockholm diabetes intervention study. Diabetes 43(2):313–317
2. Diabetes Control Complications Trial Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications of insulin-dependent diabetes mellitus. N Engl J Med 329:978–986
3. UK Prospective Diabetes Study Group (UKPDSG) (1998) Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. Lancet 352:837–853

4. Cryer PE (1999) Hypoglycemia is the limiting factor in the management of diabetes. Diabetes Metab Res Rev 15:42–46
5. Cryer PE (1994) Hypoglycemia: the limiting factor in the management of IDDM. Diabetes 43:1378–1389
6. Sumangali K, Geetika BSR, Ambarkar H (2012) Modular closed-loop control of diabetes, vol 59
7. Andy Liaw MW (2002) Classification and regression by Random Forest. R News 2.3, pp 18–22
8. Bisong E (2019) Introduction to Scikit-learn. In: Building machine learning and deep learning models on Google Cloud platform. Apress, Berkeley CA, pp 215–229
9. Asri H, Mousannif H, Al Moatassime H, Noel T (2015) Big data in healthcare: challenges and opportunities. In: 2015 International conference on cloud technologies and applications (CloudTech), pp 1–7
10. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang G-Z (2015) Big data for health. IEEE J Biomed Heal Inf 19(4):1193–1208
11. Ahmed E et al (2017) The role of big data analytics in Internet of Things. Comput Netw 129:459–471
12. Basole RC, Braunstein ML, Sun J (2015) Data and analytics challenges for a learning healthcare system. J Data Inf Qual (JDIQ) 6(2–3):1–4
13. El Jerjawi NS, Abu-Naser SS (2018) Diabetes prediction using artificial neural network. Int J Adv Sci Technol
14. Rahman F, Slepian MJ (2016) Application of big-data in healthcare analytics—prospects and challenges. In: 2016 IEEE-EMBS International conference on biomedical and health informatics (BHI), pp 13–16
15. Hasan MK, Alam MA, Das D, Hossain E, Hasan M (2020) Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access 8:76516–76531
16. Khanam JJ, Foo SY (2021) A comparison of machine learning algorithms for diabetes prediction. ICT Express
17. Chen D, Yang Y, Zhang J (2015) Methods for predicting type 2 diabetes

Study of Gate Workfunction Modulated FinFET with Effect of TID

Abhishek Ray, Alok Naugarhiya, and Guru Prasad Mishra

Abstract This article presents the study total ionizing dose (TID) effect on gate workfunction modulated silicon-on-insulator (SOI) FinFET (GWM-SOI-FinFET). The incorporation of workfunction engineering enhances the off-state performance. Radiation degrades the off-state leakage for the single metal workfunction-SOI-FinFET (SMW-SOI-FinFET) compared to GWM-SOI-FinFET. The minimal effect of TID on the proposed device is observed for 100 krad at gate bias voltage of 1 V. The GWM-SOI-FinFET mitigated the effect of TID up to the 100 krad.

Keywords FinFET · Single event effects (SEEs) · Single event upset (SEU) · Single event transient (SET) · Dose rate · Total ionizing dose (TID)

1 Introduction

The continuous advancement in semiconductor technology is enabling the scaling and miniaturization of the devices. However, the shrinking of the feature size of metal oxide semiconductor field effect transistors (MOSFETs) to nano-regime results short channel effects [1]. In the continuation of device scaling, fin field effect transistor (FinFET) is introduced as an alternative to conventional MOSFET [1]. FinFET has better controllability of gate voltage over the channel. Therefore, FinFET is also commonly used for modern integrated circuit (IC) fabrication [2]. Radiation responses like total ionizing dose effect (TID) and single event effect (SEE) are the key challenges with the scaling of devices [3]. Single event effects are classified into many types, i.e., single event transient (SET), single event upset (SEU), single event

A. Ray (✉) · A. Naugarhiya · G. P. Mishra
Department of Electronics and Communication Engineering, NIT Raipur, Raipur, India
e-mail: aray.phd2019.ece@nitrr.ac.in

A. Naugarhiya
e-mail: anaugarhiya.etc@nitrr.ac.in

G. P. Mishra
e-mail: gpscmishra.etc@nitrr.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_20

latch-up (SEL), and many more. These effects temporarily degrade the device performance, whereas TID causes the permanent failure of the device [4]. The radiation sensitivity of MOS devices was first discovered at the Naval laboratory in 1965 [5]. The effect of TID was tolerated in short and narrow channel devices. But in the bulk FinFET, degradation of the device is more prominent as compared to long-channel device [6]. In 2020, a work on modified gate stack FinFET is reported, which reduces the hole trapping in gate oxides and improved leakage current after irradiation [7]. In n-type FinFET, InGaAs is used as alternative channel material, because of their high electron mobility [8]. The effect of radiation can also be minimized by changing the fin material from silicon (Si) to indium gallium arsenide (InGaAs) [9]. Under radiation, in SOI devices, the threshold voltage shift is observed due to accumulation of interface trap charges [6]. The effect of TID with different temperatures is quite similar in the manner of charge trapping. However, lower temperature shows better subthreshold swing and drive current [10]. After the irradiation, bulk FinFET shows a huge increment in off current as compared to SOI FinFET. In radiation enrich environment, device current reduces with scaling of fin width. [11]. In 10 keV x-ray chamber, a InGaAs device was investigated with the different gate lengths [12].

Recently, linear modulation of workfunction of gate electrode instead of single workfunction was reported due to its better switching ratio and subthreshold swing [13]. A comparison of single metal workfunction and dual metal workfunction FinFET was presented. In the reported article, improved results are obtained in dual metal workfunction FinFET [14].

This work proposed an n-type gate workfunction modulated-SOI-FinFET (GWM-FinFET). It is proposed to study the effect of TID under radiation enrich environment. Visual TCAD tool is used for the evaluation of device under normal and radiation environment [15, 16]. The rest of the paper is organized as: Sect. 2 gives a brief overview of the device structure; Sect. 3 discusses the results, and last Sect. 4 conclusion.

2 Device Structure

The 2D schematic of the conventional and proposed device is shown in Fig. 1a and b, respectively. The gate electrode is divided into five equal parts with different metal workfunction. The physical parameters of the proposed device are given in Table 1. Figure 2 presents the simulated 3D structure. The Lombardi mobility, Fermi, Band to Band Tunneling, Hot carrier, SRH recombination, Band and Advanced TID model of Visual TCAD are incorporated in simulation.

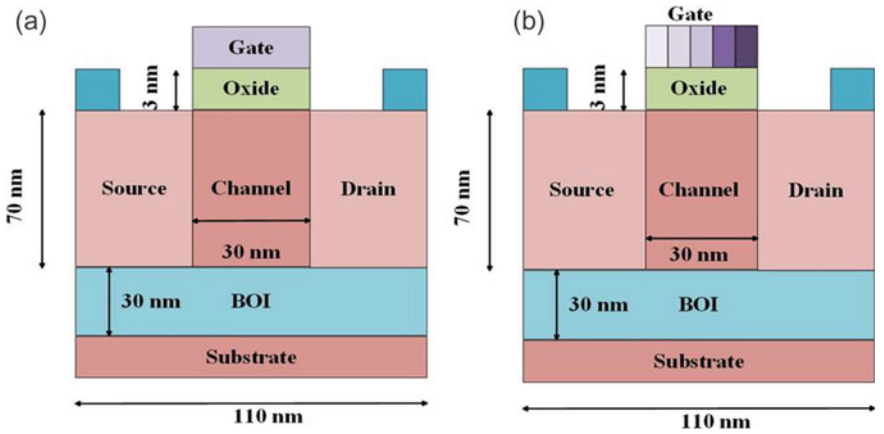


Fig. 1 **a** Conventional SOI FinFET and **b** proposed SOI FinFET

Table 1 Parameters of the proposed device

Parameters		Considered values
Gate length, L_g (nm)		30
Fin width (nm)		10
Fin height (nm)		70
Channel doping, N_A (cm^{-3})		1×10^{15}
Source and drain doping, N_D (cm^{-3})		2×10^{18}
Oxide thickness, t_{ox} (nm)	SiO_2	3
Drain voltage, V_{DS} (V)		0.5
Gate voltage, V_{GS} (V)		1
Annealing temperature (°K)		295
Dose rate (krad)		10
Annealing time (h)		6

3 Results and Discussion

Figure 3 shows the transfer characteristics of SMW-SOI-FinFET and GWM-FinFET with bias condition ($+V_{gs} = 1$ V). The workfunction of GWM-FinFET varies from 4.85 (maximum at drain side) to 4.65 (minimum at source side). The current switching ratio in proposed device improved by factor 3 as compared to SMW-SOI-FinFET. This improvement is obtained because of workfunction engineering technique.

Figure 4 shows the transfer characteristics of conventional and proposed device with and without radiation at $V_{gs} = 1$ V and $V_{DS} = 0.05$ V. The proposed workfunction engineering improves the OFF-state performance of the device than the conventional

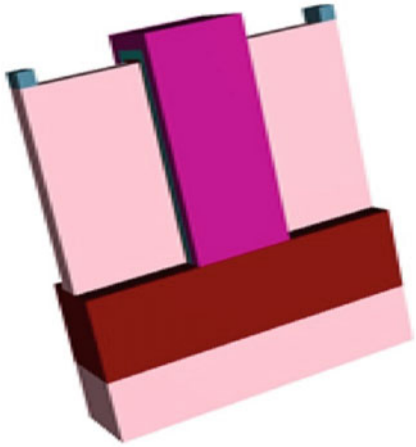


Fig. 2 Simulated 3D SOI FinFET

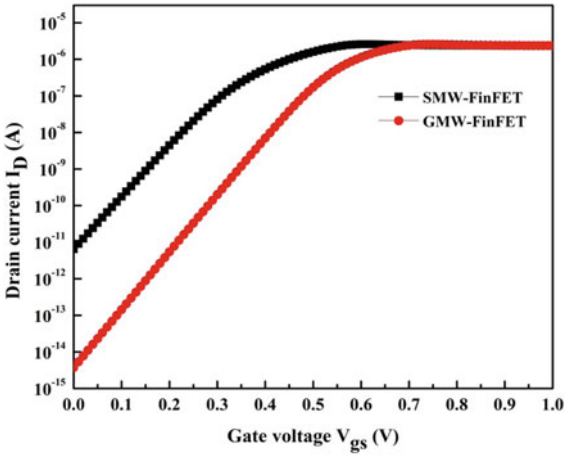


Fig. 3 I_{DS} versus V_{gs} characteristics of SMW-SOI-FinFET and GMW-SOI-FinFET

device. Workfunction modulation engineering increases electron mobility and electric field, which are proportional to the device current, whereas on-state current (I_{on}) of both the device is almost equal.

In the semiconductor/insulator, interface trap charges are found after the radiation induced on the device as shown in Fig. 5a and b, respectively. The fixed trap charges in the proposed are reducing because of the engineering techniques (workfunction modulation) applied to the proposed device. This reduction also enhancing the device performance after radiation. The effect of radiation on electric field is shown in Fig. 6. After the irradiation of device, some unwanted spikes are found in the electric field. Figure 7 shows the effect of trap charges on transfer characteristics of proposed and conventional device. Effect of radiation on transconductance (g_m) is given in Table 2.

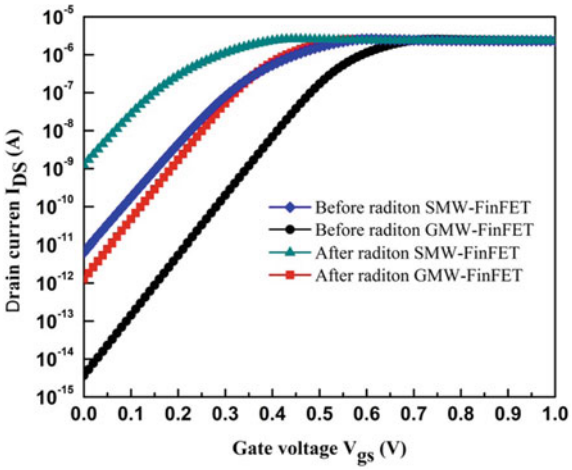


Fig. 4 Effect of TID on I_{DS} versus V_{gs} characteristics of SMW-SOI-FinFET and GMW-SOI-FinFET

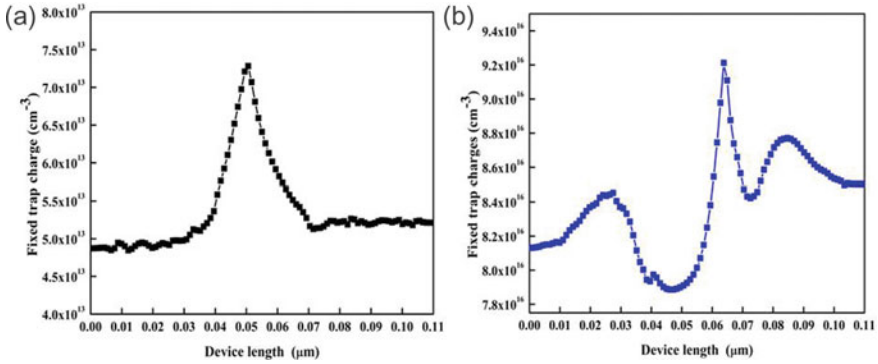


Fig. 5 Fixed oxide charge of **a** GMW-SOI-FinFET and **b** SMW-SOI-FinFET

4 Conclusion

The TID response of GWM-SOI-FinFET has been investigated with channel length 30 nm, 70 nm fin height, and fin width of 10 nm. The implementation of workfunction engineering improved the current switching ratio. The fixed trap charges in BOI and interface of gate oxide are found, which contribute to shifting of threshold voltage, whereas compared to SMW-SOI-FINFET, trap charges are minimum in the proposed device. Because of this the proposed device can able mitigate the effect of TID up to certain krad and improve overall performance of device before and after radiation. This article gives a complete insight into the defects of device after the irradiation.

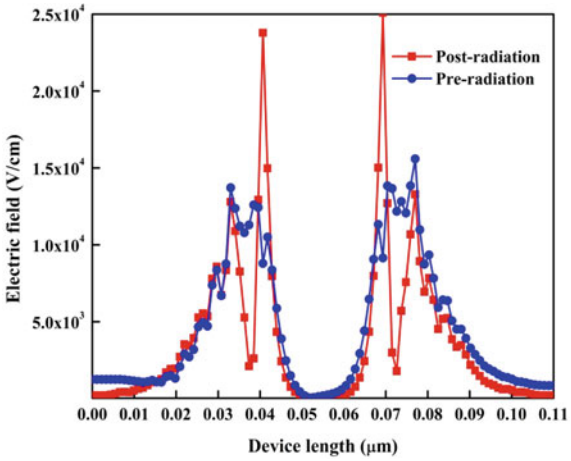


Fig. 6 Electric field across x-axis before radiation and after radiation

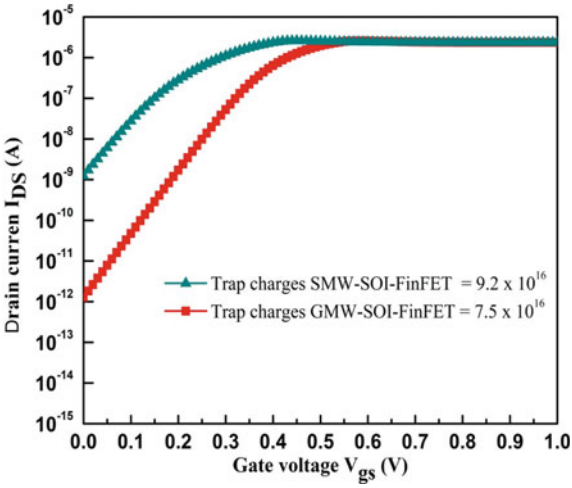


Fig. 7 Effect of trap charges on transfer characteristics of SMW-SOI-FinFET and GMW-SOI-FinFET

Table 2 Effect on g_m before and after radiation

Type	Considered values (μ siemens)
Pre-radiation	
SMW-SOI-FinFET	1.5
GMW-SOI-FinFET	10
Post-radiation	
SMW-SOI-FinFET	1.25
GMW-SOI-FinFET	1.75

References

1. Colinge J-P (ed) (2008) FinFETs and other multi-gate transistors, vol 73. Springer, New York
2. Wilson L (2013) International technology roadmap for semiconductors (ITRS). Semiconductor Industry Association, 1
3. Hughes HL, Benedetto JM (2003) Radiation effects and hardening of MOS technology: devices and circuits. IEEE Trans Nucl Sci 50(3):500–521
4. Gaillardin M et al (2017) Total ionizing dose effects in multiple-gate field-effect transistor. Semicond Sci Technol 32(8):083003
5. Raymond J, Steele E, Chang W (1965) Radiation effects in metal-oxide-semiconductor transistors. IEEE Trans Nucl Sci 12(1):457–463
6. Gorchichko M et al (2019) Total-ionizing-dose effects and low-frequency noise in 30-nm gate-length bulk and SOI FinFETs with SiO_2/HfO_2 gate dielectrics. IEEE Trans Nucl Sci 67(1):245–252
7. Zhao SE, Bonaldo S, Wang P, Zhang EX, Waldron N, Collaert N, Putcha V, Linten D, Gerardin S, Paccagnella A, Schrimpf RD (2019) Total-ionizing-dose effects on InGaAs FinFETs with modified gate-stack. IEEE Trans Nucl Sci 67(1):253–259
8. McEwen A, Cassimally H (2013) Designing the internet of things. Wiley
9. Bonaldo S, Zhao SE, O'Hara A, Gorchichko M, Zhang EX, Gerardin S, Paccagnella A, Waldron N, Collaert N, Putcha V, Linten D (2019) Total-ionizing-dose effects and low-frequency noise in 16-nm InGaAs FinFETs with HfO_2/Al_2O_3 dielectrics. IEEE Trans Nucl Sci 67(1):210–220
10. Haeffner TD, Keller RF, Jiang R, Sierawski BD, McCurdy MW, Zhang EX, Mohammed RW, Ball DR, Alles ML, Reed RA, Schrimpf RD (2019) Comparison of total-ionizing-dose effects in bulk and SOI FinFETs at 90 and 295 K. IEEE Trans Nucl Sci 66(6):911–917
11. Wang J, An X, Ren Z, Li G, Zhang W, Huang R (2019) A novel body-on-insulator (BOI) FinFET with excellent TID tolerance and scaling capability. Semicond Sci Technol 34(10):105008
12. Zhao SE, Bonaldo S, Wang P, Jiang R, Gong H, Zhang EX, Waldron N, Kunert B, Mitard J, Collaert N, Sioncke S (2019) Gate bias and length dependences of total ionizing dose effects in InGaAs FinFETs on bulk Si. IEEE Trans Nucl Sci 66(7):1599–1605
13. Daga M, Mishra GP (2021) Subthreshold performance improvement of underlapped FinFET using workfunction modulated dual-metal gate technique. Silicon 13(5):1541–1548
14. Daga M, Mishra GP (2021) Improvement in electrostatic efficiency using workfunction modulated dual metal gate FinFET. Mater Today: Proc 43:3443–3446
15. Simulator SD, User V (2013) VisualTCAD
16. Model M, Manual R (2016) Genius semiconductor device simulator

Prediction of Accuracy and Performance of Crop Production by Using Machine Learning Algorithm

Ramashankar Yadav, Pawan Kumar Singh, Ajay Sikandar, Manoj Kumar Tyagi, and Dinesh Kumar Yadav

Abstract Machine learning is an emergent field of research in crop production analysis. The present study aims at shedding light on machine learning in agriculture along with "crop management," "water management," "soil management," "fertilizer management." Production prediction is a very important issue in agriculture. Any farmer wants to know what his expected production is about to expect. In the past, production predictions were made by considering farmers' experience in specific fields and crops. Based on existing data, performance prediction is a major problem to be solved. Machine learning technology is the best choice for this purpose. Different machine learning techniques are used and evaluated in agriculture to estimate crop productions in the coming year. The aim of this paper proposes and implements a system that can predict crop productions based on previous data. This is achieved by applying machine learning algorithms like Naïve Bayes, generalized linear model, deep learning, decision tree, and random forest to agricultural data and recommending suitable fertilizers for each specific crop and it has been observed that random forest algorithm achieved better accuracy and performance to predict future crop productions.

R. Yadav (✉) · P. K. Singh · A. Sikandar
Department of Computer Science and Engineering, G. L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: ramashankar.khushi@gmail.com

P. K. Singh
e-mail: pawan8832006@gmail.com

A. Sikandar
e-mail: ajay.sikandar@gmail.com

M. K. Tyagi
Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: manojankitsuhani@gmail.com

D. K. Yadav
Department of Electronics and Communication Engineering, IIMT College of Engineering, Greater Noida, India
e-mail: dinesh.vlsi.mod@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_21

Keywords Agriculture data · Production prediction · Machine learning algorithm · Soil types · Fertilizers

1 Introduction

Agriculture is the main source of livelihood and plays a vital role in the economies of most countries. It guarantees the food safety of a country and produces materials for the industry. Since ancient times, agriculture has been regarded as the main and the most important crop cultivated in India. The ancient people cultivated their territory and so they were to meet their requirements. Since the development of new advanced technologies and systems in the agricultural sector gradually increases. Therefore, invention abundant people focus on cultivating artificial products which are hybrid products leading to an unhealthy life. Nowadays, modern man is not aware of planting at the right time and in the correct location. As a result of these farming techniques, seasonal climatic conditions are also changed from basic assets such as soil, water, and air resulting in food insecurity.

Machine Learning algorithms include supervised learning, unsupervised learning, and reinforced learning, each having its own set of benefits and drawbacks.

Supervised Learning: This method generates a mathematical model from a set of data comprising the desired inputs and outputs.

Unsupervised Learning: This program generates a mathematical model from a data set that just contains inputs and does not have the necessary output labels.

Semi-supervised Learning: This approach creates mathematical models from partial training data, in which some of the sample input has no labels.

The goal of this paper is to boost crop production using various ways and to prescribe appropriate fertilizers for each yield.

2 Literature Survey

This work is to analyze the main characteristics of the soil, such as organic matter, plant essential nutrients, and micronutrients that affect crop growth, and to understand the appropriate relationship percentages between these characteristics through supervised learning and back-propagation neural networks [1].

This document seeks to increase agricultural output in a variety of methods and proposes fertilizer that is appropriate for each crop [2].

In this work, RapidMiner 5.3 was used for experiments. Two important and well-known classification algorithms K-Nearest Neighbor (k-NN) and Naive Bayes (NB) are applied to the soil data set taken from the Jabalpur Soil Analysis Laboratory. The accuracy of M.P. is obtained by evaluating the data set. Each algorithm was run

on the training data set, and its performance on the accuracy and its prediction on the test data sets were evaluated [3].

This article selects the nutrients Naive Bayes, decision trees, and hybrid Naive Bayes method, as well as decision trees to study soil supplementation, such as nitrogen, phosphorus, potassium, calcium, magnesium, sulfur, iron, zinc. The ranking algorithm's performance is compared based on the two factors: accuracy and performance time [4].

This document consists of a soil database of farm holdings, crops supplied by farmers, and soil parameters obtained by means of laboratory soil analysis. The data is given to the recommendation system from the soil analysis laboratory. You will use the data collected and the voting technology for a majority. To recommend site-specific high-parameter cultures with exact high precision and efficiency, use the vector support and artificial neural network (SVM) as students [5].

This article is characterized by soil databases from farms, crops made available by farm experts and soil parameters obtained from laboratory data sets. Soil analysis for the recommendation system and dataset was access from Kaggle. You are using the data you collect and voting technology by majority voting. Use SVM as an instructor to recommend site-specific culture with high precision and efficiency [5] high parameters of culture [6].

For agricultural research or knowledge, the proposed work seems useful to learn to eliminate adversity and achieve a better performance of the various soil nutrients and climactic factors. Although the work on this subject focuses on soil nutrients and crop production in the middle and highlands, other types of soil and additional soil content (e.g., water level, soil humidity, soil type, and fertilizers) can also be considered [7].

3 Classification of Machine Learning Algorithms

Following machine learning algorithms have been used to predict crop yield prediction are as follows:

3.1 Naive Bayes Classifier

The Naive Bayes classifier is a collection of simple "probabilistic classifiers" based on Bayes' theorem and the assumption of strong (naive) independence between attributes in statistics. They belong to the simplest Bayesian network model [1], but combined with the estimation of kernel density, they can achieve a higher level of precision [2, 3]. We applied data set into Naïve Bayes classifier and observed that accuracy is 79.30% and classification error is 20.7% as shown in Fig. 1. This model is highly suggested to loamy for sugarcane is 21% and black soil 89% or loamy is 21% and black soil 89% for oil seeds or sandy is 76% and black is 24% for maize to crop

Fig. 1 Naïve Bayes accuracy

yield production as given in Table 1 and performance of Naïve Bayes model shown in Fig. 2 and confusion matrix as shown in Fig. 3. This model calculates accuracy and error as shown in Fig. 4 and crop production model of Naïve Bayes as shown in Fig. 5.

3.2 Generalized Linear Model (GLM)

In order to unify the linear regression, logistical reconstruction, and Poisson regression, the GLM was formulated by Nelder and Weddingburn [8]. They proposed an iterative less squared method for the maximum probability evaluation of the model parameters. They presented an iterative, weighed less square method for estimating the maximum likelihood of the model parameters. Maximum likelihood estimation is still widely used, and many statistical software packages use it as the default technique. We applied data set into GLM and observed that accuracy is 78.00% and classification error is 22.00% as shown in Fig. 6. This model is highly suggested to loamy is 12% and black soil is 88% for sugarcane or loamy is 12% and black soil is 88% for oil seeds to crop yield production as given in Table 2 performance of GLM model is shown in Fig. 7 along with confusion matrix as shown in Fig. 8. This model calculates accuracy and error as shown in Fig. 9 and crop production model of GLM as shown in Fig. 10.

3.3 Deep Learning Model (DLM)

Recent deep learning patterns include artificial neural networks, especially convolutional neural networks in deep learning model (CNNs). Deep generational models like deep belief networks and machine learning nodes include layer-structured proposal formulae or latent variables [9]. Each level of deep learning learns to turn the data it receives into a slightly abstract and compound representation.

Table 1 Naïve Bayes prediction of crop production

S. No.	Soil type	Soil prediction	Confident (sandy)	Confident (loamy)	Confident (black)	Confident (red)	Confident (clayey)	Fertilizer	Temperature	Humidity	Moisture	Nitrogen	Potassium	Phosphorus	Cost	Crop type
1	Red	Red	0.109764	7.25E-07	0.243162	0.647073373	2.97E-08	28–28	32	62	34	22	0	20	0.294146746	Tobacco
2	Sandy	Sandy	0.756267	7.25E-07	0.243732	3.23E-08	3.15E-08	17–17-17	26	52	35	12	10	13	0.512533546	Barley
3	Loamy	Loamy	0.109764	0.64753248	0.242704	2.90E-08	2.84E-08	Urea	33	64	50	41	0	0	0.295064961	Wheat
4	Clayey	Clayey	0.109764	7.25E-07	0.24259	2.87E-08	0.647645014	DAP	27	54	28	13	0	40	0.295290028	Pulses
5	Clayey	Clayey	0.109764	7.25E-07	0.242705	2.90E-08	0.647530604	28–28	32	62	41	24	0	22	0.295061207	Paddy
6	Clayey	Clayey	0.109764	7.25E-07	0.24259	2.87E-08	0.647645014	Urea	33	64	34	38	0	0	0.295290028	Pulses
7	Sandy	Sandy	0.756267	7.25E-07	0.243732	3.23E-08	3.15E-08	28–28	25	50	39	21	0	19	0.512533546	Barley
8	**Loamy**	**Black**	**0.109764**	**3.63E-06**	**0.890232**	**5.39E-08**	**5.28E-08**	**17–17–17**	**34**	**65**	**53**	**12**	**14**	**12**	**0.780464726**	**Sugarcane**
9	Red	Red	0.109764	7.25E-07	0.243162	0.647073373	2.97E-08	DAP	35	68	33	11	0	37	0.294146746	Tobacco
10	**Black**	**Black**	**0.109764**	**7.25E-07**	**0.890235**	**3.10E-08**	**3.03E-08**	**DAP**	**37**	**70**	**32**	**12**	**0**	**39**	**0.78047066**	**Oil seeds**
11	Sandy	Sandy	0.756267	7.25E-07	0.243732	3.23E-08	3.15E-08	14–35-14	36	68	38	7	9	30	0.512533546	Barley
12	Black	Sandy	0.548819	7.25E-07	0.45118	2.32E-08	2.27E-08	Urea	28	54	35	41	0	0	0.097638575	Millets
13	Loamy	Black	0.109764	7.25E-07	0.763997	0.126238133	4.51E-08	14–35-14	30	60	61	8	10	31	0.527994401	Cotton
14	**Loamy**	**Black**	**0.109764**	**3.63E-06**	**0.890232**	**5.39E-08**	**5.28E-08**	**14–35-14**	**30**	**60**	**58**	**10**	**7**	**32**	**0.780464726**	**Sugarcane**
15	Red	Red	0.109764	7.25E-07	0.243162	0.647073373	2.97E-08	28–28	32	62	34	22	0	24	0.294146746	Tobacco
16	**Black**	**Black**	**0.109764**	**3.63E-06**	**0.890232**	**5.39E-08**	**5.28E-08**	**Urea**	**34**	**65**	**60**	**35**	**0**	**0**	**0.780464726**	**Sugarcane**
17	Loamy	Loamy	0.109764	0.64753248	0.242704	2.90E-08	2.84E-08	14–35-14	38	70	48	8	8	28	0.295064961	Wheat
18	**Black**	**Black**	**0.109764**	**7.25E-07**	**0.890235**	**3.10E-08**	**3.03E-08**	**20–20**	**26**	**52**	**32**	**12**	**0**	**8**	**0.78047066**	**Oil seeds**
19	Clayey	Clayey	0.109764	7.25E-07	0.242705	2.90E-08	0.647530604	28–28	29	58	43	24	0	18	0.295061207	Paddy
20	Red	Red	0.109764	7.25E-07	0.242705	0.647530604	2.84E-08	Urea	30	60	29	41	0	0	0.295061209	Groundnuts
21	Sandy	Sandy	0.756837	7.25E-07	0.243162	3.04E-08	2.97E-08	14–35-14	33	64	51	5	9	29	0.513674451	Maize

(continued)

Table 1 (continued)

S. No.	Soil type	Soil prediction	Confident (sandy)	Confident (loamy)	Confident (black)	Confident (red)	Confident (clayey)	Fertilizer	Temperature	Humidity	Moisture	Nitrogen	Potassium	Phosphorus	Cost	Crop type
22	**Black**	**Black**	**0.109764**	**7.25E-07**	**0.890235**	**3.10E-08**	**3.03E-08**	**20–20**	**36**	**68**	**33**	**13**	**0**	**14**	**0.78047066**	**Oil seeds**
23	Clayey	Clayey	0.109764	7.25E-07	0.24259	2.87E-08	0.647645014	Urea	28	54	38	40	0	0	0.295290028	Pulses
24	Sandy	Sandy	0.756267	7.25E-07	0.243732	3.23E-08	3.15E-08	DAP	30	60	47	12	0	42	0.512533546	Barley
25	Red	Black	0.109764	7.25E-07	0.763997	0.126238133	4.51E-08	17–17-17	31	62	63	11	12	15	0.527994401	Cotton
26	Loamy	Loamy	0.109764	0.64753248	0.242704	2.90E-08	2.84E-08	Urea	34	65	54	38	0	0	0.295064961	Wheat
27	**Loamy**	**Black**	**0.109764**	**3.63E-06**	**0.890232**	**5.39E-08**	**5.28E-08**	**17–17-17**	**25**	**50**	**56**	**11**	**13**	**15**	**0.780464726**	**Sugarcane**
28	Red	Red	0.109764	7.25E-07	0.242705	0.647530604	2.84E-08	DAP	32	62	34	15	0	37	0.295061209	Groundnuts
29	Black	Black	0.109764	7.25E-07	0.763997	0.126238133	4.51E-08	28–28	34	65	64	24	0	20	0.527994401	Cotton
30	Sandy	Sandy	0.756267	7.25E-07	0.243732	3.23E-08	3.15E-08	10–26-26	28	54	47	5	18	15	0.512533546	Barley
31	Clayey	Clayey	0.109764	7.25E-07	0.24259	2.87E-08	0.647645014	10–26-26	25	50	40	6	19	16	0.295290028	Pulses
32	Red	Red	0.109764	7.25E-07	0.243162	0.647073373	2.97E-08	Urea	31	62	32	39	0	0	0.294146746	Tobacco
33	**Black**	**Black**	**0.109764**	**7.25E-07**	**0.890235**	**3.10E-08**	**3.03E-08**	**Urea**	**27**	**53**	**34**	**42**	**0**	**0**	**0.78047066**	**Oil seeds**
34	Clayey	Clayey	0.109764	7.25E-07	0.242705	2.90E-08	0.647530604	14–35-14	29	58	42	9	10	22	0.295061207	Paddy
35	Sandy	Sandy	0.756837	7.25E-07	0.243162	3.04E-08	2.97E-08	28–28	30	60	47	22	0	21	0.513674451	Maize
36	Red	Black	0.109764	7.25E-07	0.763997	0.126238133	4.51E-08	DAP	34	65	63	14	0	38	0.527994401	Cotton
37	Clayey	Clayey	0.109764	7.25E-07	0.242705	2.90E-08	0.647530604	14–35-14	28	54	43	10	8	29	0.295061207	Paddy
38	Black	Black	0.109764	7.25E-07	0.763997	0.126238133	4.51E-08	DAP	29	58	65	14	0	35	0.527994401	Cotton
39	**Loamy**	**Black**	**0.109764**	**3.63E-06**	**0.890232**	**5.39E-08**	**5.28E-08**	**20–20**	**26**	**52**	**59**	**11**	**0**	**9**	**0.780464726**	**Sugarcane**
40	Sandy	Sandy	0.756837	7.25E-07	0.243162	3.04E-08	2.97E-08	14–35-14	28	54	25	9	10	30	0.513674451	Maize

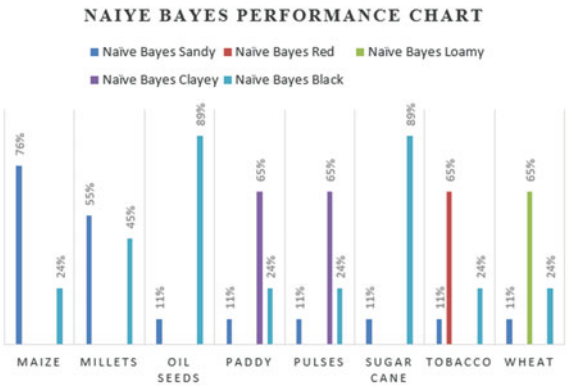


Fig. 2 Naïve Bayes performance

Confusion Matrix

	true Sandy	true Loamy	true Black	true Red	true Clayey	class precision
pred. Sandy	6	0	1	0	0	85.71%
pred. Loamy	0	3	0	0	0	100.00%
pred. Black	0	3	6	2	0	54.55%
pred. Red	0	0	3	4	0	100.00%
pred. Clayey	0	0	0	0	4	100.00%
class recall	100.00%	50.00%	85.71%	66.67%	100.00%	

Fig. 3 Naïve Bayes confusion matrix chart

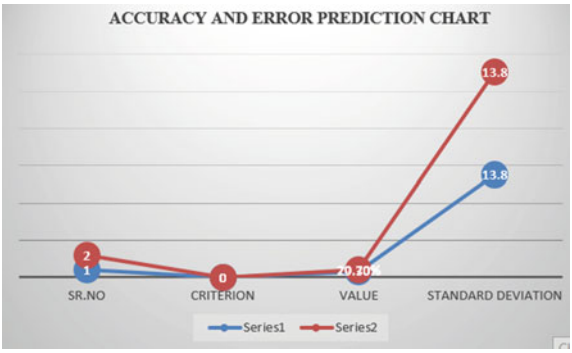


Fig. 4 Accuracy versus error prediction

We applied data set into DLM auto model of rapid miner tool and observed that accuracy is 78.70% and classification error is 21.30% as shown in Fig. 11. This model is highly suggested to black soil is 74% and loamy soil is 16% and sandy soil is 10% for oil seeds to crop production as shown in Fig. 12 along with confusion matrix

Fig. 5 Production model

Fig. 6 GLM accuracy

as shown in Fig. 14. This model calculates accuracy and error as shown in Fig. 13 and crop production model of DL as shown in Fig. 16 along with crop production prediction as shown in Fig. 15.

3.4 Decision Tree (DT)

A decision tree is a tree including a set of nodes which aims to decide on the value affiliation of a certain class or the statistical value estimation. For a particular attribute, every node represents a divisional rule. This rule separates the values belonging to various categories and, to minimize the standard error of the selected parameter, it differentiates the value of the values for regression.

Repeat the build of new nodes until they are stopped. The prediction of the class attribute of the tag depends on most of the examples that come in the leaf during generation and averaging the values on the leaf results in the value estimate. We

Table 2 GLM prediction of crop production

S. No.	Soil type	Soil prediction	Confident (sandy)	Confident(loamy)	Confident (black)	Confident (red)	Confident (clayey)	Fertilizer	Temperature	Humidity	Moisture	Nitrogen	Potassium	Phosphorus	Cost	Crop type
1	Red	Red	0.123153	1.89E-06	0.305519	0.571326786	5.33E-08	28–28	32	62	34	22	0	20	0.142653571	Tobacco
2	Sandy	Sandy	0.673755	1.89E-06	0.326242	6.48E-07	9.80E-08	17–17-17	26	52	35	12	10	13	0.347510115	Barley
3	Loamy	Loamy	0.123153	0.591224558	0.285623	5.27E-08	3.90E-08	Urea	33	64	50	41	0	0	0.182449116	Wheat
4	Clayey	Clayey	0.123153	1.89E-06	0.280007	4.87E-08	0.596838543	DAP	27	54	28	13	0	40	0.193677087	Pulses
5	Clayey	Clayey	0.123153	1.89E-06	0.286277	5.34E-08	0.590568482	28–28	32	62	41	24	0	22	0.181136964	Paddy
6	Clayey	Clayey	0.123153	1.89E-06	0.280007	4.87E-08	0.596838543	Urea	33	64	34	38	0	0	0.193677087	Pulses
7	Sandy	Sandy	0.673755	1.89E-06	0.326242	6.48E-07	9.80E-08	28–28	25	50	39	21	0	19	0.347510115	Barley
8	**Loamy**	**Black**	**0.123153**	**9.44E-06**	**0.876838**	**2.10E-07**	**1.61E-07**	**17–17-17**	**34**	**65**	**53**	**12**	**14**	**12**	**0.753675127**	**Sugarcane**
9	Red	Red	0.123153	1.89E-06	0.305519	0.571326786	5.33E-08	DAP	35	68	33	11	0	37	0.142653571	Tobacco
10	**Black**	**Black**	**0.123153**	**1.89E-06**	**0.876845**	**1.47E-07**	**5.84E-08**	**DAP**	**37**	**70**	**32**	**12**	**0**	**39**	**0.753690632**	**Oil seeds**
11	Sandy	Sandy	0.673755	1.89E-06	0.326242	6.48E-07	9.80E-08	14–35-14	36	68	38	7	9	30	0.347510115	Barley
12	Black	Sandy	0.615763	1.89E-06	0.384235	4.93E-08	4.10E-08	Urea	28	54	35	41	0	0	0.231525286	Millets
13	Loamy	Black	0.123153	1.89E-06	0.49348	0.383365134	1.13E-07	14–35-14	30	60	61	8	10	31	-0.013039508	Cotton
14	**Loamy**	**Black**	**0.123153**	**9.44E-06**	**0.876838**	**2.10E-07**	**1.61E-07**	**14–35-14**	**30**	**60**	**58**	**10**	**7**	**32**	**0.753675127**	**Sugarcane**
15	Red	Red	0.123153	1.89E-06	0.305519	0.571326786	5.33E-08	28–28	32	62	34	22	0	24	0.142653571	Tobacco
16	**Black**	**Black**	**0.123153**	**9.44E-06**	**0.876838**	**2.10E-07**	**1.61E-07**	**Urea**	**34**	**65**	**60**	**35**	**0**	**0**	**0.753675127**	**Sugarcane**
17	Loamy	Loamy	0.123153	0.591224558	0.285623	5.27E-08	3.90E-08	14–35-14	38	70	48	8	8	28	0.182449116	Wheat
18	**Black**	**Black**	**0.123153**	**1.89E-06**	**0.876845**	**1.47E-07**	**5.84E-08**	**20–20**	**26**	**52**	**32**	**12**	**0**	**8**	**0.753690632**	**Oil seeds**
19	Clayey	Clayey	0.123153	1.89E-06	0.286277	5.34E-08	0.590568482	28–28	29	58	43	24	0	18	0.181136964	Paddy
20	Red	Red	0.123153	1.89E-06	0.285569	0.591276324	3.90E-08	Urea	30	60	29	41	0	0	0.182552647	Ground Nuts
21	Sandy	Sandy	0.694279	1.89E-06	0.305719	1.28E-07	5.35E-08	14–35-14	33	64	51	5	9	29	0.388557863	Maize

(continued)

Table 2 (continued)

S. No.	Soil type	Soil prediction	Confident (sandy)	Confident(loamy)	Confident (black)	Confident (red)	Confident (clayey)	Fertilizer	Temperature	Humidity	Moisture	Nitrogen	Potassium	Phosphorus	Cost	Crop type
22	**Black**	**Black**	**0.123153**	**1.89E-06**	**0.876845**	**1.47E-07**	**5.84E-08**	**20–20**	**36**	**68**	**33**	**13**	**0**	**14**	**0.753690632**	**Oil seeds**
23	Clayey	Clayey	0.123153	1.89E-06	0.280007	4.87E-08	0.596838543	Urea	28	54	38	40	0	0	0.193677087	Pulses
24	Sandy	Sandy	0.673755	1.89E-06	0.326242	6.48E-07	9.80E-08	DAP	30	60	47	12	0	42	0.347510115	Barley
25	Red	Black	0.123153	1.89E-06	0.49348	0.383365134	1.13E-07	17–17-17	31	62	63	11	12	15	-0.013039508	Cotton
26	Loamy	Loamy	0.123153	0.591224558	0.285623	5.27E-08	3.90E-08	Urea	34	65	54	38	0	0	0.182449116	Wheat
27	**Loamy**	**Black**	**0.123153**	**9.44E-06**	**0.876838**	**2.10E-07**	**1.61E-07**	**17–17-17**	**25**	**50**	**56**	**11**	**13**	**15**	**0.753675127**	**Sugarcane**
28	Red	Red	0.123153	1.89E-06	0.285569	0.591276324	3.90E-08	DAP	32	62	34	15	0	37	0.182552647	Ground Nuts
29	Black	Black	0.123153	1.89E-06	0.49348	0.383365134	1.13E-07	28–28	34	65	64	24	0	20	-0.013039508	Cotton
30	Sandy	Sandy	0.673755	1.89E-06	0.326242	6.48E-07	9.80E-08	10–26-26	28	54	47	5	18	15	0.347510115	Barley
31	Clayey	Clayey	0.123153	1.89E-06	0.280007	4.87E-08	0.596838543	10–26-26	25	50	40	6	19	16	0.193677087	Pulses
32	Red	Red	0.123153	1.89E-06	0.305519	0.571326786	5.33E-08	Urea	31	62	32	39	0	0	0.142653571	Tobacco
33	**Black**	**Black**	**0.123153**	**1.89E-06**	**0.876845**	**1.47E-07**	**5.84E-08**	**Urea**	**27**	**53**	**34**	**42**	**0**	**0**	**0.753690632**	**Oil seeds**
34	Clayey	Clayey	0.123153	1.89E-06	0.286277	5.34E-08	0.590568482	14–35-14	29	58	42	9	10	22	0.181136964	Paddy
35	Sandy	Sandy	0.694279	1.89E-06	0.305719	1.28E-07	5.35E-08	28–28	30	60	47	22	0	21	0.388557863	Maize
36	Red	Black	0.123153	1.89E-06	0.49348	0.383365134	1.13E-07	DAP	34	65	63	14	0	38	-0.013039508	Cotton
37	Clayey	Clayey	0.123153	1.89E-06	0.286277	5.34E-08	0.590568482	14–35-14	28	54	43	10	8	29	0.181136964	Paddy
38	Black	Black	0.123153	1.89E-06	0.49348	0.383365134	1.13E-07	DAP	29	58	65	14	0	35	-0.013039508	Cotton
39	**Loamy**	**Black**	**0.123153**	**9.44E-06**	**0.876838**	**2.10E-07**	**1.61E-07**	**20–20**	**26**	**52**	**59**	**11**	**0**	**9**	**0.753675127**	**Sugarcane**
40	Sandy	Sandy	0.694279	1.89E-06	0.305719	1.28E-07	5.35E-08	14–35-14	28	54	25	9	10	30	0.388557863	Maize

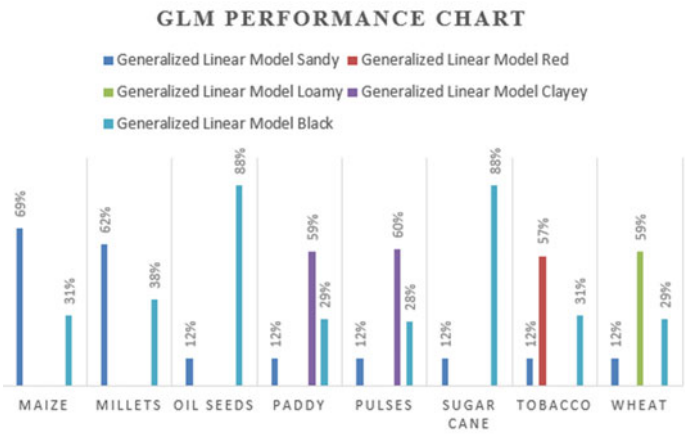


Fig. 7 Performance of GLM

Confusion Matrix

	true Sandy	true Loamy	true Black	true Red	true Clayey	class precision
pred. Sandy	5	1	1	0	0	71.43%
pred. Loamy	0	3	0	0	0	100.00%
pred. Black	0	3	5	1	0	55.56%
pred. Red	0	0	0	4	0	100.00%
pred. Clayey	0	0	0	0	5	100.00%
class recall	100.00%	42.86%	83.33%	80.00%	100.00%	

Fig. 8 Confusion matrix of GLM

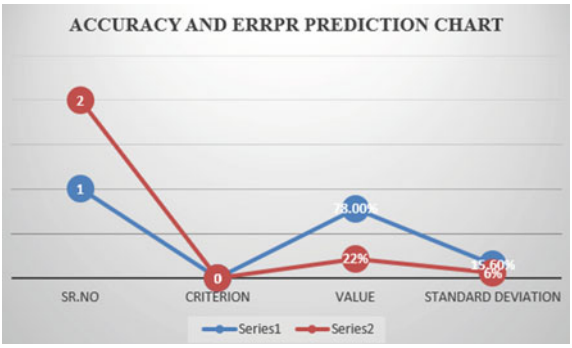


Fig. 9 Accuracy versus error prediction

applied data set into DT auto model perform on rapid miner and observed that accuracy is 79.30% and classification error is 20.70% as shown in Fig. 17. The decision tree is highly suggested to sandy soil is 81% or black soil is 19% for maize crop, black soil is 100% for oil seeds, clayey soil is 81% or black soil is 19% for paddy or black soil is 100% for sugarcane, black soil is 19% for pulses or red soil is 81% or black is 19% for tobacco, loamy soil is 81% or black is 19% for wheat

Generalized Linear Model - Production Model

Attribute	Coefficient San...	Coefficient Loa...	Coefficient Black	Coefficient Red	Coefficient Clay...	Std. Coefficient ...	Std. Coefficient ...	Std. Coefficient ...	Std. Coefficient ...	Std. Coefficient...
Crop Type.Barley	1.837	0	0	0	0	1.837	0	0	0	0
Crop Type.Cotton	-0.545	0	0.152	0.575	-0.357	-0.545	0	0.152	0.575	-0.357
Crop Type.Groun...	0	0	0	1.848	0	0	0	0	1.848	0
Crop Type.Maize	1.837	0	0	0	0	1.837	0	0	0	0
Crop Type.Millets	1.109	-0.229	0	-0.298	-0.132	1.109	-0.229	0	-0.298	-0.132
Crop Type.Oil se...	0	0	1.809	0	0	0	0	1.809	0	0
Crop Type.Paddy	0	0	-0.955	0	2.011	0	0	-0.955	0	2.011
Crop Type.Pulses	0	0	0	0	2.395	0	0	0	0	2.395
Crop Type.Sugar...	-0.395	1.617	0	-0.357	-0.188	-0.395	1.617	0	-0.357	-0.188
Crop Type.Tobac...	0	0	0	2.006	0	0	0	0	2.006	0
Crop Type.Wheat	0	2.244	0	0	0	0	2.244	0	0	0
Crop Type.WISS...	0	0	0	0	0	0	0	0	0	0
Intercept	-1.650	-1.708	-1.447	-1.671	-1.681	-1.650	-1.708	-1.447	-1.671	-1.681

Fig. 10 Production model of GLM

Deep Learning - Performance

Profits

Profits from Model: 16 Profits for Best Option (Loamy): -14 Gain: 30 Show Costs / Benefits...

Performances

Criterion	Value	Standard Deviation
Accuracy	78.7%	± 6.9%
Classification Error	21.3%	± 6.9%

Fig. 11 Accuracy of deep learning

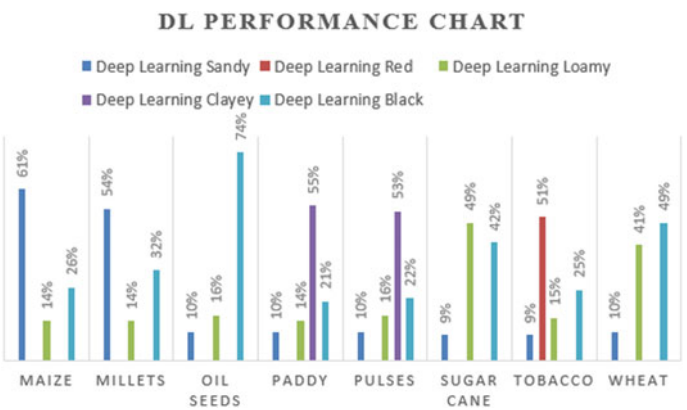


Fig. 12 Performance of DL

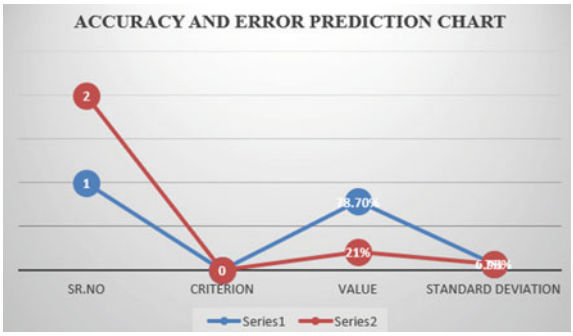


Fig. 13 Performance of DL

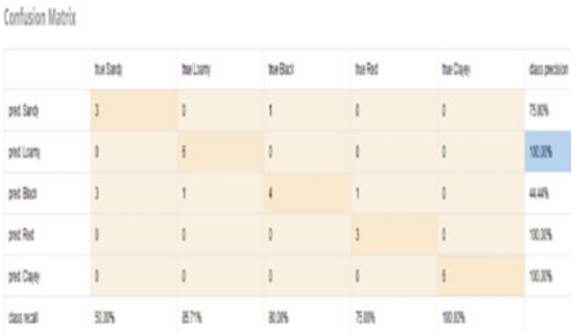


Fig. 14 Performance of DL

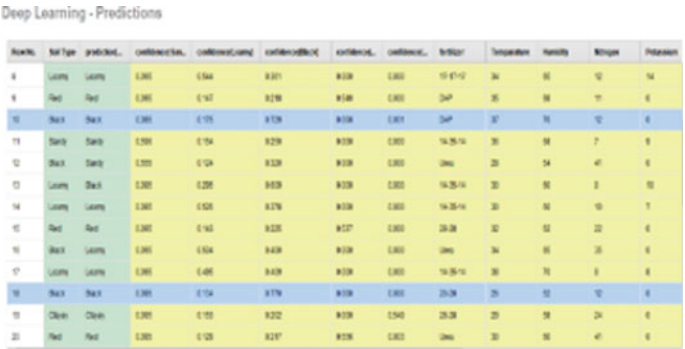


Fig. 15 Prediction of DL

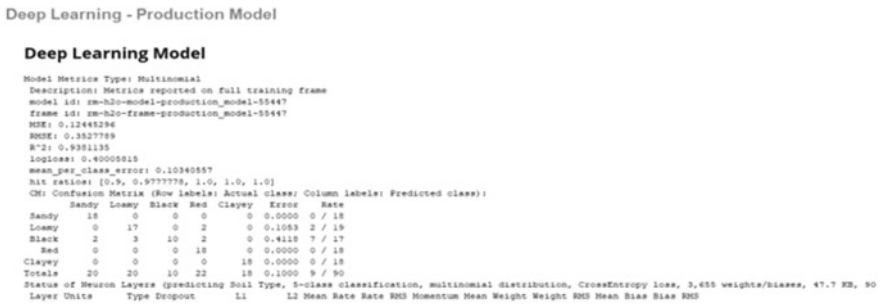


Fig. 16 Production model of DL

Fig. 17 Accuracy of DT

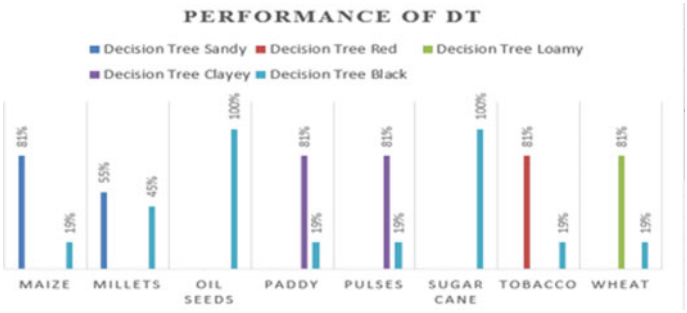


Fig. 18 Performance of DT

to crop yield production as shown in Fig. 18 along with confusion matrix is shown in Fig. 20. This model calculates accuracy and error as shown in Fig. 19 and crop production model of DT as shown in Fig. 22 along with crop production prediction as shown in Fig. 21.

4 Proposed Work

The goal of the proposed system is to help farmers grow crops to increase production. The selected crop is maize, millets, oil seeds, paddy, pulses, sugarcane, tobacco, and wheat. In this proposed methodology, I have proposed random forest algorithm to

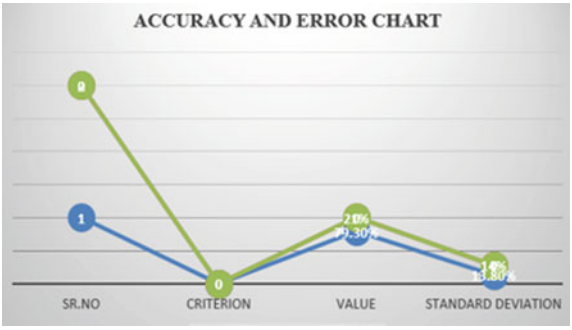


Fig. 19 Accuracy and error of DT

Confusion Matrix

	true Sandy	true Loamy	true Black	true Red	true Clayey	class precision
pred. Sandy	6	0	1	0	0	85.71%
pred. Loamy	0	3	0	0	0	100.00%
pred. Black	0	3	6	2	0	54.55%
pred. Red	0	0	0	4	0	100.00%
pred. Clayey	0	0	0	0	4	100.00%
class recall	100.00%	50.00%	85.71%	66.67%	100.00%	

Fig. 20 Confusion matrix of DT

Row ..	Soil Type	prediction(S..	confidence(..	confidence(..	confidence(..	confidence(..	confidence(..	Prize..	cost	Crop Type	fertilizer	Temperat..	Humid..	Moisture
1	Red	Red	0.091	0	0.166	0.870	0.833	0	0.625	Tobacco	28-28	32	62	34
2	Sandy	Sandy	0.071	0	0.100	0.000	0.833	10	0.625	Barley	17-17-17	26	52	35
3	Loamy	Loamy	0.962	0.010	0.028	0.000	0.833	0	0.625	Wheat	Urea	33	64	50
4	Clayey	Clayey	0.091	0	0.166	0.000	0.870	0	0.625	Pulses	DAP	27	54	28
5	Clayey	Clayey	0.091	0	0.166	0.000	0.870	0	0.625	Paddy	28-28	32	62	41
6	Clayey	Clayey	0.962	0	0.028	0.000	0.870	0	0.625	Pulses	Urea	33	64	34
7	Sandy	Sandy	0.071	0	0.100	0.000	0.833	0	0.625	Barley	28-28	25	58	30
8	Loamy	Black	0.091	0.090	0.000	0.000	0.833	14	0.000	Sugarcane	17-17-17	34	65	53
9	Red	Red	0.091	0	0.166	0.870	0.833	0	0.625	Tobacco	DAP	28	58	33
10	Black	Black	0.091	0	0.000	0.000	0.833	0	0.000	Oil seeds	DAP	37	70	32
11	Sandy	Sandy	0.071	0	0.100	0.000	0.833	9	0.625	Barley	14-35-14	26	56	36
12	Black	Sandy	0.719	0	0.209	0.000	0.833	0	0.421	Millets	Urea	28	54	35
13	Loamy	Black	0.091	0	0.054	0.135	0.833	10	0.727	Cotton	14-35-14	30	68	61

Fig. 21 Crop production prediction of DT

Fig. 22 Production model of DT

predict better crop production. In this model, we have applied data set of crop yield from Kaggle websites, and it will be observed that 80% accuracy along with less classification error 20% trained on rapid miner tool of above models. There are three steps in proposed work.

- Soil Classification: Soil nutrient data can be used for soil classification. The five machine learning algorithms used for soil classification are NB, GLM, DL, DT, and RF. These algorithms will classify and display performance, confusion matrix, production model, accuracy and error prediction chart, prediction, and at the end accuracy in percentage as output.
- Crop Yield Prediction**:** The prediction of crop yields can be done with crop yield data, nutrient information, and location. These inputs are transferred to machine algorithms like NB, GLM, DL, DT, and RF. These algorithms predict crops based on their current input.
- Fertilizer Recommendation**:** The recommendation for fertilizers can be made with data of fertilizer, crop figures, and location. We recommend appropriate crops and fertilizers for each crop in this section.

4.1 Random Forest (RF)

Random forest is a supervised machine teaching technique based on learning. Learning together is a learning way of building a more effective predictive model in many versions of the same algorithm. Several similar techniques are including the random forest algorithm. For classification and regression issues, the random forest approach might be employed. We applied data set into RF model, and it will observe that accuracy is 80.00% and classification error is 20.00% as shown in Fig. 23. The random forest is highly suggested to red soil is 1%, loamy soil is 16%, black soil is 83% for oil seed or red soil is 1%, loamy soil is 76% and black soil is 23% for oil sugarcane and also red soil is 1%, loamy soil is 78% and black soil is 21% to crop production as shown in Fig. 24 along with confusion matrix and performance and error prediction as shown in Figs. 25 and 28. The production of random forest model and prediction of crop production as shown in Figs. 26 and 27. Pseudocode of random forest and process flow of crop yield prediction are shown in Fig. 29.

Advantages of random forest algorithm:

- As several trees exist and a subdivision of data is used to train each tree, the random forest technology is impartial.
- If new data points are introduced into the data set, the random forest algorithm is stable, and the overall algorithm is not affected.

To generate c classifiers:
for i = 1 to c **do**
  Randomly sample the training data D with replacement to produce D_i
  Create a root node, N_i containing D_i
  Call BuildTree(N_i)
end for

BuildTree(N):
if N contains instances of only one class **then**
  return
else
  Randomly select x% of the possible splitting features in N
  Select the feature F with the highest information gain to split on
  Create f child nodes of N, N_1,..., N_f, where F has f possible values (F_1, ... , F_f)
  for i = 1 to f **do**
    Set the contents of N_i to D_i, where D_i is all instances in N that match F_i
    Call BuildTree(N_i)
  end for
end if

Fig. 23 Performance of RF

Random Forest - Performance

Performances

Criterion	Value	Standard Deviation
Accuracy	80.0%	± 7.5%
Classification Error	20.0%	± 7.5%

Fig. 24 Production model of DT

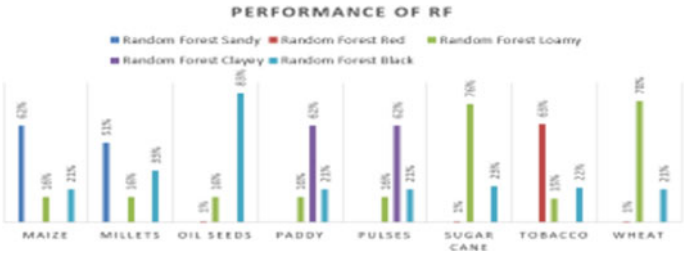


Fig. 25 Production model of DT

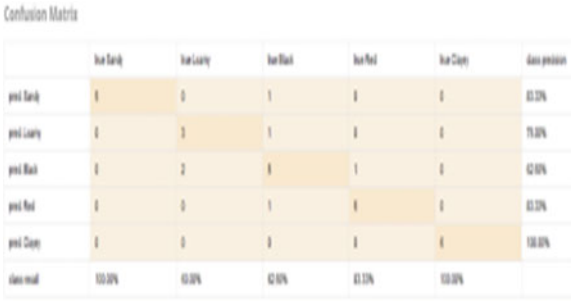


Fig. 26 Production model of DT

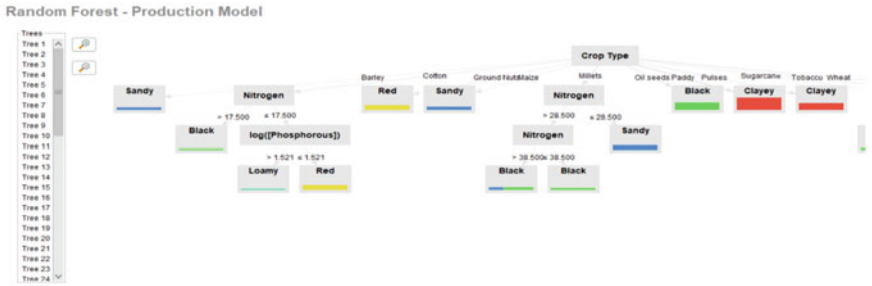


Fig. 27 Production model of DT

Fig. 28 Production model of DT

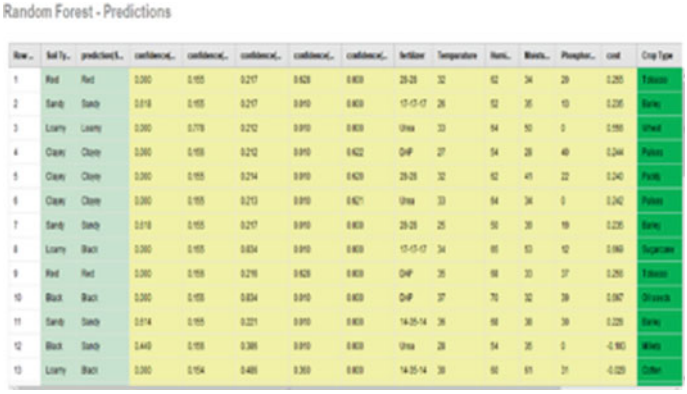


Fig. 29 Pseudocode of RF

5 Experimental Analysis

The complete system is designed using machine learning auto models of rapid miner tools. Crop yield data sets, soil and crop nutrients, fertilizer data sets, and other

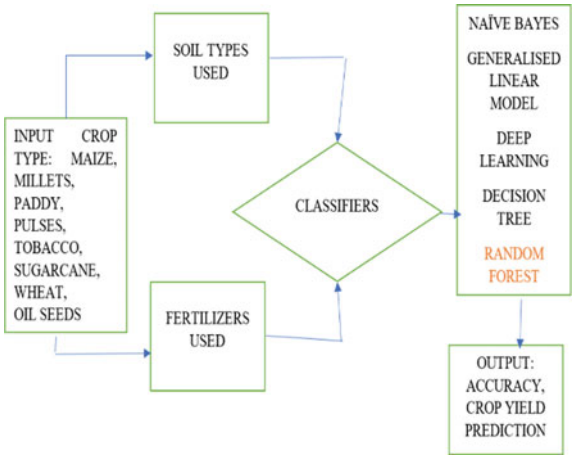


Fig. 30 Comparison of models

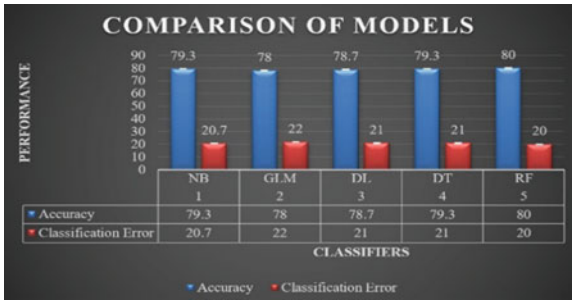


Fig. 31 Working flow of models

data sets come from the Kaggle website. During our practical work and compare our above models, it will observe that random forest algorithm has been achieved 80% of accuracy and having less error is 20% as shown in Fig. 30. The flow process of these models is shown in Fig. 31. By applying above all machine learning algorithms, it will predict crop type along with soil type prediction as given in Table 3. By evaluating of models, we observed the training and scoring time of models as shown in Fig. 32.

6 Conclusion and Future Work

Predicting crop production using Kaggle websites and properly implementing algorithms for automotive rapid mining have shown the ability to produce higher crops. We have concluded from above that random forest is good with accuracy for soil classification and crop classification of approximately 80%, and the error classification rate in comparison with NB, DT, DL, and GLM is 20%. For crop production

Table 3 Simulation of models

SI. No	Crop type/classifiers	Maize (%)	Millets (%)	Oil seeds (%)	Paddy (%)	Pulses (%)	Sugarcane (%)	Tobacco (%)	Wheat (%)
1	NB	Sandy = 76 B1ack = 24	Sandy = 55 Black = 45	Sandy = 11 B1ack = 89	Sandy = 11 Clayey = 65 B1ack = 24	Sandy = 11% Clayey = 65% B1ack = 24%	Sandy = 11 Black = 89	Sandy = 11 Red = 65 B1ack = 24	Sandy = 11 Loam = 65 B1ack = 24
2	GLM	Sandy = 69 Black = 31	Sandy = 62 B1ack = 38	Sandy = 12 B1ack = 88	Sandy = 12 Clayey = 59 B1ack = 29	Sandy = 12% Clayey = 60% B1ack = 28%	Sandy = 12 Black = 88	Sandy = 12 Red = 57 Black = 31	Sandy = 12 Loam = 59 Black = 29
3	DL	Sandy = 61 Loamy = 14 B1ack = 26	Sandy = 54 Loamy = 14 B1ack = 32	Sandy = 10 Loamy = 16 B1ack = 74	Sandy = 10 Loamy = 14 Clayey = 55 Black = 21	Sandy = 10% Loamy = 16% Clayey = 53% B1ack = 22%	Sandy = 9 Loifilly = 49 Black = 42	Sandy = 9 Red = 51 Loamy = 15 B1ack = 25	Sandy = 10 Loamy = 41 Black = 49
4	DT	Sandy = 81 Black = 19	Sandy = 55 B1ack = 45	Black = 100	Clayey = 81 Black = 19	Clayey = 81% Black = 19%	Black = 100	Red = 81 Black = 19	Loamy = 81 Black = 19
5	RF	Sandy = 62 Loamy = 17 Black = 21	Sandy = 51 Loamy = 16 B1ack = 33	Red = 1 Loamy = 16 B1ack = 83	Loamy = 16 Clayey = 62 Black = 21	Loamy = 16% Clayey = 62% Black = 21%	Red = 1 Loamy = 76 B1ack = 23	Red = 63 Loamy = 15 B1ack = 22	Red = 1 Loamy = 78 Black = 21

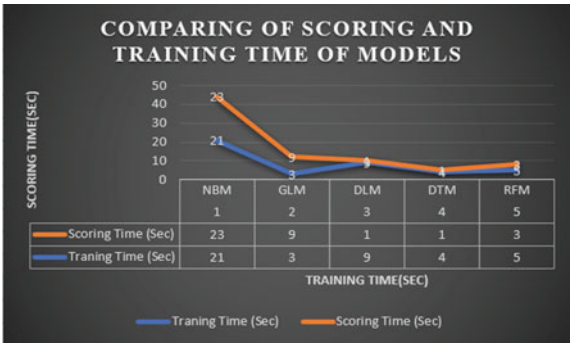


Fig. 32 Time comparison

prediction, it will observe that the random forest algorithm is achieved better performance as compared to above machine learning models. This work can be further expanded to include the following. A mobile application can be created for farmers by uploading pictures from the farm. Detection of crop diseases by image processing where users obtain pesticides based on images of disease. To achieve higher yields, implement smart irrigation systems.

References

1. Ghosh S, Koley S (2014) Machine learning for soil fertility and plant nutrient management using back propagation neural networks. Int J Recent Innov Trends Comput Commun 2(2):292–297
2. Bondre DA, Mahagaonkar S (2019) Prediction of crop yield and fertilizer recommendation using machine learning algorithms. Int J Eng Appl Sci Technol 4(5):371–376
3. Paul M, Vishwakarma SK, Verma A (2015) Analysis of soil behaviour and prediction of crop yield using data mining approach. In: 2015 International conference on computational intelligence and communication networks (CICN). IEEE, pp 766–771
4. Manjula E, Djodiltachoumy S (2017) Data mining technique to analyze soil nutrients based on hybrid classification. Int J Adv Res Comput Sci 8(8)
5. Rajak RK, Pawar A, Pendke M, Shinde P, Rathod S, Devare A (2017) Crop recommendation system to maximize crop yield using machine learning technique. Int Res J Eng Technol 4(12):950–953
6. Veenadhari S, Misra B, Singh CD (2014) Machine learning approach for forecasting crop yield based on climatic parameters. In: 2014 International conference on computer communication and informatics. IEEE, pp 1–5
7. Afrin S, Khan AT, Mahia M, Ahsan R, Mishal MR, Ahmed W, Rahman RM (2018) Analysis of soil properties and climatic data to predict crop yields and cluster different agricultural regions of Bangladesh. In: 2018 IEEE/ACIS 17th International conference on computer and information science (ICIS). IEEE, pp 80–85
8. Nelder N, Wedderburn R (1972) Generalized linear models. J Roy Stat Soc Ser A (General) 135(3):370–384. JSTOR 2344614. S2CID 14154576. https://doi.org/10.2307/2344614
9. Bengio Y (2009) Learning deep architectures for AI. In: Foundations and trends in machine learning, vol 2, no 1. https://doi.org/10.1561/2200000006

Cognitive Load Classification in Alcoholic and Control Subjects Performing Delayed Matching Task Using Optimizable Classifier Networks

Sweeti

Abstract Alcohol abuse is a major problem that has short-term as well as long-term effects on the cognitive demand of the individual. Electroencephalogram (EEG) signals acquired from the scalp are capable of analyzing these effects with the help of cognitive tasks assigned to the subject. It has been found in our previous study that when alcoholic and control subjects are given cognitive delayed matching task, they show the difference in feature distribution over the scalp for beta frequency. Also, subjects show a significant difference in the non-matching task condition which is a higher cognitive demand condition as compared to the other two conditions. These results are further explored here to find measures to classify the alcoholic and control subjects using optimizable classifiers. Results show classification accuracy of 94.8%, 96.9%, and 95% for the single object, non-matching, and matching task conditions, respectively. Results suggest that the cognitive demand of individuals gets affected by long-term alcohol abuse.

Keywords Alcohol · Delayed matching task · Entropy · Classification

1 Introduction

Alcohol abuse is a serious problem and if we look at the Indian statistics, The National Mental Health Survey of India found a 9% prevalence of alcohol use disorder (AUD) in grown men in the year 2015–16. There is a 5.4% alcohol-attributable fraction (AAF) of all-cause deaths and approximately 62.9% of all the deaths due to liver cirrhosis were because of alcohol use [1]. It is observed that alcohol abuse can affect memory and motor control when taken in high doses for a very long span. Studies have shown the effect of alcohol abuse on the cognitive problem-solving approach as well [2]. Electroencephalogram (EEG) signals are capable of analyzing this effect with the help of cognitive tasks assigned to the subject. Literature suggests a higher

Sweeti (✉)
Medical Electronics Engineering, M S Ramaiah Institute of Technology, Bangalore, India
e-mail: sweeti.bme@gmail.com; sweeti.bme@msrit.edu

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_22

entropy value in the alcoholics which shows more randomness in data [3]. Signals-based source localization shows more beta activity in control subjects than alcoholic subjects in the middle frontal gyrus and superior occipital gyrus with the given cognitive task [4, 5]. More beta activity shows the activation of attentional networks with a given activity. Other than these features, higher complexity and lower energy are also observed in the alcoholics [6].

The literature suggests various methods of EEG signal classification for different applications like brain–computer interface, including K-nearest neighbor, linear discriminant analysis, logistic regression, radial basis function, support vector machine, and shallow and deep artificial neural network [7–10]. These methods utilize features extracted from some of the common feature extraction methods that involve autoregression models, band power difference, power spectral density, standard deviation, variance, and other time–frequency domain features [11, 12]. Other studies in this area performing classification of the alcoholic and control subjects suggested an accuracy of 70% and a selectivity of 67.2% with the power of the theta frequency band [13]. SVM classifier with a polynomial kernel suggests 91.7% accuracy [2]. Classification accuracy of 92.5% and 90.3% is reported using SVM and KNN over 13 EEG channels [14].

In the present work, the focus is on the high-frequency bands, i.e., beta frequency band between 16 and 32 Hz. This work aims to detect alcoholism in human subjects using optimizable classification approaches.

2 Materials and Methods

EEG data is acquired from the alcoholic and control subjects performing the delayed matching task. It is processed, and features are extracted to classify the alcoholic and control subjects. Figure 1 shows the methodology followed.

2.1 Task and Dataset Description

The task presents the pictures of the objects taken from 1980 Snodgrass and Vanderwart picture set. These pictures are the black and white line drawings standardized on four variables of central relevance including name agreement, image agreement, familiarity, and visual complexity. Task provides the three task conditions. These are single object, matched, and non-matched conditions. In single object condition, subjects were exposed to the single picture stimulus. In matched condition, subjects were exposed to two stimuli where stimulus 1 was identical to stimulus 2, while in non-matched condition stimulus 1 differs the stimuli 2. Stimuli were presented at the center of the monitor screen, and subjects were instructed to keep the center fixation and respond according to the three task conditions. This task is used to find the EEG correlates of different cognitive loads in the center visual field. The same dataset

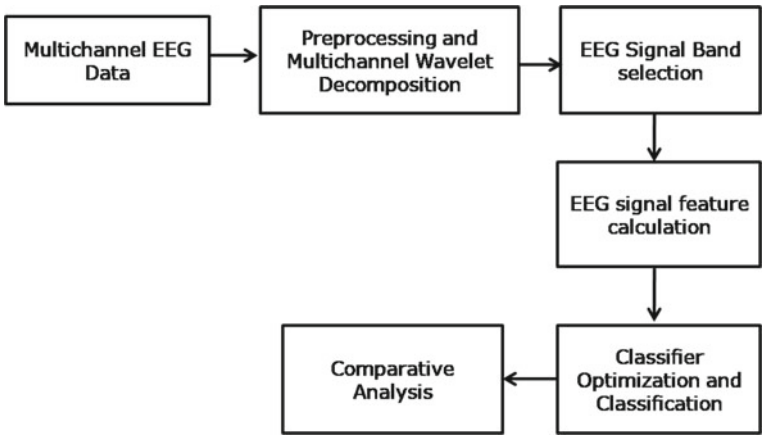


Fig. 1 Block diagram

was earlier tested to find the correlates for the difference in normal and alcoholic subjects.

The EEG data of alcoholic and control subjects is taken from the online repository of UCL. It is recorded from 64 electrodes with a sampling frequency of 256 Hz. The electrode placement is shown in Fig. 2. It consists of data of the subjects performing delayed matching task in three task conditions. These three task conditions as discussed previously provide different cognitive loads to the subjects. The data is divided into three datasets corresponding to the single object, match, and non-match task conditions for both classes. Table 1 gives the dataset information.

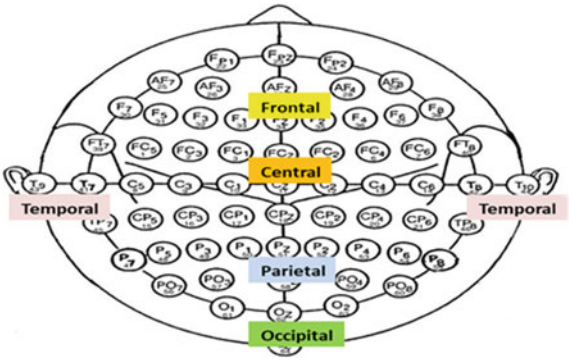


Fig. 2 Surface electrode placement

Table 1 Dataset information for each subject

Dataset			
	A	B	C
Task condition	Single object	Match	Non-match
Eye fixation	Center	Center	Center
Electrode type	Surface	Surface	Surface
Electrode placement	International 10–20 placement system	International 10–20 placement system	International 10–20 placement system
Sampling rate	256 Hz	256 Hz	256 Hz
Number of subjects	10	10	10
Number of electrodes	64	64	64

2.2 Preprocessing and Feature Calculation

EEG data is preprocessed and filtered in 0.1–40 Hz frequency range. Wavelet and PCA-based de-noising is performed to get a clear and useful signal to get good results. Data is decomposed to the frequency domain using db2 wavelet function. The sampling frequency of the EEG data is 256 Hz; the data is decomposed to five levels to get the desired frequency bands. Approximations at level 5 give the delta, details at level 5 give the theta, details at level 4 give the alpha, and details at level 3 give the beta band. The focus of the study is on the high-frequency band to compare the activity level in the two categories of subjects. Features are calculated for the beta band that is obtained at the third level of decomposition at D3. Entropy features are explored to get the significant classification accuracy, which represents the uncertainty, randomness in data, and is also sometimes related to consciousness.

The analysis is carried out with the entropy features calculated using the formula:

$$\text{Entropy} = -\sum pi \log_2(pi) \tag{1}$$

where pi = Probability Density Function

Figure 3 shows the entropy distribution over the 64 electrodes for alcoholic and control subjects. This distribution suggests that data uncertainty increases in alcoholic subjects as compared to the control subjects when subjected to higher cognition load [3].

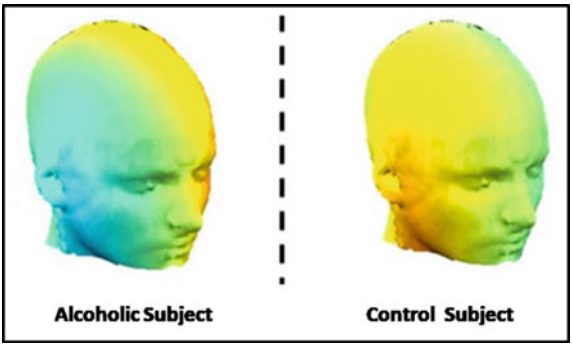


Fig. 3 Entropy distribution over the 64 electrodes for alcoholic and control subject for non-match task condition

2.3 Classification

Three optimizable supervised classifiers are selected and compared to get the best classification results following the Bayesian optimization technique. The optimization algorithm tries to find the tuned hyperparameters based on the acquisition function as follows:

- It will first build a probability model of the objective function.
- The hyperparameters are then tuned for this model and the objective function is evaluated.
- The probability model is now updated with the new hyperparameters.
- The process is repeated until the maximum number of iterations is reached.

(1) K-nearest neighbor (KNN) classifier: It works on the basis of the similarity function between the nearest neighbors. The classifier assigns weights according to the contribution of the neighbor also considering the distance to the neighbor. There is no clear training step for K, but neighbors whose class is already known from the training dataset are selected. It makes the decision based on the majority vote given by selected nearest neighbors. The hyperparameters like the number of neighbors, distance metric, and distance weight can be optimized.

(2) Support vector machine (SVM): SVM tries to find a hyperplane between the two classes. The optimization approach will try to tune the hyperparameters so that the distance between two support vectors can be maximized. The hyperparameters like kernel function and kernel scale can be tuned to get the best classification result.

(3) Discriminant analysis (DA): It is a covariance matrix-based method like logistic regression. This method is generally used for data or dimensionality reduction. It looks for the combination of variables to get the best classification score so that predictions can be made for two classes. The optimized classifier selects the quadratic hyperparameters. These classification algorithms are implemented with MATLAB 2020.

Table 2 Comparison of classifier accuracy (%) for three task conditions

Task conditions	Optimizable classifiers accuracy (%)		
	KNN	SVM	Discriminant analysis
Single object	92.5	**94.8**	93.4
Match	90	96.3	**96.9**
Non-match	89.9	93.3	**95**

3 Results

This section discusses the classification results obtained using the entropy feature to classify the cognitive load in alcoholic and healthy subjects for three different task conditions. Table 2 gives a comparison of the classification accuracy of three classifiers for single object, match, and non-match task conditions.

S. C. et al. reported an accuracy of 91.7% using SVM classifier with a polynomial kernel [2]. W. Mumtaz et al. showed that integration of different EEG band powers can be used to classify the alcoholic and control subjects. They reported an accuracy of 89.3% based on EEG band powers [15]. S. A. Abdulrahman suggested classification accuracy of 92.5% and 90.3% is reported using SVM and KNN over 13 EEG channels [14]. In our present study, we are able to achieve the highest classification accuracy of 94.8% in single object task condition over 64 electrode channels using SVM. DA gives the highest classification accuracy of 96.9% and 95% for the match and non-match task conditions, respectively. Thus, we can classify the alcoholic and control subjects with a maximum classification accuracy of 96.9%.

Table 3 presents the summary of the classification performance. It gives the sensitivity (true positive rate), specificity (true negative rate), accuracy, precision (positive predictive value), AUC (area under the receiver operating characteristic curve), and minimum classification error. The classifiers can classify the alcoholic and control subjects over different task conditions or different cognitive loads. The single object and non-match task conditions classify the two classes with a maximum classification accuracy of 94.8% and 95%, respectively. The match task condition gives the highest classification accuracy of 96.9%, sensitivity of 98.4%, and specificity of 95.4% with a minimum classification error of 0.031 over 16 iterations.

Figure 4 gives the confusion matrix and AUC for the three best-performing classifiers for three task conditions.

4 Conclusion

This paper discusses the classification among two categories of subjects when subjected to high cognitive demand in the delayed non-matching task. The prime motive of the work is to classify the alcoholics from the non-alcoholics. We can

Table 3 Summary of best classification performance

Parameters	Task conditions		
	Single object	Match	Non-match
Sensitivity	94.8%	98.4%	96.6%
Specificity	94.8%	95.4%	93%
Accuracy	**94.8**%	**96.9**%	**95**%
Precision	94.8%	95.3%	93.3%
AUC	0.99	0.98	0.97
Minimum classification error	0.05	0.031	0.05

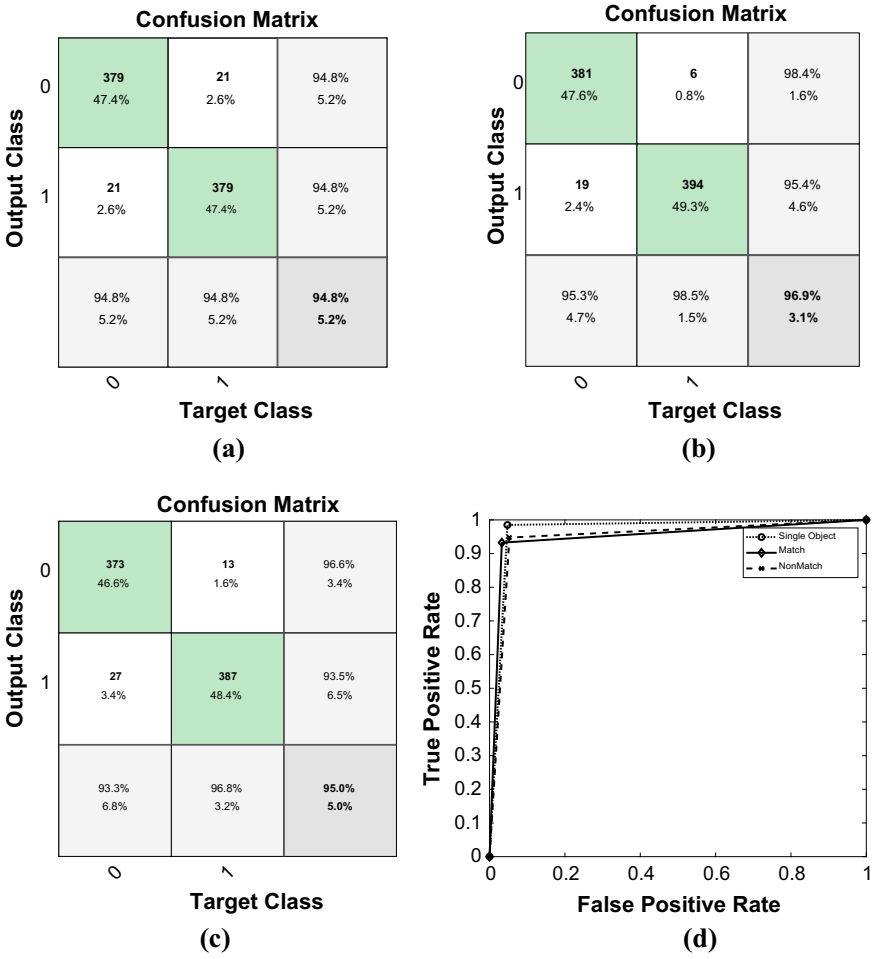


Fig. 4 **a**–**c** Confusion matrices for single object, match, and non-match task conditions, respectively; **d** ROC curve

achieve the highest classification accuracy of 96.9% using optimizable discriminant analysis for match task condition. Based on the high accuracy obtained, we can conclude that long-term alcohol abuse can highly affect cognitive processing capabilities.

In the future, the classifier can be trained with data from up to 100 such subjects to increase the robustness of the model. Also, the data from different sources can further add to the effectiveness of the system.

Acknowledgements We would like to acknowledge Henri Begleiter at the Neurodynamics Laboratory at the State University of New York Health Center at Brooklyn for sharing the dataset.

References

1. Eashwar VA, Umadevi R, Gopalakrishnan S (2020) Alcohol consumption in India—an epidemiological review. J Fam Med Primary Care 9(1):49
2. Acharya UR, Sree SV, Chattopadhyay S, Suri JS (2012) Automated diagnosis of normal and alcoholic EEG signals. Int J Neural Syst 22(3):1–11
3. Panigrahi BK, Santhosh J, Anand S, Sharma R (2015) Entropy feature distribution in delay-matching task for alcoholic and control subjects. Int Conf Comput Commun Autom (ICCCA2015):968–973
4. Sweeti, Singh N, Godiyal AK, Panigrahi BK, Anand S, Santhosh J (2015) Source localization in alcoholic and control subjects to estimate cognitive load using EEG signal. IEEE Int Conf Comput Commun Control, IC4 2015:1–5
5. Sun Y, Ye N, Xu X (2006) EEG analysis of alcoholics and controls based on feature extraction. ICSP2006 Proceedings, pp 6–9
6. Systems E, Park W, Kingdom U, Analysis E (2007) On the complexity and energy analyses in EEG. Int J Comput Intell Appl 2(48):1–15
7. Jia H (2011) Neural network in the application of EEG signal classification method. In: Proceedings of 2011 7th international conference on computational intelligence security CIS 2011, pp 1325–1327
8. Subasi A, Erçelebi E (2005) Classification of EEG signals using neural network and logistic regression. Comput Meth Programs Biomed 78(2):87–99
9. Zhang H, Silva FHS, Ohata EF, Medeiros AG, Rebouças Filho PP (2020) Bi-dimensional approach based on transfer learning for alcoholism pre-disposition classification via EEG signals. Front Hum Neurosci 14:1–13
10. Fayyaz A, Maqbool M, Saeed M (2019) Classifying alcoholics and control patients using deep learning and peak visualization method. ACM Int Conf Proc Ser:1–6
11. Hosni SM, Gadallah ME, Bahgat SF, Abdelwahab MS (2007) Classification of EEG signals using different feature extraction techniques for mental-task BCI. In: 2007 international conference on computer engineering & systems-2007, pp 220–226
12. Murugappan M (2010) Classification of human emotion from EEG using discrete wavelet transform. J Biomed Sci Eng 03(04):390–396
13. Guntaka R, Tcheslavski GV (2013) On the EEG-based automated detection of alcohol dependence. Int J Bioautomation 17(3):167–176
14. Abdulrahman SA (2019) Comparison between alcoholic and control subjects in EEG signals using classification methods. J Mech Contin Math Sci 14(4):93–101
15. Mumtaz W, Vuong PL, Xia L, Malik AS, Rashid RBA (2017) An EEG-based machine learning method to screen alcohol use disorder. Cogn Neurodyn 11(2):161–171

Energy Detection in Cognitive Radio Networks to Optimize the Spectrum Utilization Through Cooperative Spectrum Sensing Using Weights

Rahul Gupta and P. C. Gupta

Abstract Cognitive radio network provides an efficient solution for spectrum utilization. The cooperative spectrum sensing is one of the sensing techniques with which the secondary users of cognitive radio network (CRN) could learn about the availability of the spectrum. In this article, authors had designed and developed their own CRN simulation model in which the RF spectrum range of 54 MHz to 648 MHz is divided into various channels. Secondary users are kept moving in these various channels along with the primary user. The presence of primary user is detected by applying weights on the energy received by local secondary users. The information about the energy of primary user is collected locally by SU and provided to the fusion center for PU detection. This article discusses how the secondary user detects the energy of primary user using weights.

Keywords Cognitive radio networks (CRN) · Primary users (PU) Secondary users (SU) · Dynamic spectrum sensing (DSA) · Radio frequency (RF) · Cooperative spectrum sensing (CSS) and Fusion center (FC)

1 Introduction

Cognitive radio networks [1, 2] emerged as a potential solution for optimizing the wireless spread spectrum because of their ability to cognitively learn and sense the availability of the spectrum. The major advantage of cognitive radio networks is that they exist with current RF spectrum like 3G, 4G and are also compatible with upcoming 5G and 6G spectrum technologies. The major research requirement in the field of cognitive radio networks [3] is to understand the network protocols, architecture, devices and the communication among them. The major challenge in front of researchers is to enhance the learning and sensing capability of secondary users of

R. Gupta (✉) · P. C. Gupta
Department of Computer Science and Informatics, University of Kota, Kota, Rajasthan, India
e-mail: rg12kota@gmail.com

P. C. Gupta
e-mail: dr.pcgupta@uok.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_23

CRN. For detection of the occupancy of the spectrum by the primary user, researchers come across with various spectrum sensing techniques [4, 5] and classified them into non-cooperative and cooperative spectrum sensing. The non-cooperative sensing such as spectrum holes detection, match filter detection [6], cyclostationary feature detection [7] and energy detection does not require sharing of information by the SU. In the cooperative sensing techniques [8, 9], the researchers proposed that spectrum detection could be more efficient if secondary users share their spectrum sensing information with each other. The cooperative spectrum sensing has a controlling device called fusion center (FC) for controlling the information sharing of the SU. In cooperative spectrum sensing, major proposals were theoretical discussions [9] and the decision making was around a central controlling device called fusion center (FC). Researchers' further classified cooperative spectrum sensing [10] as centralized cooperative spectrum sensing, distributed cooperative spectrum sensing and relay-assisted spectrum sensing. This article discusses various spectrum sensing techniques for cognitive radio networks with respect to centralized cooperative spectrum sensing. To implement centralized cooperative spectrum sensing, the authors had designed and developed their own CRN simulation model in which SU detects the presence of PU by applying the weights to the energy received from signals of PU. This article contributes in the following manner:

(1) It discusses various aspects of cognitive radio networks.
(2) It discusses various spectrum sensing techniques available for cognitive radio networks.
(3) A simulation model is developed for cooperative spectrum sensing and the primary user is detected by the secondary users when they sense the signals received from the PU, and the weights are applied to energy received from the signals of PU.

This article is divided into various sections: Sect. 2 discusses about various aspects of cognitive radio networks. Section 3 discusses the spectrum sensing techniques along with cooperative spectrum sensing available for cognitive radio networks. Section 4 discusses simulation model, data collection and feature extraction. Section 5 discusses the experiment performed, the results and the graphs obtained after successful execution of the CSS simulation model in MATLAB. Section 6 concludes the article.

2 Cognitive Radio Networks and Its Various Aspects

Cognitive radio networks [3, 11] turned out to be breakthrough research in wireless networks as they provided the intelligence to the devices in the network. Cognitive radio networks consist of two kinds of users (1) primary user which are assigned the license band by the regulating authorities in their areas. (2) Secondary user which identifies the availability of the spectrum, when the primary users are not using it. Researchers found that primary users do not use their license bands frequently,

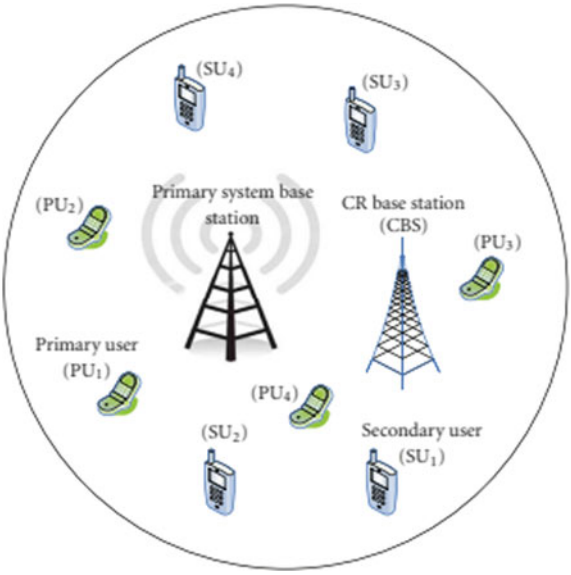


Fig. 1 Cognitive radio network [12]

which remain available for secondary users most of the time. When the primary user wants to occupy its license band it preempts the secondary users. The secondary user applies dynamic spectrum access to detect the availability of spectrum, when the primary user has not occupied it. For dynamic spectrum access (DSA), the secondary users must process three capabilities: Cognitive capability, reconfigurability capability and learning capability. (1) Cognitive capability is the ability of the secondary user to sense the environment and gather the information like transmission frequency, power modulation, bandwidth, etc., to identify the best possible spectrum. (2) Reconfigurability means the ability of the secondary user to rapidly adapt to the learned parameters and get optimum performance. (3) Learning capability means the ability of the secondary users to learn from their past experience, build and enhance a model for efficient decision making (Figs. 1 and 2).

2.1 Cognitive Cycle

The secondary user goes through the cognitive cycle in order to dynamically access the spectrum. The cognitive cycle for the dynamic spectrum access has following phases:

- Spectrum sensing
- Spectrum analysis
- Reasoning
- Adaptation

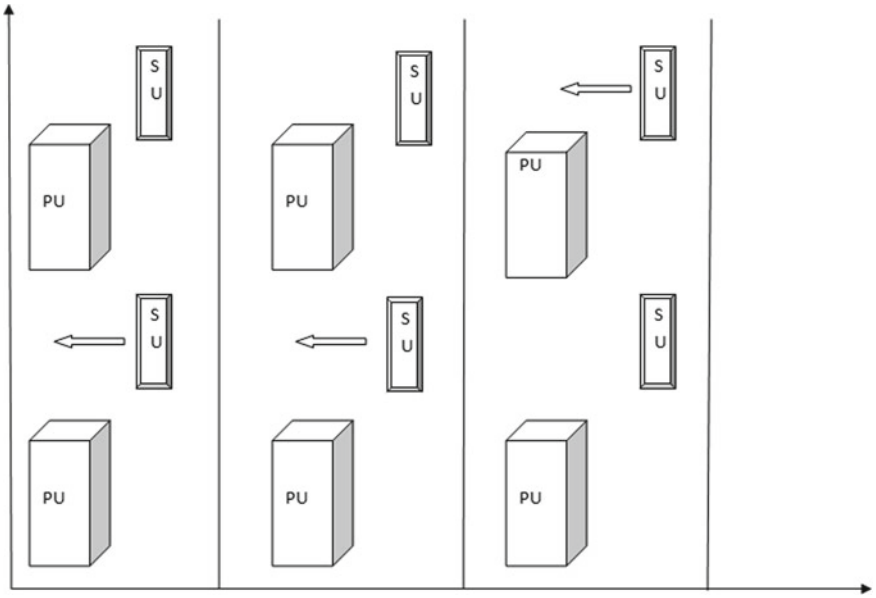


Fig. 2 Cognitive radio channels

Spectrum Sensing: The secondary users must identify when the primary users are not using the spectrum.

Spectrum Analysis: The SU should analyze the spectrum dynamically and collect all relevant information like transmission frequency, bandwidth, power modulation, etc.

Reasoning: The SU should select the best mechanism to share the spectrum.

Adaptation: The SU should modify its parameters according to the sensed environment and get the optimum performance.

3 Spectrum Sensing Techniques

Spectrum sensing plays a vital role in cognitive radio networks as it enables secondary users to obtain information about the environment and availability of the spectrum. The spectrum sensing is classified as non-cooperative spectrum and cooperative spectrum sensing (Fig. 3).

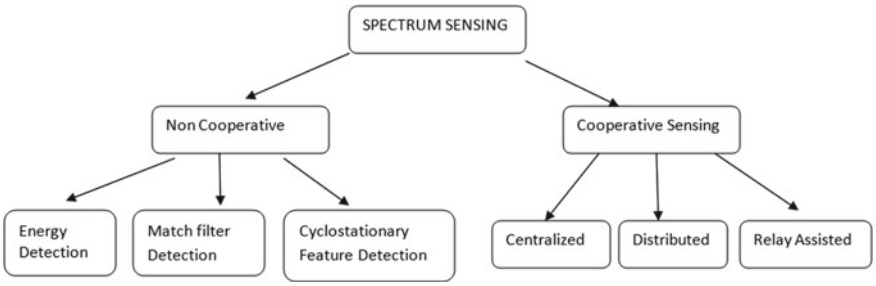


Fig. 3 Spectrum sensing techniques

3.1 Cooperative Spectrum Sensing

In various spectrum sensing techniques, the secondary users receive the signal from the primary users, which can be actual signal combined with noise or noise signal only. Every secondary user senses the presence or absence of primary user on the basis of energy of signals received from them. The secondary user could be near to primary user or far from it. On the basis of the distance from PU in the spectrum, the secondary user receives different signal strength from the PU, which creates doubt in sensing accuracy of secondary user. This limitation created the interest of researchers and they came up with cooperative spectrum sensing technique which is further classified into various categories like centralized, distributed and relay assisted. This article focuses mainly on centralized cooperative spectrum sensing and accordingly simulation model is developed. Cooperative spectrum sensing consists of a fusion center, which is a central station to decide which frequency bandwidth in spectrum is available for detection of primary user. It instructs the local secondary user for energy detection. The fusion center collects the results of energy sensed by local secondary users combines them and decides whether primary user is present in the particular channel of spectrum or not. The fusion center after detecting the presence or absence of primary user sends the information back to cooperating secondary users.

The process of centralized cooperative spectrum sensing is quite simple and is executed in 3 steps local sensing, reporting and data fusion. First, the fusion center selects a particular channel of interest in the spectrum having a certain frequency range and instructs all secondary users to perform local sensing. In the second step, the secondary user senses the energy level of signal from primary user and report the sensing result to fusion center through control channel. In the third step, the fusion center collects the local sensing information from secondary users combines the information, detect whether primary user is present or not and send the result back to all cooperative secondary users.

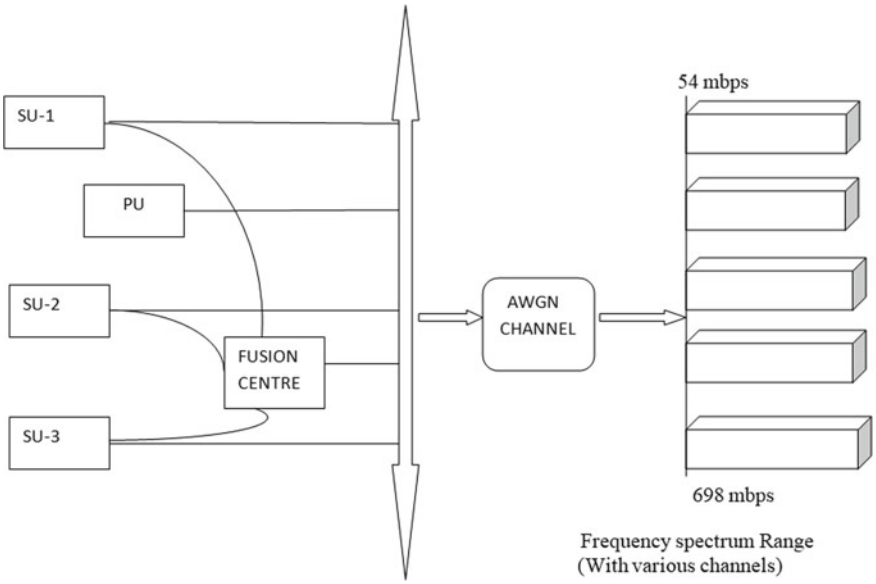


Fig. 4 Simulation model for cooperative spectrum sensing using weights

4 Simulation Model, Data Collection, Feature Extraction

This section discusses the simulation model used; sample data collected and feature extraction.

4.1 Simulation Model

Simulation model had been prepared and coded in MATLAB [13–17] for performing experiments on centralized cooperative spectrum sensing using weights. In this model, one primary user has license band in the frequency range 54 MHz to 698 MHz. As per requirement of user, up to 8 secondary users could be connected in the same frequency band, which are also connected with a fusion center as shown in Fig. 4.

4.2 Data Collection

In a particular trial run, 3 secondary users are connected with a fusion center (FC) and a frequency band. The primary user is connected with a frequency band. The sample data collected during the trial run of the simulation model is shown in various tables.

Table1 Energy of a user at particular channel (CH)

CH	1	2	3	4	5	6	7	8
User1	0.11898	0.24207	0.43420	0.58682	0.45932	1.40328	0.53179	0.49136
User2	0.07457	0.37610	0.13325	0.89452	0.80249	0.47479	1.55080	2.54261
User3	1.14803	0.18538	0.13247	0.28607	0.36819	0.84997	3.67353	2.21976

In the sample data, 3 secondary users using 8 different channels received information about energy of PU.

Table 1 gives the energy received by a local secondary user from the signals of primary user at a particular channel. In the sample CRN simulation run, the information of energy on 8 different channels of spectrum from signal of PU is received by local SU. The energy of the signal received will be greater at a particular channel if the SU in that channel is near to PU, else the energy will be less.

Table 2 gives the decision parameter taken by local secondary user when it analyzes the information of the energy received from the signal of PU. The local SU sends the value of energy and the decision parameter calculated by it to the FC for taking the final decision on the presence of PU in that particular channel. When the energy received by a local secondary user at a particular channel is greater than a particular threshold, the SU put the decision value to be 1 else the SU puts the decision value to be -1. The threshold $= 3$ in this sample run.

Table 3 gives the weights calculated by fusion center (FC) on the basis of information sensed by local secondary users on a particular channel. The FC calculates the weights using the equation

$$W_i = \text{energy}(i, j)/\text{avg-energy}(j) \tag{1}$$

where

Table 2 Value of decision parameter when Energy > threshold

Channels	1	2	3	4	5	6	7	8
User1	−1	−1	−1	−1	−1	−1	−1	−1
User2	−1	−1	−1	−1	−1	−1	−1	−1
User3	−1	−1	−1	−1	−1	−1	1	−1

Table 3 Weights calculated by FC for local secondary user at particular channel (CH)

CH	1	2	3	4	5	6	7	8
W1	0.66673	1.44876	3.31209	2.31116	2.01527	3.66817	0.54970	0.64013
W2	0.41789	2.25088	1.01647	3.52301	3.52094	1.24110	1.60302	3.31242
W3	6.43313	1.10950	1.01047	1.12667	1.61547	2.22183	3.79723	2.89181

$$\text{avg-energy} = \sum_{i=0}^{NN} \text{energy}(i, j) \Big/ NN \tag{2}$$

where i is the ith node in jth channel and NN is number of nodes in jth channel.

4.3 Feature Extraction

The above sample data generated from a particular sample run of simulation model shows that when the local secondary user receives energy at a particular channel greater than a particular threshold (threshold = 3), the decision parameter changes its value to 1 else the value remains −1. When the fusion center receives the energy and decision parameters, sensed by the local secondary user, it calculates weights on the local secondary users. The final decision value is calculated on the basis of weights. If the final decision value is greater than 1, the FC informs that PU has been detected by a local secondary user. For example, local secondary user 3 detected that the energy received at channel 7 (=3.673539) is greater than the threshold (=3); therefore, the decision parameter changes to 1 of user 3 at channel 7. The weight calculated by FC for this local secondary user is 3.797231. On the basis of this information, the final decision value calculated by FC is 2.023931. FC uses the function:

$$fd = fb + pd + w(i) * d(i, j) \tag{3}$$

where, fd is the final decision, pd is the previous decision, where $pd = -1.7733$, w is weight on particular user i in channel j, $d(i,j)$ is decision parameter calculated by user i at channel j and fb is final bias calculated on the basis of function:

$$fb = b + \alpha t \tag{4}$$

Here the initial bias is 1, the learning rate $\alpha = 1$ and $t = 1$ is a constant.

The learning rate α is variable [1, 2, 3, etc.], which could be selected as per the requirement.

5 Experiments and Results

In a particular trial run, 3 secondary users are connected with a fusion center (FC) and frequency band, the primary user is connected with frequency band. The PU signals pass through additive white Gaussian noise (AWGN) channel [18] for combining noise to signal of the primary user. The sampling frequency taken for the primary user signal is 1400 MHz, the secondary user signals are generated using the cosine

function [19].

$$x = \cos(2 * pi * 1000 * t). \tag{5}$$

To plot the power spectral density of primary user and secondary user following amplitude modulation function [20, 21] was used:

$$\text{PUS} = \text{ammod}(x, \text{PUF}, \text{Fs}) \tag{6}$$

The weights are applied to the energy received by the local secondary users, and the sum of all the energy is calculated by fusion center along with average weight. Final energy is calculated as a function of average energy and average weight. On the basis of this, the fusion center takes the decision of the presence or absence of PU in that frequency channel.

Figure 5a shows that the primary user is active and has occupied a slot in frequency band of the range around 350–400 MHz. The power spectral density graph shows the power of the PU signal peaking in the specified frequency range.

Figure 5b shows the position of secondary users in the spectrum when they detect the absence of primary user in a particular run. The X and Y coordinates of all the secondary users are shown in the figure.

Figure 5c shows the energy detection graph of the particular run, when SU detects the absence of PU. This graph suggests that the decision parameter for the particular energy level is either 0 or −1 which indicates the absence of energy of primary user in that particular frequency band in the particular run.

The fusion center collects the information of energy sensed by the local secondary users via reporting channel, applies weights on the energy and compares with particular threshold. When the energy is greater than that particular threshold, it detects the presence of primary user in a particular run and informs all the secondary users.

Figure 5d shows the presence of primary user with red X. This graph shows the position of PU and SU when SU detects the presence of the PU.

Figure 5e shows the energy detection graph of the fusion center in that particular run. This graph shows that there were two instances when the energy of PU signal was greater than the particular threshold, which raise the decision parameter to 1, and therefore, the fusion center was able to detect the presence of PU. Thus, it is concluded that by using the weighted approach on information received by local secondary users the detection of presence of primary user by fusion center is quite easy and very effective.

6 Conclusion

The cognitive radio networks had emerged as a technology for optimum utilization of radio frequency spectrum. Various techniques exist for secondary users to sense the environment and detect the presence of primary user like energy detection,

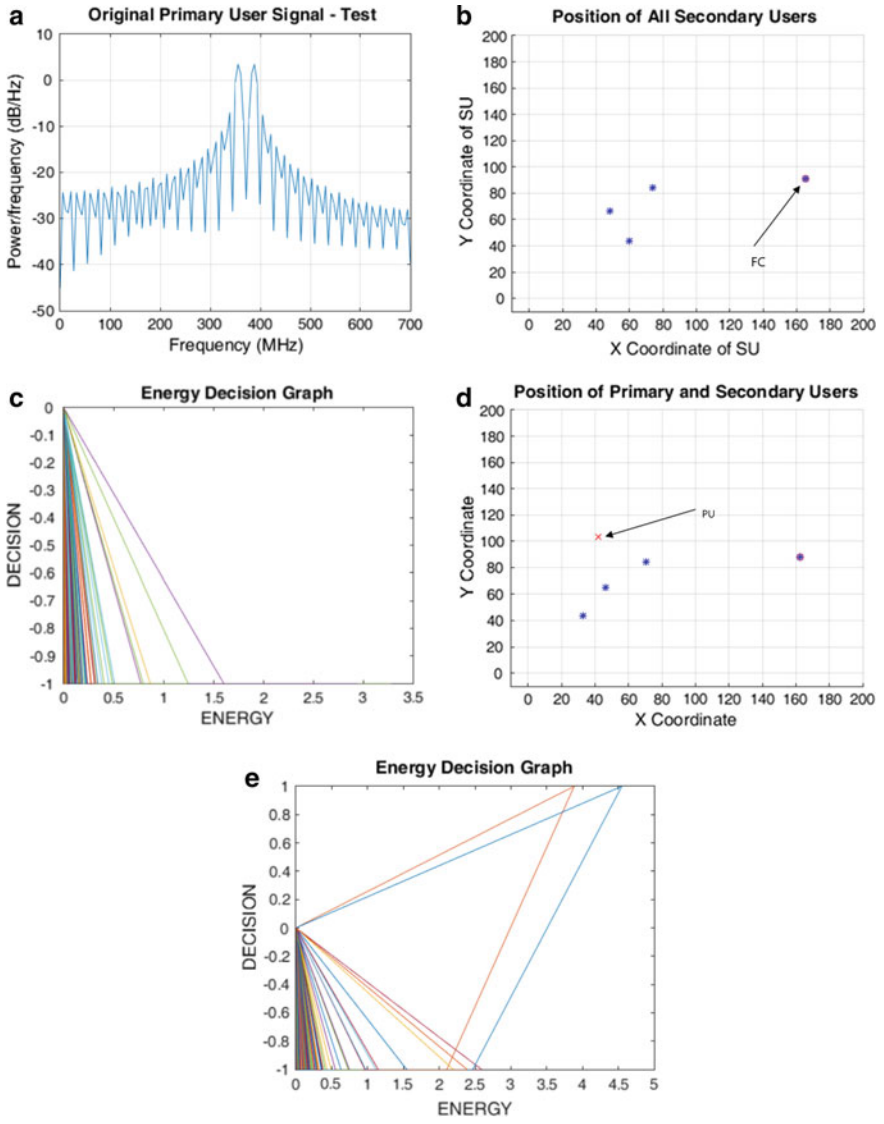


Fig. 5 **a** Slot occupied by primary user in license band, **b** secondary users when detected absence of primary user by fusion center, **c** energy decision graph when SU detected absence of PU, **d** secondary users when detected presence of primary user, **e** energy decision graph when SU detected presence of PU

matched filter detection cooperative spectrum sensing (centralized, distributed and relay assisted). This article discusses the introduction of cognitive radio networks, the various techniques available for primary user detection and develops the simulation model focusing on centralized cooperative spectrum sensing using weight approach. The simulation model was developed and encoded in MATLAB. For optimum results, the weights were applied to the energy detected by local secondary users and the result was forwarded to the fusion center via reporting channel. The fusion center combines the sensing results received and detects the presence or absence of PU in a given frequency range. The results concluded that by applying weights on energy detected by secondary user, better efficiency for primary user detection could be obtained. Various graphs and output generated in MATLAB were presented and discussed in support of the result achieved from the simulation model. In the future, various machine learning techniques will be applied by the authors on the simulation model to achieve a higher level of reliability and optimum results for spectrum utilization by secondary users of cognitive radio networks.

References

1. Nadine A, Youseef N, Karim A (2015) Recent advances on artificial intelligence and learning techniques in cognitive radio networks. EURASIP J Wirele Commun Netw 174
2. Cognitive Radio Networks Wikipedia. https://en.wikipedia.org/wiki/Cognitive_radio
3. Beibei W, Liu KJ (2011) Advances in cognitive radio networks: a survey. IEEE J Sel Top Sign Proces 5:5–23
4. Shewangi, Garg R (2017) Review of cooperative sensing and non-cooperative sensing in cognitive. Int J Eng Technol Sci Res 4(5):229–234
5. Zeng Y, Liang YC, Hoang AT, Zang R (2010) A review on spectrum sensing for cognitive radio: challenges and solutions. EURASIP J Adv Sign Proces
6. kockaya K, Develi I (2020) Spectrum sensing in cognitive radio networks: threshold optimization and analysis. EURASIP J Wirele Commun Networking 255
7. Aparna PS, Jayasheela M (2012) Cyclostationary feature detection in cognitive radio using different modulation schemes. Int J Comput Appl 47(21)
8. Lu Y, Wang D, Fattouche M (2016) Cooperative spectrum-sensing algorithm in cognitive radio by simultaneous sensing and Ber measurements. EURASIP J Wirele Commun Networking 136
9. Akyildiz IF, Brandon F, Balakrishnan R (2010) Cooperative spectrum sensing in cognitive radio networks: a survey. Elsevier, Atlanta G.A. 4:40–62
10. Wang J, Wu Q (2009) Cooperative spectrum sensing. Intech Open Book Series
11. IEEE 802.22 Wikipedia. https://en.wikipedia.org/wiki/IEEE_802.22
12. Shaat M, Bader F (2010) Computationally efficient power allocation algorithm in multicarrier -based cognitive radio networks: OFDM and FBMC systems. EURASIP J Adv Sign Proces
13. Gupta R, Gupta PC (2019) A comparative study of various network simulators available for cognitive radio networks. Vindhya Bharti Res J 1:66–72
14. Matlab Tutorial. https://in.mathworks.com/support/learn-with-matlab-tutorials.html
15. Matlab Tutorial. https://www.tutorialspoint.com/matlab/index.html
16. Dongy Q, Cheny Y, Liy X, Zengz K (2018) A survey on simulation tools and test beds for cognitive radio networks study. arXiv
17. Matrix and Array in Matlab. https://www.javatpoint.com/matrices-and-arrays-in-matlab
18. AWGN Channel Wikipedia. https://en.wikipedia.org/wiki/Additive_white_Gaussian_noise

19. Gupta R, Gupta PC (2019) Cognitive radio network implementation for spectrum utilization in Hadoti (Rajasthan) Region. Vindhya Bharti Res J 2(18):13–18
20. Amplitude Modulation. https://www.mathworks.com/help/comm/ref/ammod.html
21. Amplitude Modulation Wikipedia. https://en.wikipedia.org/wiki/Amplitude_modulation

Real-Time Tracking of Health Parameters Using IOT and Data Logger: Application to COVID-19

Insha Zehra, Rajat Mehrotra, M. A. Ansari, Rajeev Agrawal, and Aruna Pathak

Abstract There is an alarming upward trend in COVID-19 cases, and the existing healthcare system is unable to cater the day-to-day requirements from testing followed by appropriate patient care. The use of monitoring devices has found the importance of detecting the presence of the virus early. In order to expand the exciting monitoring system, every technocrat needed to come forward and donate his money to this. Authors as a responsible persons deemed it proper to work on the development design of multipurpose all in one reliable device to help the front like warriors in identifying persons with COVID-19 symptoms like temperature, heartbeat, humidity and positioning real time. After putting in an extra effort, author has succeeded in translating my idea into action. A device that is so well-designed and will increase the medical brotherhood effort to continue monitoring patient health parameters using a variety of sensors connected to the Arduino board. The generated data is then transmitted via Wi-Fi module to IOT platform, i.e., ThingSpeak and which can be monitored on devices like desktop, laptop or smartphone.

Keywords Internet of Things (IOT) · Arduino board · Wi-Fi module · Sensor · ThingSpeak

I. Zehra (✉) · M. A. Ansari
Department of Electrical Engineering, Gautam Buddha University, Gr. Noida, India
e-mail: inshazehra15@gmail.com

M. A. Ansari
e-mail: ma.ansari@ieee.org

R. Mehrotra · R. Agrawal
Department of Electrical and Electronics Engineering, GL Bajaj Institute of Technology & Management, Gr. Noida, India
e-mail: rajatlmp@gmail.com

R. Agrawal
e-mail: rajkeed@gmail.com

A. Pathak
Department of Electronics and Communication Engineering, Govt. Eng. College, Bharatpur, Rajasthan, India
e-mail: aruna.pathak@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_24

1 Introduction

Adequate availability of attainable health facilities is one of the basic fundamental rights of every human being without caste, color and creed. Governments have a responsibility toward the health of their people which can be achieved by providing appropriate healthcare facilities. In spite of the best possible facilities, COVID-19 has posed challenges even to these countries both at technical and administrative levels. Due to invisible presence of the virus, people are scared to have one to one contact in absences of PPE's, same is true in case of medical fraternity which is also interested to have minimum contact with the people just to avoid risk of infection but there is an imperative need to eliminate these apprehensions by introducing a mechanism which will help both to counter the impact of virus [1]. In order to cope up with this scenario, authors have worked on a device that has a mechanism to avoid direct contact between healthcare workers and persons having COVID-19-like symptoms. The device so designed and developed will enable healthcare workers to know about the health status, viz. COVID-19-like temperature, humidity level, position and heart rate by using different sensors connected to the Arduino board [2], and same will be sent in real-time data to the doctor with the help of IOT/Android application called ThingSpeak, patients from remote/far flung areas can also have a facility to reach to the renowned doctors for sorting out their COVID-19-related issues and early detection and diagnosis of this fatal disease by using this technology. Data logging is the process that includes the analyses of different physical parameters and also keeping them as a record for an increased span of time. The data that has been collected can be accessed and then it can be observed and analyzed even at some other date. We should not get confused between the data acquisition and data logging, while data logging is only a part of data acquisition and it use a lower rate for sampling (seconds, minutes, hours, or days), whereas data acquisition system samples the data at higher speeds (GS/s, MS/s, KS/s) for the real-time value monitoring, analysis and recording.

2 Related Work

This paper gets us through the innovation of technology straightaway from the bottom to the top and still growing. In this paper, our main requirement is to share the data to remote locations and we are enhancing the Data Logger technology in order to make it more convenient and more precise. The data stored in the memory can be viewed in GUI and it can also be analyzed if needed [3]. Various sensors that are used in measuring different physical entities like temperature, humidity, wind speed, overloading, etc., are directly connected to the Arduino Uno. The generic serial LCD is used as the display device. Here, we use IOT, i.e., Internet of Things to find a way into next generation data logging equipment [4], and we will come to know about the data logging fundamentals and how the new devices are being adopted into IOT.

We have made a data logger that is mainly based on components like WI-FI module ESP8266, LCD, Arduino, temperature, humidity, heartbeat and accelerometer sensor. Arduino reads the sensor readings and then it sends this signal to a server connected through the internet with the help of WI-FI module. The Internet server which we are using here is ThingSpeak, an open-source application using IOT that turns out to be an Internet or the Intranet Server. It is used as for the logging of data, for tracking the location and getting the system updates by providing the HTTP API [5].

3 System and Overview

Several sensors, display methodology, IOT and core controller together constitute the system. The sensors of this system estimate pulse rate temperature, humidity and fall detection. This acts as input to core controller which in turn assesses and forwards them to cloud computing through Wi-Fi while parallelly displaying them also [6]. A simple computer or mobile makes available the data to the users through IOT platform. An instant visualization can be created by the users from the device. Various sensors and controllers are used depending on the requisite specifications for the healthcare purpose. The core controller processes analog signals and converts into digital signal for further analysis [7] (Figs. 1 and 2).

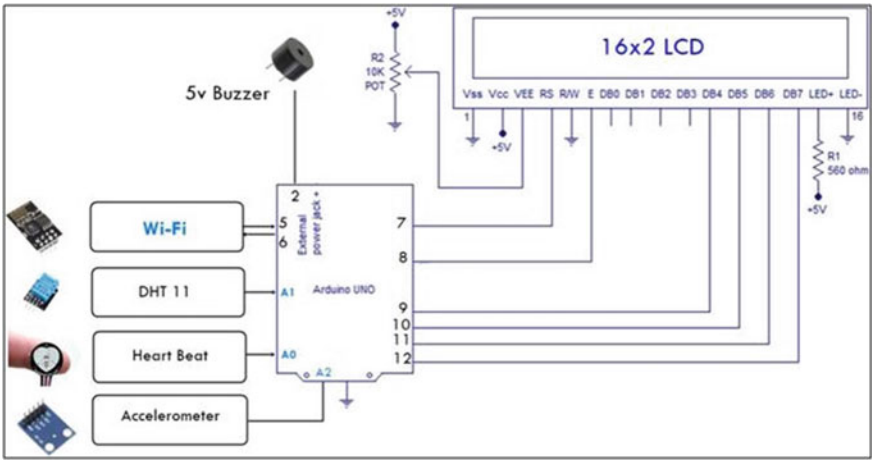


Fig. 1 Circuit diagram of IOT-based healthcare system

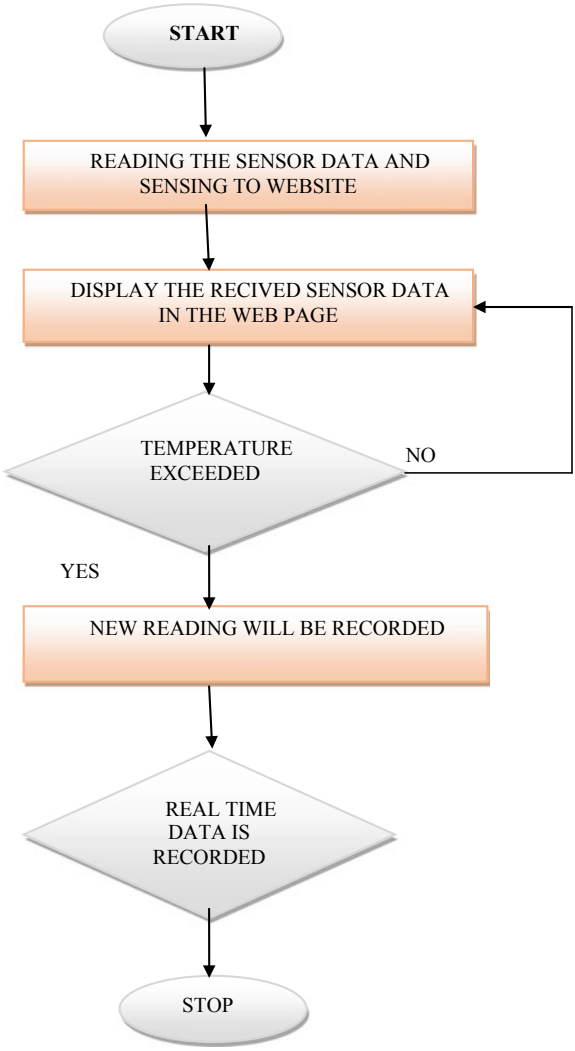


Fig. 2 Flowchart of the data logger using IOT

4 Description of System Components

4.1 DHT11 Sensor

This module depicts a complex of temperature and humidity with a calibrated digital signal output. An accurate value is given by DTH11 that of temperature and humidity with an assurance of long-term stability and greater reliability. A resistive type humidity measurement component is found in the sensor. Further, an 8-bit inbuilt,

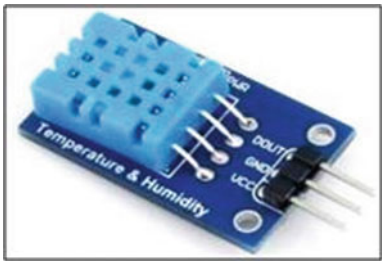


Fig. 3 DHT11 sensor

microcontroller NTC type temperature measurement component that is available in 4-pin single-row package, ensure cost-effective and fast response.

It has a large display and an indication of alarm that makes it efficient to sense the air humidity, temperature. This measurement lasts long by using it continuously and high accurate method for keeping the record of temperature and humidity can be kept in our mobile phones. We can get an instant notification if any changes take place, and this method is most convenient as the growing technology allows us to maintain our pace with it [3] (Fig. 3).

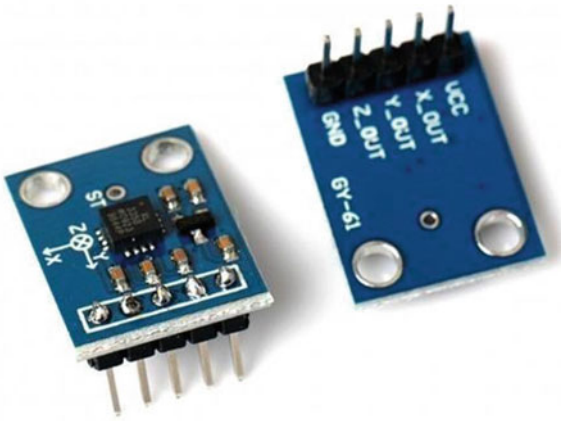
4.2 Heartbeat Sensor

Heartbeat sensor is a device which is used to measure any changes in the amount of blood flowing through a finger that causes a change in light as it applies to the principle of photo plethysmography. This sensor is directly connected to an Arduino that measures heart rate per minute. The heartbeat consists of a detector and a light emitting diode or four image mode as the heartbeat is responsible for the blood circulation to each parts of the body as we place the finger on the sensor, and the light illuminating from finger tissue is transmitted and detected by the sensor. The required light output that has been transmitted to the sensor is measured in the form of an electric signal and is equal to the heart rate. The IC LM358 is used for this sensor. The light absorbed depends on the volume of blood in that tissue [4].

A remote monitoring system using Android applications is used to develop the heartbeat sensor system. The heartbeat sensor is directly connected to the Arduino. The Arduino will convert this transition into the heartbeat per minute. With the help of sensors and bright LED, there is variation in veins in the finger [8]. Arduino UNO collects the data and is displayed on the LCD, thereby saving the doctor's time and improving the quality and responsiveness health services [9] (Fig. 4).

Fig. 4 Heartbeat sensor

Fig. 5 Accelerometer sensor



4.3 Accelerometer Sensor

It is the type of sensor which is used to measure the velocity of the body or acceleration in its own rest frame, by using this sensor, we can monitor the patient's inclination and any other body moment. It works on a principle of piezoelectric, i.e., whenever we apply force on the object, it will produce electric charge that will be proportional to the amount of force applied to it. Accelerometer sensor is used to detect whether a patient is in a state of being (standing or sitting) or lying down (sudden fluctuations in position). This sensor provides digital output. The primary function of accelerometers is tracking physical activity level of a person. It is utilized in health care and mitigates the risk of chronic diseases (Fig. 5).

4.4 IOT in Data Logging

Products of IOT like microcontrollers, sensors and wireless circuits are used to create more effective data loggers. IOT products are cheaper, smaller, consume larger computing power and use less power consumption. Most of the chips now use Bluetooth, WI-FI or Zigbee networks [10]. By the help of the IOT, the whole system is connected to the application made by the user on the phone or any other device,

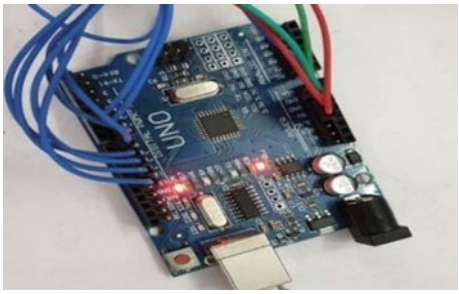


Fig. 6 Arduino Uno

where username and password have to be given to access that application. All the data recorded by IOT would come on the device once logged in [5]. So, this is how it has become so advanced that the memory cards have been replaced by the IOT using WI-FI.

4.5 Arduino

Arduino board consists of digital and analog input/output pins which can be used for multiple combinations or expansion boards. They can be used for other circuits or bread boards. Other features include Universal Serial Bus (USB) on various models used on your computers to download the program. Most Arduino boards include the Atmel AVR 8-bit microcontroller with a number of pins, flash memory. These boards use single-row or sometimes double-sided pins to facilitate installation and installation connections with other circuits. Many and well-placed shoes can be targeted individually with a moving bus. Boards include a 5 V regulator and a 16 MHz oscillator. Arduino's small controls must be pre-configured with a bootloader that simplifies program loading to light up memory. Boards are usually loaded with system code by connecting to another computer. The current boards are arranged not on a moving bus but on a Universal Serial Bus (USB) [2] (Fig. 6).

4.6 Wi-Fi Module

The Wi-Fi module connects a separate SOC coordinating TCP/IP stack and providing access to Wi-Fi organize. A pre-customized Wi-Fi module is available with an AT order set firmware, thereby enabling a connection to Arduino. The Wi-Fi module has a colossal growing network and financial prudence. This module has a pioneering ability to coordinate with sensors and other gadgets to GPIOs with very less stacking. It is on chip reconciliation filters outside hardware, front-end module and PCB region.

5 Software and Hardware Requirement

5.1 ThingSpeak

It is an open-source application enabled by IOT including HTTP API, and it captures the data from the things through the Internet. With the help of this, we can create a network of various things which can be further used for the logging of data, data tracking, etc. The most primary thing we have to do is to open and make an account on ThingSpeak website [2] (Figs. 7 and 8) (Table 1).

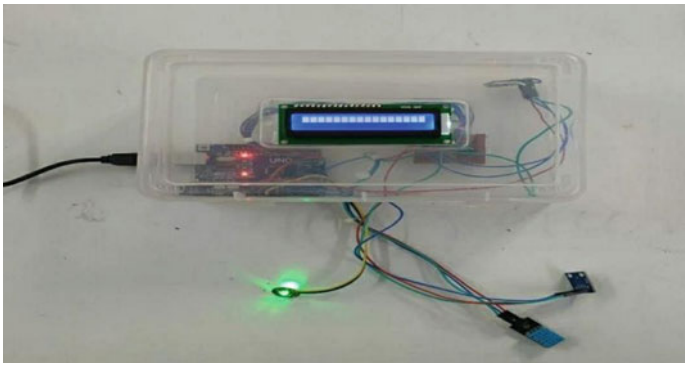


Fig. 7 Hardware components for measuring healthcare parameters

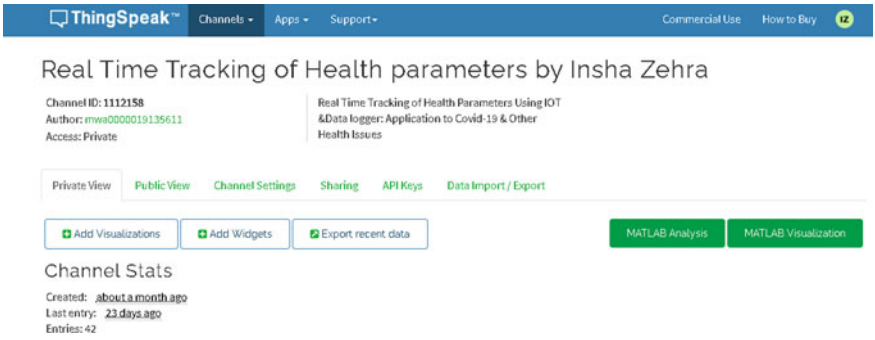


Fig. 8 ThingSpeak channel for sensor data storage

Table 1 Readings of components of the IOT-based smart healthcare system [11]

S. No.	Components	Ratings
1	Arduino UNO	6 analog inputs, 14 digital input/output,16 MHz quartz crystal, 3.3–5 V
2	Wi-Fi module (ESP 8266)	32-bit microcontroller, 80 MHz, 16 GPIO pins, input voltage-3.3–3.6 V
3	DHT11 sensor	Voltage: 3.5–5.5 V, Current: 0.3 mA 60uA, Temperature range: 0–50 °C, Humidity Range: 20–90%, Accuracy: ±1 °C and ±1%
4	Heartbeat sensor	Diameter = 0.625", Operating voltage = 3–5 V, Operating current = ~4 Ma at 5 V
5	Accelerometer sensor	Working voltage:1.8–3.6 V, working current: 350μa, Sensing range: ±3 g (full scale), Temperature range: 40–85 °C
6	LCD	8-bit data pins, supply voltage-5 V

5.2 *Results and Discussion*

(a) Heartbeat sensor is a device which is designed to give us digital output of our heartbeat. Whenever we place a finger on it, the digital output can be directly connected to Arduino such as, to measure the heartbeat per minute (BPM).

(b) Accelerometer sensor is used to measure the motion of a patient. Whenever the force is applied, it will produce an electric charge that will be proportional to the amount of force applied to it. This sensor also provides us digital output which is connected with Arduino to measure current position of the patient (Figs. 9 and 10)

(c) The DHT11 is a device that measures temperature and humidity. The humidity is measured by the electric resistance between two electrodes, and temperature is measured by using thermistor.

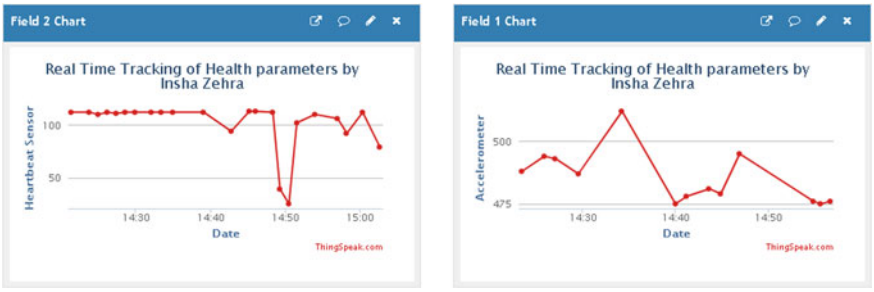


Fig. 9 Measurement by accelerometer and heartbeat sensor

Fig. 10 Measurement by digital temperature and humidity sensor

6 Conclusion

By the help of this system, the doctors or caretakers will be able to diagnose, monitor and advise their patients accordingly, as their real-time data is displayed online on their respective smart phones by using an Android app. The major advantages of my system are that patients from isolated areas can be monitored apprehensively by healthcare professionals. The progressive work of this designed system is very competent in order to make it more accurate and advanced. In the current designed system, we can enhance its working by placing other sensor by which we can measure other vital health parameters, and we can also select advanced microcontrollers for high-speed implications and this would also increase the efficiency of the machine on whole. As wireless data loggers can replace other data loggers as it can assure any measurements taken at the right time. Due to mass advancement in technology, we can never deny to have a data logger which would not require any external source to operate.

References

1. del Rio C, Malani PN COVID-19—new insights on a rapidly changing epidemic. JAMA 323:339–1340
2. Matlani P, Londhe ND (2013) A cloud computing-based telemedicine service. In: 2013 IEEE point-of care healthcare technologies (PHT), Bangalore, pp 326–330
3. Abrar M, Patil RR (2013) Multipoint temperature data logger and display on PC through Zigbee using psoc. Int J Adv Res Comput Commun Eng
4. Kumar R, Rajasekaran MP (2016) An IoT based patient monitoring system using raspberry Pi. In: 2016 international conference on computing technologies and intelligent data engineering (ICCTIDE'16), Kovilpatti, pp. 1–4 A
5. Kale S, Mane S, Patil P (2017) IOT based wearable biomedical monitoring system. In: 2017 international conference on trends in electronics and informatics (ICEI), Tirunelveli, pp 971–9
6. Chavan M, Pardeshi S, Khoje SA, Patil M, Study of health monitoring system. Second International

7. Santhiya K, Kamalakannam B, Muthu Gowtham S, Pandian A, Ram Pradeep R (2018) Human body health monitoring system in the fusion of IoT & cloud computing. Int J Eng Res Comput Sci Eng (IJERCSE)
8. Shallal RaAA (2018) Remote patients monitoring system(heartbeat and temperature) using arduino design of a new antenna for emerging 5G applications operating in the millimeter wave frequency range view project certificates of achievement view project remote patients monitor, researchgate.net
9. Alyasseri ZeaAA (2018) Real-time heart pulse monitoring technique using wireless sensor network and mobile application. Article Int J Electr Comput Eng
10. Kajornkasirat S, Chanapai N, Hnusuwan B (2018) Smart health monitoring system with IoT. In: 2018 IEEE symposium on computer applications & industrial electronics (ISCAIE), Penang, pp 206–211
11. Rohini, Ansari MA, Pal NS (2020) Remote monitoring of vital parameters with IoT-based sensing system

Real-Time Facial Detection Using MATLAB (Utilizing VJ Method)

Preeti Sharma and Suryansh Goel

Abstract In PC writing, face recognition has been quite possibly the most contemplated themes. Given a subjective picture, the aim of the proposed work is the decision of any appearances in the picture along with the detection of eyes and the upper body. While this seems to be a trifling assignment for the individuals, yet it seems a testing job of PCs. The trouble related in recognition of face could be ascribed with numerous varieties of scale, area, light, impediments, and so on. Though many reports reveal approaches for face identification, the Viola and Jones (V&J) face discovery approach is the most important method, which helps to quickly prepare pictures and achieve high detection rate. In the proposed experiment, we intend to examine and comprehend the Viola–Jones calculation by carrying out the discovery outline work, perform execution and analysis to improve the presentation.

Keywords Biometric face detection · Security check · Real-time face detection

1 Introduction

Nowadays, the recognition of facial features has become an important application in computers for identification of a person from an image or a stream of video. The methods of facial detection and identification have become popular in the analysis of images and systems of computer visions [1]. The detection of the face helps to identify the characteristics from the image background, thus forming the initial step for the recognition of the faces. It further helps to obtain these features from faces amidst other obstructions from other things and objects [2]. In the past, various algorithms

P. Sharma (✉)
Chitkara University Institute of Engineering and Technology, Chitkara University, Patiala, Punjab, India
e-mail: preeti.sharma@chitkara.edu.in

S. Goel
Chitkara University Institute of Engineering and Technology, Chitkara University, Solan, Himachal Pradesh, India
e-mail: sgoel.ece18@chitkarauniversity.edu.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_25

were introduced in the area of facial detection and identification from different images and videos. However, each method had their own benefits and limitations [3].

Recently, most of the attention has also been drawn by the biometric methods for identification of human beings. The detection of faces has been used widely for security enhancement at public stations, such as, airport and railways; also, the verification of employees at commercial hubs, identifying offenders in crimes, etc., have shown good performance using these methods.

The detection and discovery of faces caused various difficulties which were accredited using the different components, such as image quality, fluctuations in radiance, facial expressions, and visual angle [4].

The images used for detection must be of optimum quality, which can be attained under various conditions. This ensured the necessary feature extractions which enhanced the robustness of the detecting system. Similarly, the changes in the light caused different facial images to appear in a dissimilar manner. The expression of faces of different people was too altered by the presence of faces. The images of the faces changed as the angles varied with respect to the movement of the axis of the camera. The detection helped to discriminate the faces from the other objects inside an image. The proposed experiment aims to contribute toward the robust detection of faces using efficient scanning of eyes, image acquisition through different postures and directions.

The organization of the paper is as follows; Sect. 2 discusses the literature. Section 3 describes the VJ algorithm followed with the results of performance in Sect. 4. Finally, the conclusion is presented in Sect. 5.

2 Review of Literature

Over the recent years, much related work has been done in the field of detection of faces. Yow and Cipolla [5] gave the multiresolution rule technique based upon the information to utilize to recognize the face by utilizing the primary characteristics of the faces. Most techniques employed the features of faces [6, 7], epidermis tone [8, 9], and other highlights [10] with improved precision. The levels of accuracy are forfeited when high speed is required. Rowley et al. [11] utilized already defined layouts and formats without using the concepts of learning. Emergent techniques gave quick and better results that recognized facial and non-facial characteristics in any natural condition [12–15]. Neutral networks models were usually utilized to achieve desirable results.

The concept of artificial neural networks (ANNs) was initially used for detection of faces [16, 17]. But, its limitations related to use of neuron model, complex computations, and difficult implementations shifted the focus toward the skin-toned methods [18, 19]. These methods were highly sensitive toward illumination under different conditions of light as well as variations in color of skin for different people.

Later, certain methods based on appearance were also suggested for facial detection due to their improved performances in restricted environments [20–22]. The

results utilizing this approach also displayed variations under cases of scaling, head positions, rotations, and illuminations. Similarly, another approach based on features used random datasets [23–25] for good performance. However, these methods gave accurate results under normalized conditions of pose, occlusion, etc., and had computational challenges too.

3 Perspective of Viola–Jones Algorithm

The VJ method is indeed an accurate and a quick process of recognizing any article using a camera. The basic flow diagram as shown in Fig. 1 illustrates the main steps. Firstly, an image is taken as an input. The features are selected using some rectangular windows in Fig. 2. An integral image is obtained by subtracting the obtained features from the white-colored window boxes in Fig. 3. Finally, using different classifiers as shown in Fig. 4, the face is detected. These four steps are used in the implementation of the algorithm as follows:

(a) The regularities of the facial characteristics are matched using Haar features. These features discriminate between light and dark image portions of the face, especially around eyes and nose.
(b) The Haar feature selection is formed as an integral image like a rectangle around the detected area. These are shown in Figs. 2 and 3, respectively.

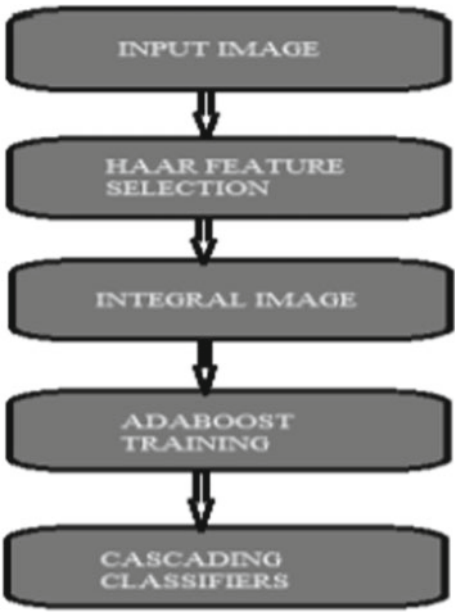


Fig. 1 Flowchart of VJ algorithm [10]

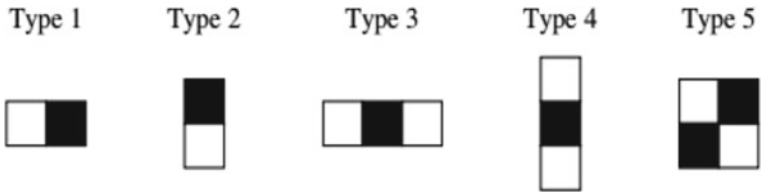


Fig. 2 Various rectangular windows for feature extraction

1	1	1
1	1	1
1	1	1

1	2	3
2	4	6
3	6	9

Fig. 3 Integral image representation

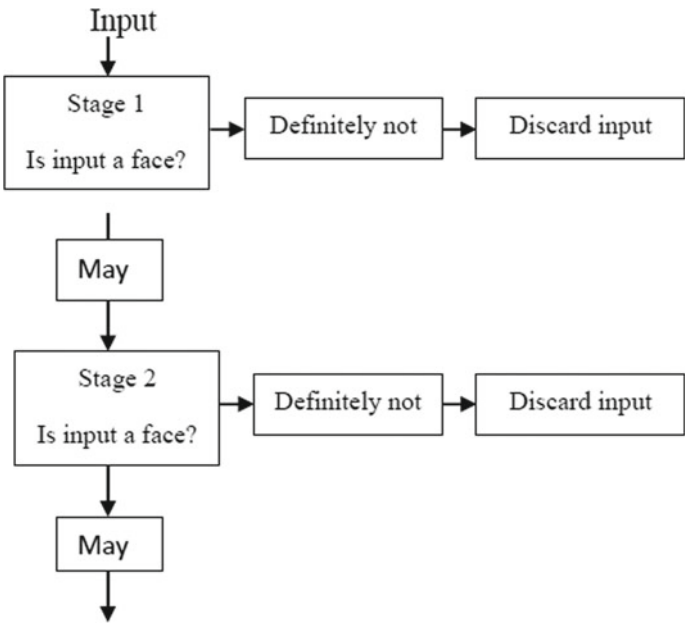


Fig. 4 Sample of a cascaded classifier

(c) The fixed sized rectangle window captures all features. However, only, a few are of significance. These features are selected using AdaBoost machine learning technique.
(d) A cascaded classifier is a series of windows which classify the detection of the face through each stage of sub-window. It uses strict parameter checks to finally give the correct output.

Each manufacturer of the advanced cells and computerized cameras adapt these methods to improve the focus [12].

4 Results of Performance

The VJ algorithm was implemented using MATLAB 2013a version. A graphical user interface was created as shown in Fig. 5.

The image was captured using the camera, wherein the face and eyes were detected in Figs. 6 and 7, respectively.

The results of experimentation given in Table 1 prove accuracy for various conditions of light, rotations, poses, etc.

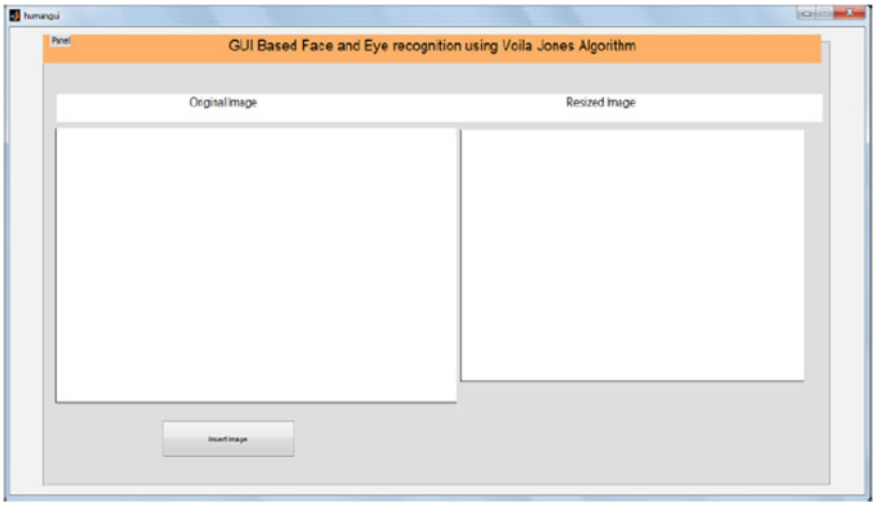
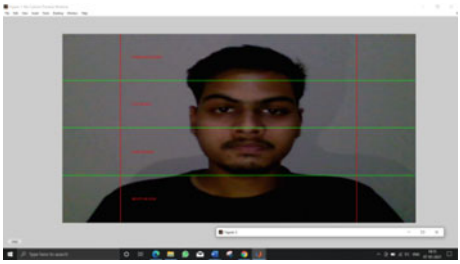
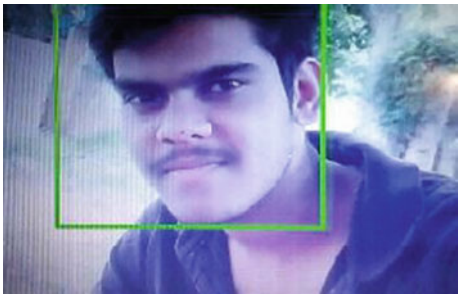


Fig. 5 GUI interface using MATLAB 2013a



(a)

(b)



(c)

Fig. 6 (**a**–**c**) Detection of faces

5 Conclusion

The paper examined the normally utilized facial identification techniques, which implied the utility and significance for various supplications. As such strategies are highly spirited, much development is being achieved to accomplish precise as well as genuine facial recognition. The methodology for recognizing faces was determined to increment the analytical proficiency using MATLAB. Haar features as well as AdaBoost learning helped in facial detection in few fractions of seconds.

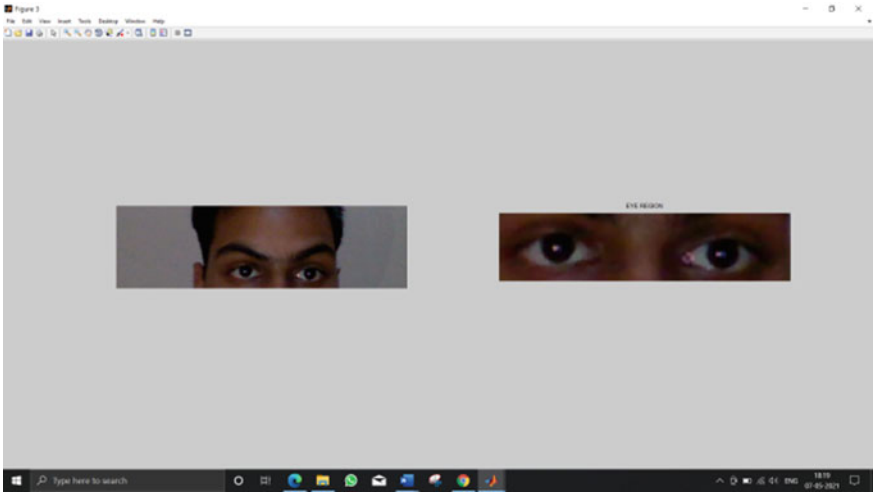


Fig. 7 Detection of eyes

Table 1 Comparative analysis of performance (in accuracy)

System for facial detection	Output (accuracy in %)
Yang et al. [23]	89.4–91.5
Roth et al. [24]	94
Hossen et al. [25]	98.97
Proposed method	98.99

References

1. Yang G, Huang TS (1994) Human face detection in complex background. Pattern Recogn 27(1):53–63
2. Leung TK, Burl MC, Perona P (1995) Finding faces in cluttered scenes using random labeled graph matching. In: Proceedings Fifth IEEE international conference computer vision, pp 637–644s
3. Ritesh Boda, Jasmine Pemeena Priyadarsini M (2016) Face detection and tracking using KLT and Viola Jones. ARPN J Eng Appl Sci 11(23):2212
4. Jain AK, Klare B, Park U (2011) Face RECOGNITION. Some challenges in forensics. In: IEEE international conference automatic face & gesture recognition and workshops (FG 2011), pp 726–733
5. Yow KC, Cipolla R (1997) Feature-based human face detection. Image Vis Comput 1 (9):713–735
6. Yang J, Waibel A (1996) A real-time face tracker. In: Proceedings of third workshop applications of computer vision, pp 142–147
7. McKenna S, Gong S, Raja Y (1998) Modelling facial colour and identity with Gaussian mixtures. Pattern Recogn 31(12):1883–1892
8. Kjeldsen R, Kender J (1996) Finding skin in color images. In: Proceedings of second international conference automatic face and gesture recognition, pp 312–317
9. Craw I, Tock D, Bennett A (1992) Finding face features. In: Proceedings of second European conference computer vision, pp 92–96

10. Lanitis A, Taylor CJ, Cootes TF (1995) An automatic face identification system using flexible appearance models. Image Vis Comput 13(5):393–401
11. Rowley H, Baluja S, Kanade T (1998) Neural network based face detection. IEEE Trans Pattern Anal Mach Intell 20(1):23–38
12. Jaspreet K, Anand S (2017) Performance analysis of face detection by using Viola-Jones algorithm. Int J Comput Intell Res ISSN 0973-1873 13(5), pp 707–717
13. Sharifara A et al. (2014) A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection. In: Biometrics and security technologies (ISBAST), 2014 international symposium on IEEE
14. Monali C, Gauresh V, Dhairya T, Malay S, Amit K (2015) Intelligent surveillance and security system 3(3):2291–2299
15. Cordiner A, Ogunbona P, Wanqing Li (2009) Face detection using generalised integral image features. In: 16th IEEE international conference on image processing (ICIP), pp 1229–1232
16. Yang MH, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. Pattern Anal Mach Intell IEEE Trans 24(1):34–58
17. Mangayarkarasi N, Padmavathi S (2017) Illumination Invariant face detection using Viola Jones algorithm: In: 2017 international conference on advanced computing and communication systems (ICACCS-2017)
18. Dikle SR, Shiurkar UD (2020) Real time face tracking and recognition using efficient face descriptor and features extraction algorithms. In: Iyer B, Deshpande P, Sharma S, Shiurkar U (eds) Computing in engineering and technology. Advances in intelligent systems and computing, vol 1025. Springer, Singapore, pp 55–63
19. Bairagi R, Ahmed R, Tisha SA, Sarder MS, Islam MS, Islam MA (2021) A real-time face recognition smart attendance system with Haar cascade classifiers. In: 2021 third international conference on inventive research in computing applications (ICIRCA), pp 1417–1425
20. Siddiqi MH (2018) Accurate and robust facial expression recognition system using real-time YouTube-based datasets. Appl Intell 48:2912–2929
21. Ngo H, Rakvic R, Broussard R, Ives R, Carothers M (2021) Architecture design for feature extraction and template matching in a real-time iris recognition system. Electronics 10:241
22. Min WY, Romanova E, Lisovec Y, San AM (2019) Application of statistical data processing for solving the problem of face recognition by using principal components analysis method. In: 2019 IEEE conference of Russian young researchers in electrical and electronic engineering (EIConRus), pp 2208–2212
23. Yang MH, Ahuja N, Kriegman D (2020) Face detection using mixtures of linear subspaces. In: Proceedings fourth IEEE international conference on automatic 24(1):70–76
24. Roth D, Yang MH, Ahuja N (2020) A Snow-based face detector. Adv Neural Inf Process Syst 12 (NIPS 12), 51(1):855–861
25. Hossen AMA, Ogla RA, Ali MM (2017) Face detection by using OpenCV's Viola-Jones algorithm based on coding eyes. Iraqi J Sci 58(2A):735–745

Hardening of EMP in Shielded Enclosure for Radiated Fields

Pathala Venkata Sai Charishma and P. V. Y. Jayasree

Abstract Telecommunication and the military are facing the major problem against the electromagnetic pulse (EMP) effect. The electromagnetic pulse is having a high amplitude of a short period of electromagnetic energy. Electromagnetic interference is a significant effect on electrical and electronic systems. The electronic systems must be hardened; the hardening of equipment from EMP should meet the requirements as per MIL-STD 188-125-1&2; standard explains about the design objective for hardening of EMP signal for fixed and movable equipment. The best way to harden the equipment is shielding. Electromagnetic interference [EMI] can be decreased to electronic systems by using shielding. The shielding effectiveness of the shielded rectangular enclosure can be improved and calculated by using the transmission line method [TLM]. This paper proposes to design a metallic enclose as per the MIL-STD-188-125-1&2 with different types of entries like honeycomb structure, single and multiple rectangular waveguides. The comparison between shielding effectiveness values of the theoretical and simulation are done.

Keywords Electromagnetic pulse (EMP) · Electromagnetic interference (EMI) · MIL-STD 188-125-1&2 · Transmission line method (TLM) · Shielding effectiveness · Rectangular enclosure

1 Introduction

Now a day, the use of technologies around the world has been increased. The use of electronic & electrical systems is increased in our daily life. Now, the research trends to minimize the size of electronic systems allow a large number of electrical components to be fixed into small methods. The bulk of electronic systems creates the EMI problem within & between the systems through 2 models radiated EMI, conducted EMI. Therefore, to protect the equipment in industrial and military sectors from the

P. V. S. Charishma (✉) · P. V. Y. Jayasree
Department of Electrical Electronics and Communication Engineering, GITAM Institute of Technology, GITAM Deemed to be University Visakhapatnam, Visakhapatnam, Andhra Pradesh 530045, India
e-mail: cpathala@gitam.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_26

radiated mode EMI, the hardening of the equipment is necessary [1]. The equipment, which comes under the military sector is signal processors, flight controls, control systems, digital engine, is mostly affected by the EMP signal. Telecommunication equipment is radar, television, etc.

The hardening of the system against EMP is done in two ways. Shielding is the most popular technique used; second method is tailored hardening, in which the most affected component in the system is redesigned to get the capability of handling EMP signals.

A high-altitude EMP is a high amplitude pulse of electromagnetic energy which has destructive effects on an unprotected system [2]. Rectangular enclosure has more wavelength than circular enclosure, and it requires less power when compared to the circular, so rectangular is more preferable [3, 4]. In practical use, the shielded enclosure is not solid. The must-have doors, waves guide, holes for cables, honeycomb structure for ventilation [5, 6]. In practical conditions, the shielding effectiveness at a higher frequency of the leakage through the aperture is negligible [7]. TLM method is applied to the EMI problem of shielded enclosures for different entries during the design of enclosures [8]. The aperture is designed. It doesn't allow emissions through it and acts as a waveguide [9, 10].

The paper explains the design and modeling of shielded enclosure for EMP hardening. Shielded enclosures with different entries are discussed; the mathematical calculations of rectangular enclosures with multiple apertures have been calculated. The comparison of simulation results with theoretical results is presented in this paper.

2 Theory and Computation of Shielding

The rectangular shielded enclosure with the same shape waveguide is taken within the enclosure. The SE of the enclosure is calculated by the transmission line method [8, 11]. The shielding is defined as the difference between the electric field inside the outside of the enclosure. It is also defined as the ratio of the incident wave to the wave reflected from the shield. The rectangular shielded enclosure with the same shape waveguide is shown in Fig. 1. The thickness of the enclosure is T, and the point of observation is Q.

Figure 2 shows the equivalent model of Robinson et al. [11], where the enclosure and waveguide are in the exact shape of rectangular.

The electric shielding effectiveness is obtained by taking the voltage at that point Q; the magnetic shielding is derived from the current at that point Q. In Fig. 2, the input voltage V_0 and the impedance I_0, impedance, and the propagation constant of the enclosure are denoted as I_g and P_g.

At first, have to calculate the impedance for the slot and then translate all the voltage and impedances to point Q using a basic transmission line theory. The characteristic impedance I_{os} is given by Gupta et al. [12].

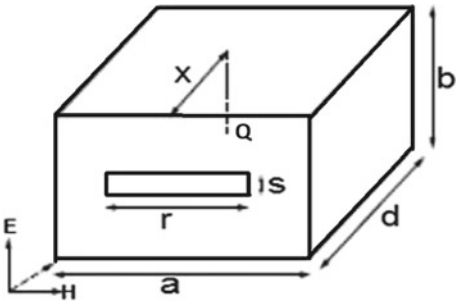


Fig. 1 Rectangular enclosure with a rectangular waveguide

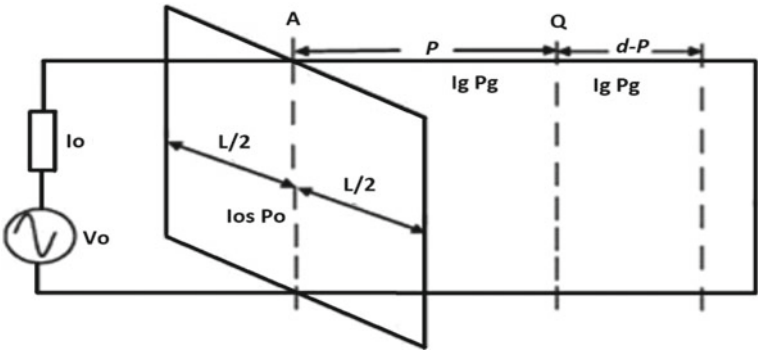


Fig. 2 Equivalent model of enclosure

$$\text{If } W_e = \frac{b}{\sqrt{2}} \quad \text{Where} \frac{W_e}{b} = h$$

$$I_{\text{OS}} = 120\pi^2 \ln\left[2\left(\frac{1+\sqrt[4]{[1-(h)]^2}}{1-\sqrt[4]{[1-(h)]^2}}\right)\right]^{-1} \tag{1}$$

Else

$$I_{\text{OS}} = 120\pi L \frac{W}{b} L' \frac{W_e}{b} \tag{2}$$

where L and L' are elliptic integrals

The effective width is given by

$$W_e = w - \frac{5t}{4\pi}(1 + \ln Q) \tag{3}$$

where $Q = 4\pi w/t$

By short-circuiting the equivalent model, the impedance of the aperture is calculated. The point A is placed at the center of the equivalent circuit. It is important to add a factor Ia for the coupling between the aperture and the enclosure.

$$I_{\text{ap}} = \frac{1}{2}\frac{l}{a} j I_{\text{os}} \tan\left(\frac{k_0 1}{2}\right) \tag{4}$$

This accounts for the connection between the transmission line and the waveguide.

3 Shielding Effectiveness of Enclosure

Shielding effectiveness is defined as the ratio of the incident wave to the wave reflected from the shield. SE of the enclosure is given in the below equation.

The electric and magnetic shielding effectiveness is

$$\text{SE} = -20 \log_{10}\left|\frac{2v_p}{v_o}\right| \tag{5}$$

$$\text{SM} = -20 \log_{10}\left|\frac{2c_p I_o}{v_o}\right| \tag{6}$$

Thevenin's theorem for the equivalent circuit.

$$V_1 = V_A = \frac{V_0 I_{\text{ap}}}{I_0 + I_{\text{ap}}} \tag{7}$$

$$I_1 = I_A \frac{I_0 I_{\text{ap}}}{I_0 + I_{\text{ap}}} \tag{8}$$

For the TE lo mode of propagation, the waveguide has characteristic impedance:
where $f = \frac{\lambda}{2a}$

$$I_g = \frac{I_O}{\sqrt{1 - f^2}} \tag{9}$$

And propagation constant:
where $P_o = \frac{2\pi}{\lambda_0}$

$$P_g = P_o \sqrt{1 - f^2} \tag{10}$$

The I_g and P_g are imaginary at frequencies below the cutoff (equal to $c_0/2a$). At point Q, the waveguide terminals are shorted and transformed by attributing an equivalent voltage V_2, source impedance I_2, and load impedance I_3.

$$V_2 = \frac{V_1}{\cos(I_g Q) + j\left(\frac{I_1}{I_g}\right)\sin(P_g Q)} \tag{11}$$

$$I_2 = \frac{I_1 + jI\tan(P_g Q)}{1 + j\left(\frac{I_1}{I_g}\right)\tan(P_g Q)} \tag{12}$$

$$I_3 = jI_g \tan P_g(d - Q) \tag{13}$$

The voltage at Q is

$$V_P = \frac{V2I3}{(I2 + I3)} \tag{14}$$

The current at Q is

$$C_P = \frac{V2}{(I2 + I3)} \tag{15}$$

The electric and magnetic shielding is

$$\text{SE} = -20\log_{10}\left|\frac{2v_p}{v_o}\right| \tag{16}$$

$$\text{SM} = -20\log_{10}\left|\frac{2c_p I_o}{v_o}\right| \tag{17}$$

3.1 Shielding Effectiveness of an Array of Waveguides

For the correct estimation of shielding, it is essential to consider the mutual coupling between the apertures.

The normalized shunt admittance is ([13, 14]) is given by

$$\frac{Y_{\text{array}}}{Y_o} = -j\frac{3H_s V_s \lambda_0}{\pi d_1^3} + j\frac{288}{\pi\lambda_0 d_1^2}\left[\sum_{\substack{m=0\\ m\neq\text{odd}}}^{\infty}\sum_{\substack{n=0\\ n\neq\text{odd}}}^{\infty}\left(\frac{\varepsilon_m n^2}{V_s^2} + \frac{\varepsilon_n m^2}{H_s^2}\right)J_1^2(X)\right] \tag{18}$$

where λ_0 and Y_0 are the free-space wavelength and intrinsic admittance, respectively, d_v and d_h are the vertical and horizontal separations between the holes, and d is the diameter of the circular hole.

The argument of the Bessel function is

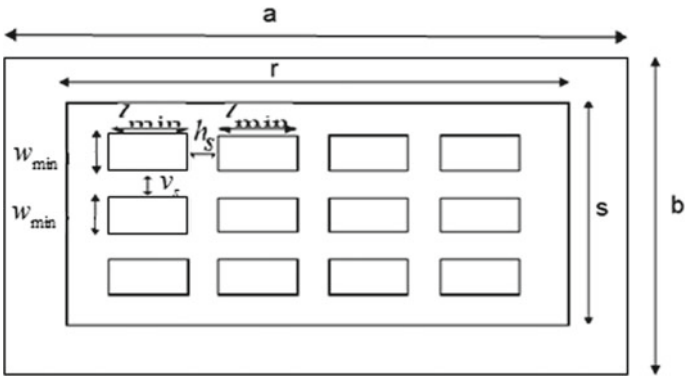


Fig. 3 An array of waveguides in enclosure

$$X = \frac{\left[\pi d\left(m^2/d_h^2 + n^2/d_v^2\right)/2\right]^{1/2}}{\left(m^2/d_h^2 + n^2/d_v^2\right)/2]^{1/2}} \tag{19}$$

where $d = 0.636\ (u + v)$ for rectangular holes

Here, 'u', 'v' are the length and width of the small rectangular waveguide

$I_{\text{ah}} = 1/Y_{\text{ah}}$ represents the array of small rectangular waveguides connecting free space [6, 7]. Figure 3 shows enclosure wall with rectangle enclosure its actual wall impedance is I in part I use a concept of an impedance ratio:

$$I'_{\text{ah}} = I_{\text{ah}}\left(\frac{WL}{ab}\right) \tag{20}$$

where length/and width w of the array are

$$L = \frac{d_h}{2} + (m-1)d_h + \frac{d_h}{2} \tag{21}$$

$$W = \frac{d_v}{2} + (n-1)d_v + \frac{d_v}{2} \tag{22}$$

Here, m and n are the numbers of small holes in the length and width of the array, respectively.

4 Shielding Effectiveness of Waveguide

When the enclosure is absent, the load impedance at 'Q' remains simply as I_g. The voltage at 'Q' is $V_P^{'} = \frac{V}{2}$, and the current is $C'_P = \frac{V_2}{2I_0}$

The electric and magnetic shielding are, therefore, given by

Table 1 SE equation for waveguide

Waveguides shape	SE_{dB}
Rectangle	$SE_{dB} = 27.3 \frac{d}{x\sqrt{1-\left(\frac{x_f}{150000}^2\right)}}$
Hexagon	$SE_{dB} = 17.5 \frac{d}{x\sqrt{1-\left(\frac{x_f}{96659}^2\right)}}$

$$SE = -20 \log_{10} \left| \frac{V}{v_o} \right| \tag{23}$$

$$SM = -20 \log_{10} \left| \frac{V_2}{v_o} \right| \tag{24}$$

Table 1 gives the SE equation for rectangle equation [15], hexagon. The equations' dimensions are present in the table; d is a waveguide, and g is a transverse dimension of the waveguide in millimeters. f is MHz frequency of operation.

5 Shielding Effectiveness Analysis of the Number of Waveguides

The Wiener–Hopf technique analyzes an endless variety of parallel platforms [16]. The resultant equation is seen below in reference [17]. The SE of a rectangular unit waveguide cell is the first term of Eq. (25), whereas the SE of the array of a matrix of parallel waveguides [18] is the second term:

$$SE_{dB} = 27.3 \frac{d}{x} - 20 \log_{10} \frac{2kx}{\pi} \cos \emptyset B \tag{25}$$

where k is the number of the wave, x is a waveguide transversal dimension, and $\emptyset$ is a wave incidence angle. When adding a new SE equation to each term in Table 1 for the second term (1), the SE of honeycombs is Eq. (26)

$$SE_{dB} = 17.5 \frac{d}{x\sqrt{1-\left(\frac{x_f}{96659}^2\right)}} - 20 \log_{10} \frac{2kx}{\pi} \cos \emptyset B \tag{26}$$

5.1 Design Procedure of Shielded Enclosure with Point of Entries

In CST Software, the shielded enclosure with a different point of entry was designed. Entries such as the aperture in the enclosure's side for cable routing, honeycomb, and 178 for heat dissipation in shielded enclosures were discussed in [20]. Access panel with waveguides and connectors in either side shielded enclosure [19]. The parameters of the enclosure are detailed in Table 2.

6 Simulation Analysis of Shielded Enclosure with Rectangular Waveguide

The shielded enclosure with a single rectangular waveguide was designed in CST software, as shown in Fig. 4.

A. Simulation Analysis of Shielded Enclosure with Rectangular Waveguide

In the cut view model for shielded enclosure, the electric field probes are placed at a distance of 1 m from inside and outside (Fig. 5).

The input of the shielded enclosure is the EMP signal of double exponential voltage pulse with amplitude: 55 kV, rise/fall time: 2.3 ns, pulse width: 22.73 ns was applied, which was shown in Fig. 6.

The electric field signal obtained inside the shielded enclosure with a single rectangular waveguide is shown in Fig. 7.

The electric field signal obtained outside the shielded enclosure with a single rectangular waveguide is shown in Fig. 8.

The shielding effectiveness is calculated by taking the electric field data without and with shielding, and plotting the simulation graph is shown in Fig. 9.

Table 2 Shielded enclosure, door, waveguide pipes, and honeycomb dimensions

Parameters	Dimensions (m)
RF shielded enclosure	
Height	2
Width	2
Length	2
Honeycomb panel	
Length	0.45
Width	0.45
Rectangular waveguide	
Length	0.075
Width	0.0375

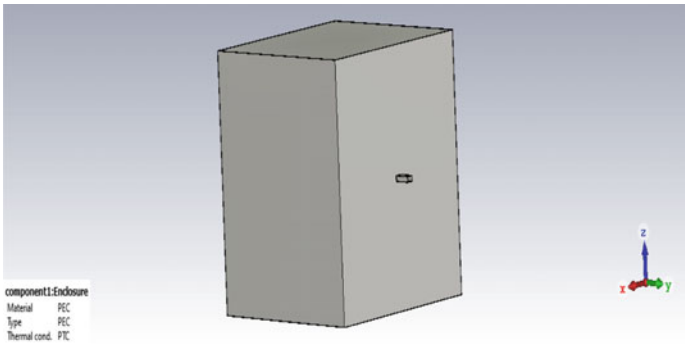


Fig. 4 Shielded enclosure with rectangular waveguide

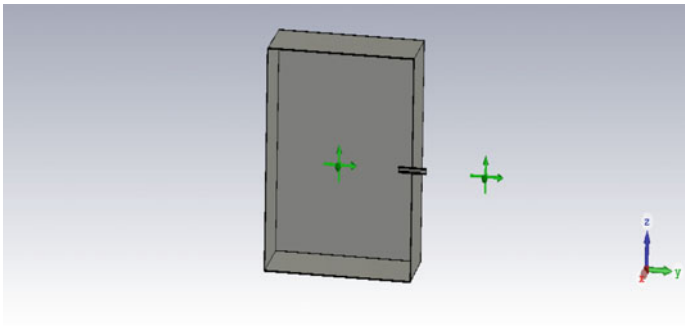


Fig. 5 Cut view model for shielded enclosure with rectangular waveguide

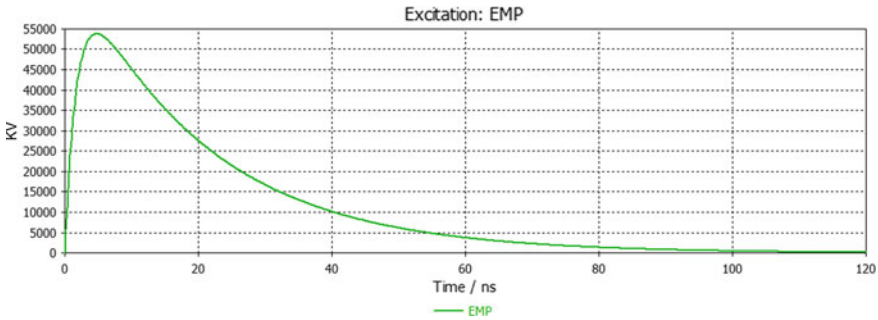


Fig. 6 Input EMP signal

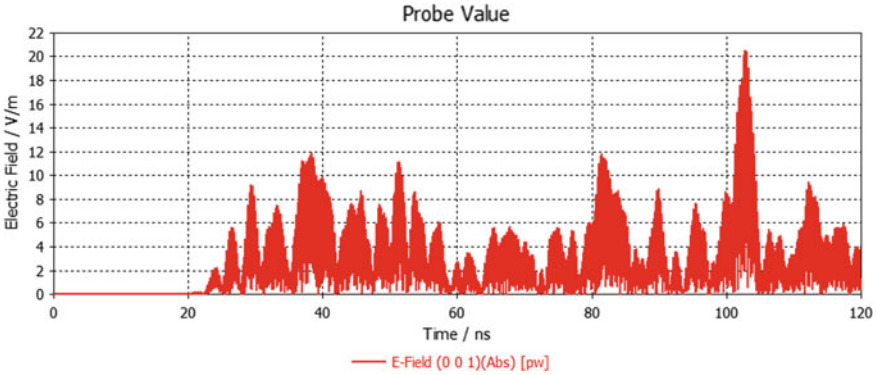


Fig. 7 E field inside the shielded chamber

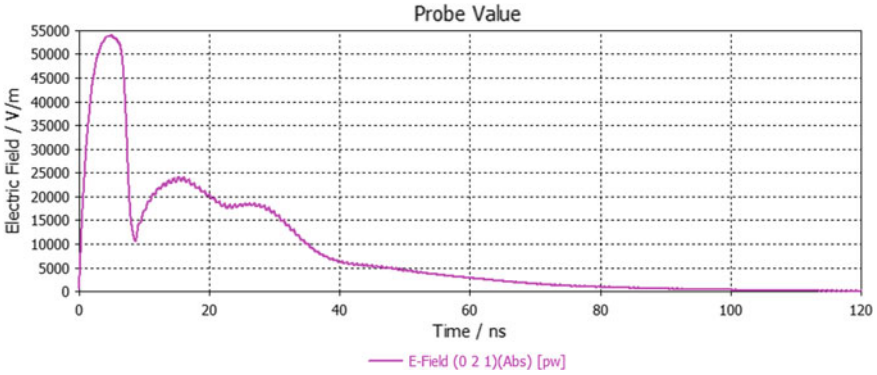


Fig. 8 E field outside the shielded chamber

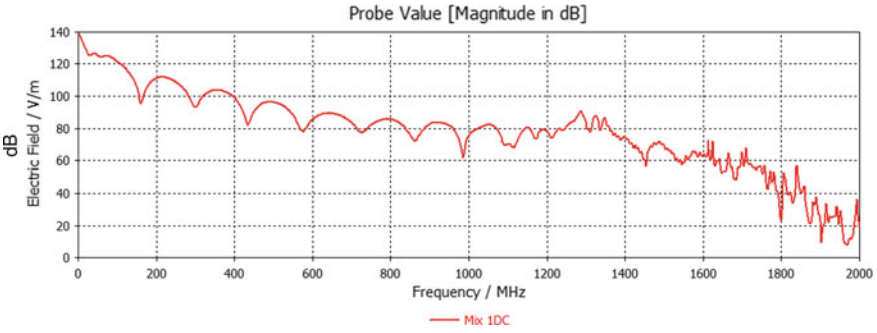


Fig. 9 Shielding effectiveness of a single waveguide is 140 dB

6.1 *Simulation Analysis of Shielded Enclosure with Four Rectangular Waveguide*

The shielded enclosure is designed by taking the array of waveguides in matrix form shown in Fig. 10.

The electric field signal obtained inside the shielded enclosure with four rectangular waveguides is shown in Fig. 11.

The electric field signal obtained outside the shielded enclosure with four rectangular waveguides is shown in Fig. 12 (Fig. 13).

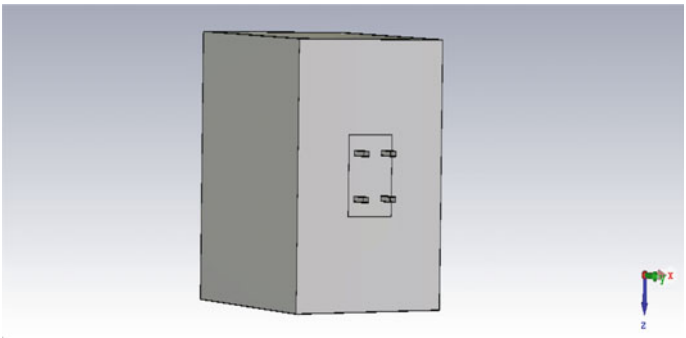


Fig. 10 Enclosure with four waveguides

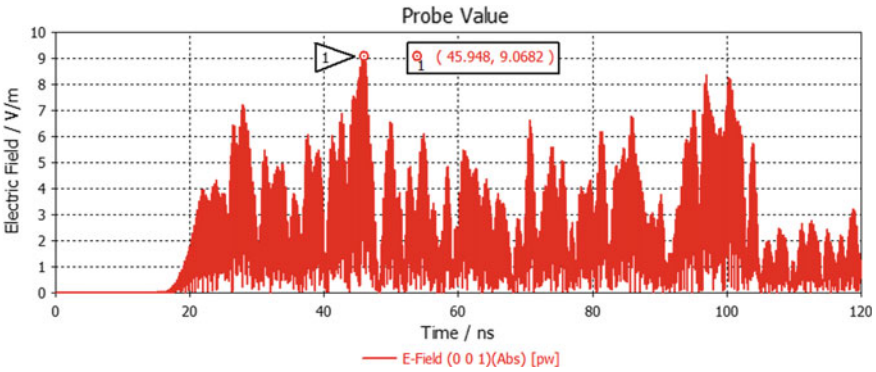


Fig. 11 Electric field inside the enclosure

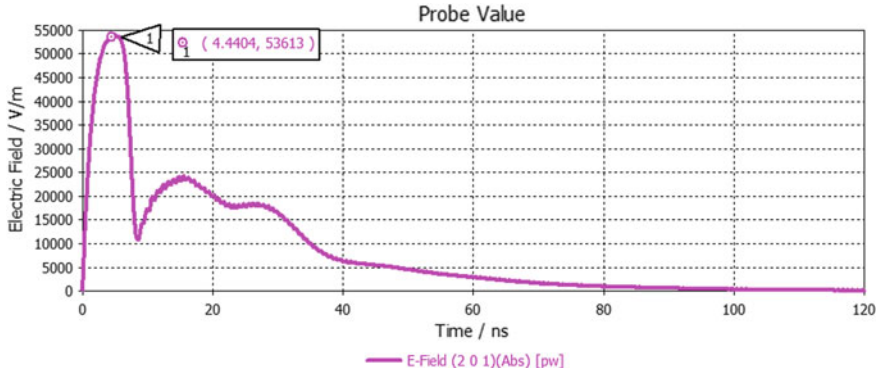


Fig. 12 Electric field outside the enclosure

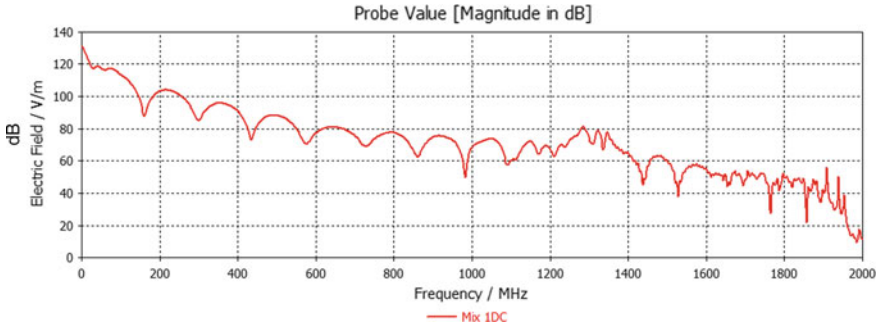


Fig. 13 Shielding effectiveness value for four waveguides is 130 dB

6.2 Simulation Analysis of Shielded Enclosure with Honeycomb Structure

The shielded enclosure is designed by taking honeycomb is shown in Fig. 14 (Fig. 15).

7 Results

7.1 Electric Shielding Effectiveness of Single Waveguide

The shielded enclosure was designed with dimensions of (200 × 200 × 200) cm3; the rectangular slot is placed in the shielded enclosure. The enclosure is excited by an EMP wave with normal incidence for studying the SE of a rectangular shielded enclosure. Figure 4 shows the shielding effectiveness of rectangular waveguides by using the TLM method (Fig. 16).

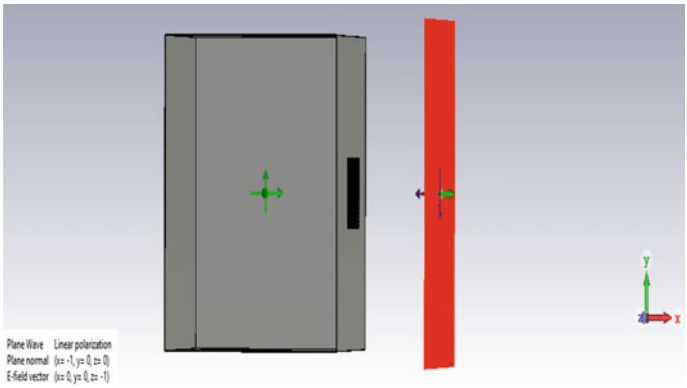


Fig. 14 Enclosure with honeycomb structure

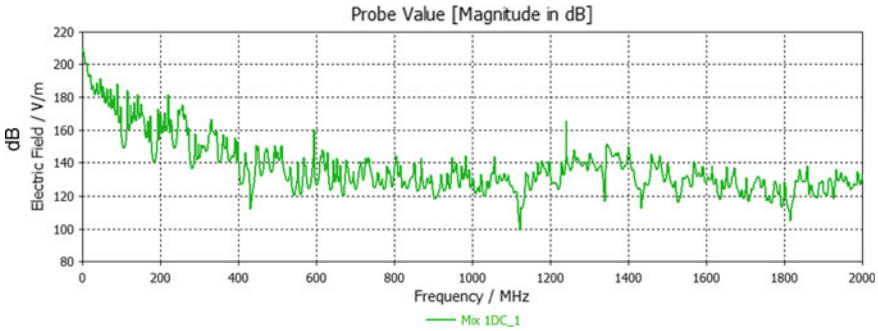


Fig. 15 Shielding effectiveness value for a honeycomb structure is 210 dB

*7.2 Electric Shielding Effectiveness of 2 * 2 Matrix Waveguide*

Figure 17 shows the SE results for shielded enclosure with dimensions200 × 200 × 200 mm^3 with an array of (2 × 2), the rectangular waveguide with separation of 30 cm for both polarization. By using the TLM method, we can predict that the shielding effectiveness decreases with increasing with a number of the waveguide.

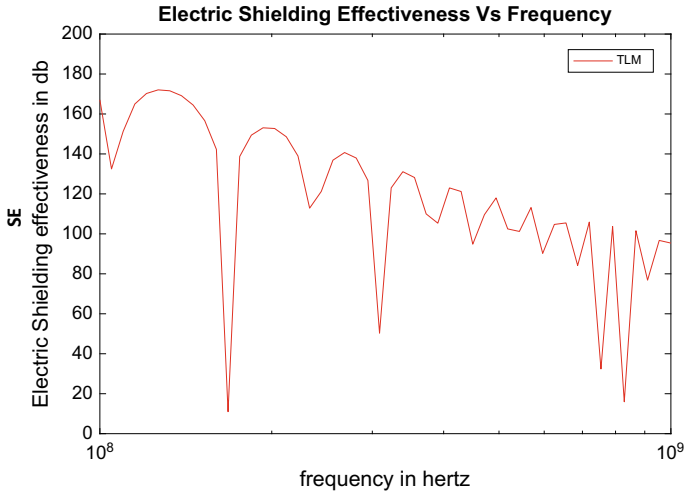


Fig. 16 Calculated SE using transmission line formulation in 200 × 200 × 200 cm^3 box with 7.5 × 3.75 cm^2 waveguide

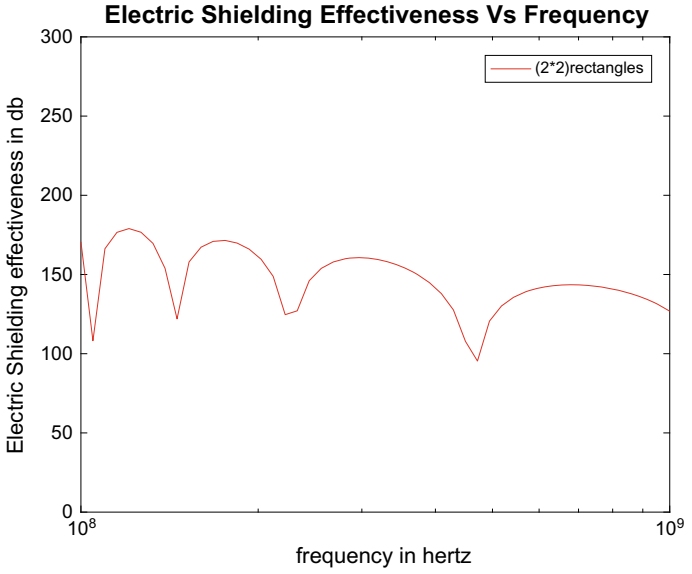


Fig. 17 SE of enclosure using TLM method for 2 * 2 matrix

8 Conclusion

Telecommunication and the military are facing the major problem against the electromagnetic pulse (EMP) effect. The electronic systems must be hardened; the hardening of equipment from EMP should meet the requirements as per MIL-STD 188-125-1&2. This paper explains about the design of metallic enclose as per the MIL-STD-188-125-1&2 with different types of entries like honeycomb structure, single and multiple rectangular waveguides. The comparison between shielding effectiveness values of the theoretical and simulation is done. The variation of SE value in theoretical and simulation is around 20 dB.

References

1. Easton E, Horton R, Bryant K, Butterfield J (2020) Assessment of EMP hardened substation protection and control module. In: 2020 IEEE international symposium on electromagnetic compatibility & signal/power integrity (EMCSI), pp 448–453. https://doi.org/10.1109/EMCSI38923.2020.9191681
2. Wraight A, Prather WD, Sabbath F (2013) Developments in early-time (E1) high-altitude electromagnetic pulse (HEMP) test methods. IEEE Trans Electromagn Compat 55(3):492–499
3. Kumar TRS, Venkatesh C (2009) Shielding effectiveness comparison of rectangular and cylindrical enclosures with rectangular and circular apertures using TLM modelling. In: 2009 applied electromagnetics conference (AEMC), pp 1–4. https://doi.org/10.1109/AEMC.2009.5430647
4. Robinson MP, Benson TM, Christopoulos C, Dawson JF, Ganley MD, Marvin AC, Porter SJ, Thomas DWP (1998) Analytical formulation for the shielding effectiveness of enclosures with apertures. IEEE Trans Electromagn Compat 40(3):240–248
5. Saymeer S, Ramana BV, Subbarao B, Anuradha MS (2019) HEMP hardening of shielded enclosure for critical infrastructure protection. In: 2019 IEEE 5th global electromagnetic compatibility conference (GEMCCON)
6. MIL-STD-188-125-1&2
7. I'm H, Jung I, Yook J, Song H (2018) Analysis of EMP penetration into an enclosure with electromagnetic shielding material. In: 2018 USNC-URSI radio science meeting (joint with AP-S symposium), pp 31–32. https://doi.org/10.1109/USNC-URSI.2018.8602763
8. Zhang X, Zhou YH, Ma RP, Zhang DP (2012) Shielding effectiveness evaluation of rectangular enclosure with high-order modes transmission line model. In: 2012 6th Asia-Pacific conference on environmental electromagnetics (CEEM), pp 292–294. https://doi.org/10.1109/CEEM.2012.6410625
9. Shim J, Kam DG, Kwon JH, Kim J (2010) Circuital modeling and measurement of shielding effectiveness against oblique incident plane wave on apertures in multiple sides of rectangular enclosure. IEEE Trans Electromagn Compat 52(3)
10. Nie B-L, Du P-A, Yu Y-T, Shi Z (2011) Study of the shielding properties of enclosures with apertures at higher frequencies using the transmission line modeling method. IEEE Trans Electromagn Compat 53(1)
11. Robinson MP, Turner JD, Thomas DWP, Dawson JF, Ganley MD, Marvin AC, Porter SJ, Benson TM, Christopoulos C (1996) Shielding effectiveness of a rectangular enclosure with a rectangular aperture. Electron Lett 32(17):1559–1560
12. Gupta KC, Grag R, Bahl IJ (1979) Microstrip lines and plotlines, Chap. 7. Artech House, Norwood, MA
13. Dehkhoda P, Tavakoli A, Moin R (2007) An efficient shielding effectiveness calculation (a rectangular enclosure with numerous apertures). IEEE Int Symp Electromagn Compat

14. Dehkhoda P, Tavakoli A, Moini R (2008) An efficient and reliable shielding effectiveness evaluation of a rectangular enclosure with numerous apertures. IEEE Trans Electromagn Compat 50(1):208–212
15. Otoshi TY (1972) A study of microwave leakage through perforated flat plates. IEEE Trans Microw Theory Tech 20(3):235–236
16. Culshaw W (1959) Reflectors for a microwave Fabry-Perot interferometer. IEEE Trans Microw Theory Tech MTT-7(2):221–228
17. Paul CR (1992) Introduction to electromagnetic compatibility. Wiley, Inc.
18. Bereuter WA, Chang D (1982) Shielding effectiveness of metallic honeycombs. IEEE Trans EMC
19. Mittra R, Lee SW (1971) Analytical techniques in the theory of guided waves. MacMillan, New York, pp 99–104
20. Lee KW, Cheong YC, Hong IP, Yook JG (2005) Design equation of shielding effectiveness of honeycomb. In: Proceedings of ISAP2005, Seoul, Korea

Review on Optimizing Text-Video Retrieval Using CLIP (Contrastive Language-Image Pre-training)

Shelly Sinha, Anupriya, and M. Rathna Chary

Abstract Text-to-video retrieval in the multi-modal research is one of the important and vastly utilized concept in many of the real-world applications. However, it is very expensive process to construct and maintain the datasets in terms of time and thus difficult to get on an outsized scale. The existing methods which are being used for problems related to text-to-video retrieval do not fully utilize the cross-modal cues existing within the video. The contrastive language-image pre-training (CLIP), an image-language pre-training model, has successfully explained the facility of learning from the visual concepts with the help of image-text datasets collected from various sources mainly the Web. During this work, we proposed to find out any such embedding from the input video data which is available with the tongue annotations within the sort of automatically transcribed narrations. We observed models on large scale using How To100M: text dataset for 2 tasks: Firstly, we have done is the standard ranking for the video annotation and the retrieval. Secondly, it was observed that the optimization of text-to-video trained using the sample data results in the higher level of results and the presence of an object in the various datasets such as YT-UGC, HowTo100, and more. In last, it was shown that the video transfers very well to the other datasets too like MSR-VTT and LSMDC. This basic idea helps us in determining the higher level of results for the retrieving of video through the datasets.

Keywords Image and text retrieval · HowTo100M · Visual language optimizing · YT-UGC

S. Sinha (✉)
Department of Computer Science and Engineering, Rajarambabu Institute of Technology, Rajaramnagar, Pune, Maharashtra, India
e-mail: shelly.sinha@ritindia.edu

Anupriya
Department of Computer Science and Engineering, Chandigarh University, Sahibzada Ajit Singh Nagar, Punjab, India

M. Rathna Chary
Department of Civil Engineering, KG Reddy College of Engineering and Technology, Hyderabad, India
e-mail: rathna.chary@kgr.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_27

1 Introduction

Video is the media type which, now-a-days, is most popular and used for almost all of the tasks either by an individual or an organization like media as it has the ability to capture the series of dynamic events, and it appeals very natural to our audio-visual senses. Many online video platforms now-a-days are also promoting the usage of video in many different fields. However, billions and millions of videos on the online platforms are unused because of the improper access to these videos, or because of their low quality, they will not be used much as compared to the other video content. We should be able to access the content efficiently by just writing a simple query in the simple natural language.

In our study paper, we analyzed text-to-video retrieval as well as video-to-text retrieval. In the process of text-to-video retrieval, we will take a question in the form of the caption, for example, "How to assemble the laptop" and in return, the videos which best explains the process of it, i.e., assembling the laptop will be retrieved as the result. Provided the test dataset of text-to-video pairs, now, we have to give the ranks to every video which can be the candidates of the question asked in the form of the query as described above in the example. So, the video which is corresponding to the caption query will be provided with the highest possible rank. On the contrary, process of video-to-text retrieval works on the principle by looking for the suitable candidates from the set of all candidates, thus giving the one which is going to describe the query video.

The commonly used approach for any kind of retrieving problems is the similarity learning [1]. In this, a function is used which generally comprises: a query through which question is being asked and the candidate which results best based on the similarity with the query. Then, all the candidates obtained are given the ranking number based on the consistency in similarity to the query for video retrieving. This ranking is done with the help of captions given to the videos as the videos are mostly represented in the 3D space, based on the similarities calculated among their representations [1]. The important thing to keep in mind here is the way how to provide the captions to the videos so that they will precisely represents both the captions of the video and the content of the video so that the basic idea of using the similarity learning concept achieved efficiently. However, existing datasets like MSR-VTT [2], DiDeMo [3], and EPIC KITCHENS [4] are having tens to many thousands of such pairs making the size very big and that too they are observed manually. Scaling of these large sized datasets manually is a tedious task and inefficient. This manual annotation process will also result in the low consistency.

Another challenge that arises in the video representation is its temporality. This can be made easy by handling the video of variable duration [5]. The current approaches discard the related information by combining the different extractions at various points within a candidate video. Here, we discussed that this discarded related information is often of utmost importance in the retrieval process. As we have seen in Fig. 1, a video of "someone walking toward the object" and "someone walking away from the object" will have an equivalent representation when stacked relatedly.

Fig. 1 Shows "someone walking to" and "someone walking away"

However, in the video query, the movement of an individual relative to the object is important.

We introduced the multi-modal transformer to address the temporal challenges that occurred in the video data. To do this, the multi-modal transformer is used which will be performing the extraction task. This extraction is done based on various features from the different modalities. Also, it is done at various points in the video and then combined during the representation in compact form. Talking about the transformer architecture [6], it is analyzed that it uses the self-attention mechanism to collect the cues both with respect to the cross-modal and the temporal too. These movements are about the events which are occurring in the video. The integration of multi-modal transformer with the already present cross-modal framework lifts both the videos and the captions of it and, hence, estimates their similarity.

2 Related Work

The work, presented here, constructs a video encoder for the process of video retrieval. Convolutional neural networks aka CNNs [7] are the deep learning models which are inspired by the biologically neuron learning models that replaces all the three stages with the pre-trained single neural network. There is comparatively little work to be done on applying CNNs to classify the videos as compared to large image data domains. Since all the successful applications of convolutional neural networks in the domain of image shares the outsized training dataset, we observed that this is sometimes due to the lack of large-scale video classification benchmarks.

2.1 Video Representation Method

As we observed from the previous works focused mainly on the field of 2D/3D spatial–temporal convolution for representing the videos, the recently ViT (Dosovitskiy et al. 2021), which works on a transformer-based image encoder, has drawn more attention. This encoder based on the transformer-based model remains in its early stages for the action classification. However, we mainly focused on the multimodal video-text retrieval using transformer-based video as the backbone [8, 9]. The indexing of video content in order to enable the retrieving of features in computer vision, the sophisticated systems are developed to seek out specific objects, the actions, their predefined categories, irregularities, and also the near duplicates [10]. The transformer-based methods, which lifts the BERT aka bidirectional encoder representations from transformers are applied to S3D works, are known to be specialized in the visual signals; they need not to be studied about the way audio signals are encoded by them.

2.2 A Vision, Language, and Speech Supervision

While most of the works focused particularly on the learning of mapping the visual semantics for both the videos as well as the language, but existing some methodologies started supporting different data forms such as images and the text as the main parameter for computation on the visual frames. Mithun et al. [11] observe that this approach is an easy way of a high-level embedding method.

Visual representation learning is quite a challenging task, and it has been mostly studied with the neural networks learning strategies using supervised or self-supervised methods [12]. Though, in the video, these methods do not contain the variety of information like background noise or any other animated effects, which are as important as other factors as they can influence to great extent in human captions. Generally, approaches deal with well annotated medium-sized datasets, and the captions, descriptive in nature, are gathered for every single clip of the video separately. This process requires great efforts in doing the human annotation, thus, making the medium-sized datasets expensive in nature and difficult to optimize. Results are shown in "Table 1."

In our analysis, the estimation of both the video embedding and the caption embedding is done which then computes their similarities. Zhang et al. performed retrieval of paragraph to video making the assumption of video and paragraph decomposition hierarchically, but we analyzed the method which is not performing the fragmentation of video into number of clips aligning with the sentences corresponding to the captions. An alternative is used which separates the spaces included for various lexeme such as the object or subject of verbs, nouns [18]. In contradiction to the present approach, we will encode the sentence directly with the help of BERT. Another study on the work [19] uses the larger number of language videos within

Table 1 Comparison of various existing datasets containing the video description

Dataset name	Clip	Videos	Text (k)	Time (h)
MSR-VTT [2]	10k	7180	167	40
HT100 [13]	**136M**	**1.222M**	**136**	**134,472**
LSMDC [14]	101k	200	101	150
MSVD [15]	49k	92	56	84
E-KITCHENS [4]	40k	432	432	55
ANET [16]	14k	20,000	69	849
DiDeMo [3]	27k	10,464	41	87
YouCook2 [17]	1.5k	2000	12	176

It is observed that the size of the dataset "HowTo100M," which is studied, bypasses the largest datasets size by greater orders of magnitude typically three orders. Here, M is representing million while k is representing thousand

the dataset "HowTo100M" but is not utilizes together, the characteristics of two entities which are video and the features in VideoBERT [20] and CBT [21]. The analyzed work rather is depending on the extractions of the videos made from the dataset "HowTo100M" videos in order to find out the relative dependencies and also able to address the mismatch in the features of language and visual elements. The experts used by the Mithun et al. for computing the text-to-video similarities for the corresponding experts, namely objects, places, and activity, however they did not show great with their respective similarities. So, one other approach was used in which already computed experts with the help of the other experts are taken into account. Here, similarity between them is obtained by means of the sum of all the input weights associated for similarity between every individual expert.

2.3 Text Retrieval Representation

Text representations which are the earlier works on text representations include Word2Vec [22]. A limitation of those representations is seen during the capturing of the sequential properties when dealing with the sentence as dataset. LSTM [23]—the deep learning models are primarily used for handling this. Other approaches which have shown significant results for text representation are the transformer architecture. It uses the self-attention mechanisms in which the sentence is broken down into words and every single word can attend to all or any of the others. This architecture is piled-up with the fully inter-connected layers and hence forms the base of the modeling network such as BERT [24]. The analysis was performed b Burns et al. using various word inclusions and the language models such as long short-term memory networks, BERT, and more, which are utilized in the vision to language tasks. From this, it was shown that the BERT model which is pre-trained does not

performs well in comparison to the LSTM or maybe considered as an average embedding model. In our analysis, it was studied that the BERT model which is pre-trained outperforms some but not all models used for video retrieval provided that they must be fine-tuned.

3 Related Work Methodology

3.1 Text-Video Retrieval

In this study, we learnt that the combined text-to-video retrieval model was presented which is done from the video clips and captions pairs formed automatically in the dataset. In this model, set of n video clips was given along-with their associated captions. The set of 'n' text and video clips given was represented by $\{(V_i, T_i)\}_{i=1}^{n}$. It denotes by $v\ \varepsilon\ \mathcal{R}$ dv and $t\ \varepsilon\ \mathcal{R}$ dt the d_v and d_t be the representation of the dimensional feature of the video clip say 'V' and text say 'T'. Provided with this information, the goal was to study about the two different functions used for mapping which are: $k : \mathcal{R}^{d_v} \rightarrow \mathcal{R}^{d}$ and $g : \mathcal{R}^{d_t} \rightarrow \mathcal{R}^{d}$. These functions will include the features of text and video, respectively, into a command, say 'd' dimensional space.

Estimation of the similarities: We studied that the similarity between the video and caption was computed as a cosine sum represented by the mathematical expression as:

$$s(v, t) = \frac{(k(v), g(t)}{||k(v)||_2 ||g(t)||_2} \tag{1}$$

It was observed that the cosine sum was high when text 'T' is describing the video clip 'V'. However, it is low if it does not describe the video clip.

Here, the class of function which is nonlinear is used to evaluate and is given by:

$$k(v) = \left(W_1^v v + b_1^v\right) \circ \sigma\left(W_2^v\left(W_1^v v + b_1^v\right) + b_2^v\right) \tag{2}$$

and

$$g(t) = \left(W_1^t t + b_1^t\right) \circ \sigma\left(W_2^t\left(W_1^t v + b_1^t\right) + b_2^t\right) \tag{3}$$

where: $W_1^v\ \varepsilon\ \mathcal{R}^{d \times d_v}$, $W_1^t\ \varepsilon\ \mathcal{R}^{d \times d_t}$, W_2^v, $W_2^t\ \varepsilon\ \mathcal{R}^{d \times d}$, b_1^v, b_1^t, b_2^v, $b_2^t\ \varepsilon\ \mathcal{R}^{d}$ are the learning rate parameter which is denoted by 'σ', and it is based on the sigmoid activation function, and $\circ$ is denoting the multiplication element by element which is aka Hadamard product. Due to this approach, the function used proved effective in other text-to-video applications as they were able to model the nonlinear multiplicative interactions between the dimensions of the input feature vector [25].

Training Strategy Loss Function: Here, the model was trained using the maximum margin ranking loss [25–28]. In this approach, in the training algorithm, the batch of caption and video clips $(V_i, T_i)_{i\,\varepsilon\,B}$, pairs which is represented as: $B = \{i_{1,}, ..., i_b\} \subset \{1, ..., n\}$, are sampled at every iteration and hence updates the parameters with a gradient step of the following loss:

$$\sum_{i \in B} \sum_{j \in N} \max(0, \delta + s_{i,j} - s_{i,i}) + \max\big(0, \delta + s_{j,i} - s_{i,j}\big),$$

where $s_{i,j} = s(V_i, T_i)$ is the similarity score between video clip V_i and text T_i, $N(i)$ is a set of negative pairs for caption-clip i and δ is the margin T.

3.2 Video-to-Text Datasets

HowTo100M [13]. This dataset is having around 1 million instructional videos from YouTube, which have speech transcriptions associated with every video, and thus, they form the video captions. These captions do not necessarily describe the visual content of the video accurately, or we can say that they are temporally misaligned. This dataset can be used only for the network model pre-training phase.

MSR-VTT [2] This dataset is mostly used by the researchers for model testing based on the text-to-video retrievals, and it consists of video segments around 10,000 in number, where 20 captions are there with every single segment. The testing has been done by collecting around 257 queries used for searching purposes and able to retrieve around 118 most suitable videos from the YouTube, for every query. The total duration of the video collected in the dataset is 42 h.

YouCook2 [17] is a dataset containing video related to cooking particularly, and these videos are also collected from the YouTube. These videos clips collected are around 14,000 in number and represent around 89 different recipes. All the clips are labeled well with the textual descriptions. In the dataset, we have the test set and the validation set, but here, in this dataset, since there was no test set clips which are provided with the test descriptions, so evaluation was done on the validation clips which are around 3500 out of the total 14,000 clips. Here, the video which is already present in the dataset HowTo100M is removed from this dataset while analyzing the results.

LSMDC [14] is another dataset which contains the clips of the movies. It stores around 101 k video clip-caption pairs. These clips are provided with the text description which comes either from the script of the movie or from the audio description. The model was evaluated on LSMDC test dataset having 1 k video-caption pairs.

ActivityNet [16] This dataset consists of 20,000 clips of the videos which in turn are having 100 k captions associated with them in total. For most of the videos, these captions cover the complete video and not part of it, but the neighbor captions may overlap. This particular situation, where overlapping occurs, may create a problem

in the testing of the text-to-video retrieval. Imagine, the 2 video-caption pairs represented as $(V_{1,}T_1)$ and $(V_{2,}T_2)$, respectively, where V_1 is the first video segment and has a non-empty overlap with the second video segment denoted by V_2.

MSVD [15] This dataset contains 1970 videos, each with a length that ranges from one to 62 s. The training dataset, test dataset, and validation dataset contain 1200, 670, and 100 videos, respectively [1]. Each video has approximately 40 associated sentences in English.

3.3 Comparison with State-of-the-Art

Different datasets will be having the video clips stored which may vary in number, captions associated with every video, the captions may be either long or short as compared to other datasets, they might have different criteria for the creation of captions and more. Due to these difference in the dataset, it is possible that some may contain greater information than others and hence needs greater time in the training of models while some may have lesser information and hence needs less time to be spent in the training of models using these datasets. However, the problem of overfitting can occur in the models if more training time is used for smaller sized dataset. In this approach, the number of examples which will be used for training the network is changed in the original training procedure selected for the purpose which shows in the result that if the training time is increased, does not gives the better performing model.

CrossTask. In this, the comparison of the trained model using HowTo100M dataset was compared against the approaches which were proposed by Alayrac et al. [29] and Zhukov et al. [30] and which was also the present most recent stage incorporated on CrossTask for not strongly supervised methods. They were having access to the action labels list at the task level, and the only form of supervised training followed by them was the narrations. They also show the upper bound form of fully supervised learning approach [30] that was achieved with the help of the model trained on action segments with annotation. This approach outperforms the mentioned current idea though it was not following the process of localization in clips. The improvement was shown by their method which was same toward all the tasks and shows that that the model trained is not having any biased behavior toward any particular field. For the task "Add Oil to a Car," the improvement observed was more than 30%. So, we can say that their approach also outperforms the upper bound of the fully supervised learning approach. Thus, it can be said that if the model is trained with a larger number of videos narrated is better if compared to the model trained using localization on a smaller but carefully annotated training dataset (Fig. 2).

YouCook2. As it does not provide any standard with which the retrieving of video clips can be compared, so the recent idea used by Klein et al. [17] for text-to-video retrieving was incorporated into this using the features according to their approach. It was observed that the network model trained using the dataset HowTo100M performs

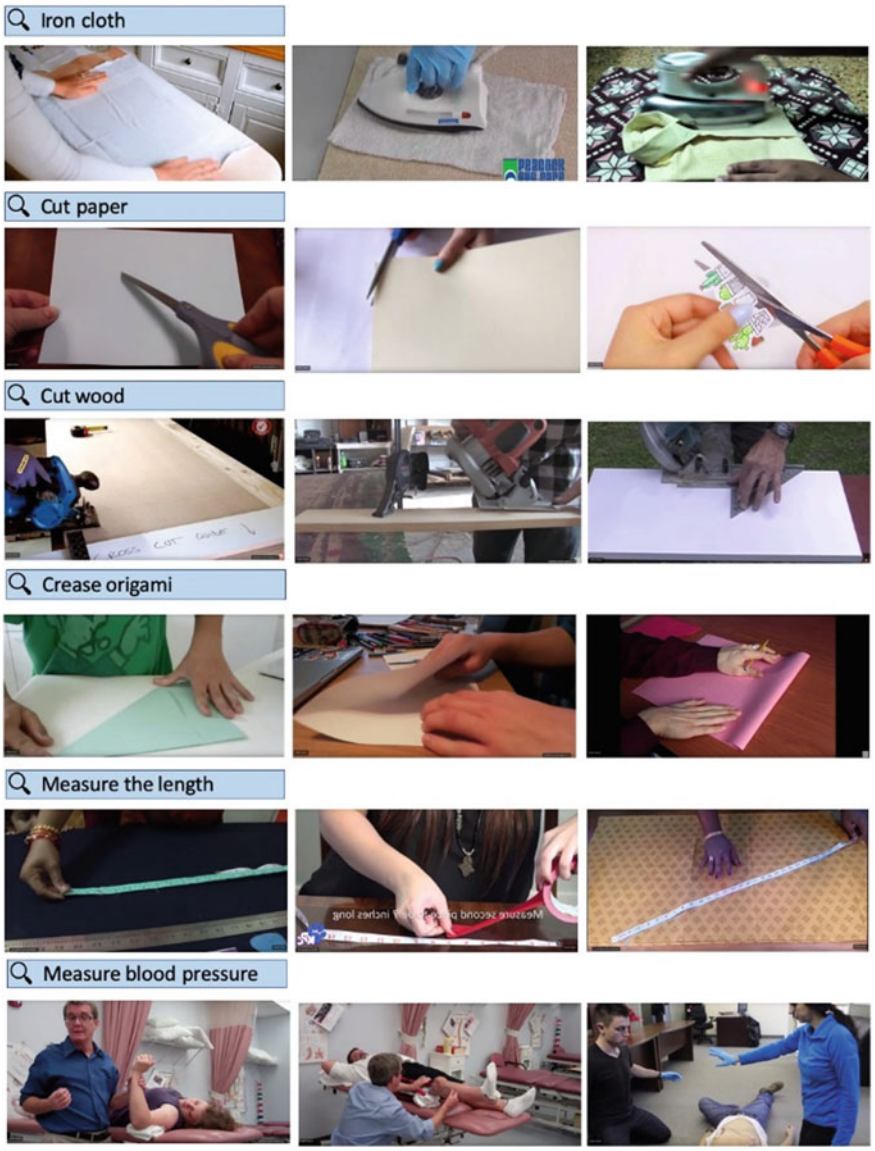


Fig. 2 Example text-video retrieval result on HowTo100M

better than the network model which was trained using YouCook2 directly [1]. This training on the dataset HowTo100M instead of YouCook2 results in a much larger improvement of 13.7%. So, we can conclude that the model trained using

HowTo100M performs better on this particular domain associated with the instructional based dataset of videos. Moreover, this model can further be improved with the help of fine-tuning.

MSR-VTT. In this, the model trained using the HowTo100M dataset only, MSR-VTT dataset only, and the model which was earlier trained on HowTo100M dataset, and after this, fine-tuned on the dataset "MSR-VTT" was compared. **It was observed** that the model trained on the dataset HowTo100M [4, 27, 28, 31, 32] performs better than the models trained on MSR-VTT. Again, it was observed that the pre-trained model using HowTo100M and then fine-tuning on MSR-VTT has improvement than the current idea on training the model using MSR-VTT only [33]. However, it was also observed that if we consider the instructional videos such as CrossTask and cooking videos of YouCook2, training the model on dataset MSR-VTT shows better performance than model trained on dataset HowTo100M. The possible reason for this can be that the video clips on MSR-VTT are YouTube generic videos which are different from the VLOG instructional videos utilized by the HowTo100M dataset videos. When analyzed about the performance of the fine-tuned pre-trained model, it was observed that the current latest idea can be achieved with just 20% of the MSR-VTT samples.

LSMDC. In last, the comparison of the current idea on LSMDC was done. The LSMDC was having distinct movie clips compared to the HowTo100M dataset and thus was a quite challenging task to do. Here, the comparison was done for the various earlier works done and reproduced [33] and are trained on LSMDC. Again, here, it was observed that the model earlier training on HowTo100M dataset and then fine-tuned on LSMDC also shows great improvements than the model which was directly trained on LSMDC.

4 Discussion

The various datasets are introduced including the "HowTo100M" which is a dataset having video clips approximately 130 M in number that are extracted using approximately 1.2 M videos from the various Web sources. These recorded videos are selected which are specially performing the complex visual tasks. This dataset is used in order to find out a combined text-to-video embedding by 130 M approximate video clip-caption pairs. Results are shown through the various experiments that our pre-trained or learned embedding can give better results compared to the other models which are trained on existing annotated but only with the help of datasets having smaller video description. We introduced multi-modal transformer, which extracted many features at various moments, using the various modalities in the video. This lifts both the temporal as well as cross-modal cues. Both of these cues are very important for the exact representation of the video. We incorporated this video

encoder along-with the caption encoder as discussed. It was done during the cross-modal framework for performing the matching of caption-video and acquiring the current approach results for the retrieval from video.

5 Conclusion

In this paper, we focused on the pre-trained CLIP as our base to unravel the process of retrieving the video clip. The experimental results show the effectiveness of the model and achieve the SOTA results on the different approaches, namely MSR-VTT, MSVC, LSMDC, Activity Net, and DiDeMo. Apart from this, the insight from our studies includes mainly the following four. First was that the image feature can also promote the text-to-video retrieval. Secondly, we observed that the post–pre-train performed on even the best image-to-text pre-trained CLIP also able to improve the performance further specially on our base of text-to-video retrieval. Thirdly, it was also observed that similarities between the 3D patch linear projection and sequential type are also giving good results and proving as the promising approaches if focused on the video retrieval task. Last, but not the least, the CLIP used on text-to-video retrieval is sensitive to the learning rate.

References

1. Chen D, Dolan WB (2011) Collecting highly parallel data for paraphrase evaluation. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, pp 190–200
2. Xu J, Mei T, Yao T, Rui Y (2016) Msr-vtt: a large video de-scription dataset for bridging video and language. In CVPR
3. Hendricks LA, Wang O, Shechtman E, Sivic J, Darrell T, Russell B (2017) Localizing moments in video with natural language. ICCV. 1, 2, 5, 12
4. Kauman D, Levi G, Hassner T, Wolf L (2017) Temporal tessellation: A unified approach for video analysis. In ICCV 7:8
5. Liu Y, Albanie S, Nagrani A, Zisserman A (2019) Use what you have: video retrieval using representations from collaborative experts. ArXiv abs/1907.13487
6. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Lu, Polosukhin I (2017) Attention is all you need. In: NIPS
7. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L (2014) Largescale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1725–1732
8. Gabeur V, Sun C, Alahari K, Schmid C (2020) Multi-modal Transformer for video retrieval. arXiv:2007.10639v1 [cs.CV] 21
9. Dzabraev M, Kalashnikov M, Komkov S, Petiushko A (2021) MDMMT: multidomain multimodal transformer for video retrieval. arXiv:2103.10699v1 [cs.CV]
10. Li Q, Li P, Mao K, Lo EY (2020) Improving convolutional neural network for text classification by recursive data pruning. Neurocomputing 414:143–152
11. Miech A, Zhukov D, Alayrac JB, Tapaswi M, Laptev I, Sivic J (2019) Howto100M: learning a text-video embedding by watching hundred million narrated video clips. arXiv preprintarXiv:1906.03327

12. Huang DA, Fei-Fei L, Niebles JC (2016) Connectionist temporal modeling for weakly supervised action labeling. In: ECCV
13. Miech A, Zhukov D, Alayrac JB, Tapaswi M, Laptev I, Sivic J (2019) Howto100m: learning a text-video embedding by watching hundred million narrated video clips. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 2630–2640
14. Rohrbach A, Rohrbach M, Tandon N, Schiele B (2015) A dataset for movie description. In: CVPR
15. Torabi A, Pal C, Larochelle H, Courville A (2015) Using descriptive video services to create a large data source for video annotation research. arXiv preprint arXiv:1503.01070. 2
16. Krishna R, Hata K, Ren F, Fei-Fei L, Niebles JC (2017) Dense- captioning events in videos. In: ICCV
17. Zhou L, Xu C, Corso JJ (2018) Towards automatic learning of procedures from web instructional videos. In: AAAI
18. Chen DL, Dolan WB (2011) Collecting highly parallel data for paraphrase evaluation. In: Proceedings of the 49th annual meeting of the association for computational linguistics, Portland, Oregon
19. Miech A, Zhukov D, Alayrac JB (2019) HowTo100M: learning a text-video embedding by watching hundred million narrated video clips
20. Sun C, Myers A, Vondrick C, Murphy K, Schmid C (2019) Videobert: a joint model for video and language representation learning. In: ICCV
21. Sun C, Baradel F, Murphy K, Schmid C (2019) Learning video representations using contrastive bidirectional transformer. arXiv 1906.05743
22. Mikolov T, Chen K, Corrado GS, Dean J (2013) Efficient estimation of word representations in vector space. In: ICLR
23. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8)
24. Devlin J, Chang MW, Lee K, Toutanova K (2019) Bert: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT
25. Miech A, Laptev I, Sivic J (2018) Learning a text-video embedding from incomplete and heterogeneous data. arXiv:1804.02516. 1, 2, 4, 5
26. Karpathy A, Joulin A, Li FFF (2014) Deep fragment embeddings for bidirectional image sentence mapping. In: NIPS. 5
27. Wang L, Li Y, Huang J, Lazebnik S (2018) Learning two-branch neural networks for image-text matching tasks. PAMI 1(2):5
28. Yu Y, Ko H, Choi J, Kim G (2016) Video captioning and retrieval models with semantic attention. In: ECCV LSMDC2016 workshop. 5, 7, 8
29. Alayrac J-B, Bojanowski P, Agrawal N, Laptev I, Sivic J, Lacoste-Julien S (2016) Unsupervised learning from narrated instruction videos. In CVPR 2(6):7
30. Zhukov D, Alayrac JB, Cinbis RG, Fouhey D, Laptev I, Sivic J (2019) Crosstask weakly supervised learning from instructional videos. In: CVPR
31. Yu Y, Ko H, Choi J, Kim G (2017) End-to-end concept word detection for video captioning, retrieval, and question answering. In CVPR 7:8
32. Kiros R, Salakhutdinov R, Zemel RS (2014) Unifying visual-semantic embeddings with multimodal neural language models. TACL 7:8
33. Yu Y, Kim J, Kim G (2018) A joint sequence fusion model for video question answering and retrieval. In: ECCV. 1, 2, 6, 7, 8
34. Dong J, Li X, Snoek CGM (2018) Predicting visual features from text for image and video caption retrieval. IEEE Trans Multimedia, arXiv:1709.01362v3 [cs.CV]
35. Bulent Sariyildiz M, Perez J, Larlus D (2020) Learningvisual representations with caption annotations. arXive-prints
36. Luo H, Ji L, Zhong M, Chen Y, Lei W, Duan N, Li T (2022) CLIP4Clip: an empirical study of CLIP for end to end video clip retrieval, arXiv:2104.08860 [cs.CV]
37. Portillo JA, Ortiz JC, Terashima-Marin H (2021) A straightforward framework for video retrieval using clip. arXiv:2102.12443v1 [cs.CV]

38. Parcalabescu L, Gatt A, Frank A, Calixto I (2021) Testing the cross-modal capabilities of pretrained V & L models on counting tasks. arXiv:2012.12352v3 [cs.CV] 24
39. Xu J et al. (2016) MSR-VTT: a large video description dataset for bridging video and language. In: IEEE international conference on computer vision and pattern recognition (CVPR)
40. Miech A, Laptev I, Sivic J (2017) Learnable pooling with context gating for video classification. arXiv preprint arXiv:1706.06905
41. Krishna R et al. (2017) Dense-captioning events in videos. In: International conference on computer vision (ICCV)
42. Torabi A et al. (2015) Using descriptive video services to create a large data source for video annotation research. arXiv: 1503.01070 [cs.CV]
43. Chen D, Dolan WB (2011) Collecting highly parallel data for paraphrase evaluation. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies. pp 190–200
44. Rohrbach A, Torabi A, Rohrbach M, Tandon N, Pal C, Larochelle H, Courville A, Schiele B (2017) Movie description. IJCV 2, 5, 6

Argument Mining: A Categorical Review

Sakshi Arora, Ajay Rana, and Archana Singh

Abstract Argument mining is a complex arena built up on natural language processing as its fundamental engine. We may consider it as a specialized branch of natural language processing. There are various aspects involved in argumentation mining, which involve meaning of the input corpus; sentiment of the input corpus; extraction of evidence present in the input corpus; persuasiveness of the input corpus. Each of the aspect is very challenging and has its own presence. But when all of these are put together, these can solve real-world problems. Thereby, making argument mining a lucrative avenue of research. In this paper, we have studied and analyzed various papers with respect to argument mining and its related mechanisms. We have categorically separated various aspects of argument mining and summarized the usage of these aspects in studied papers.

Keywords Argument mining · Natural language processing · Sentiment analysis · Evidence extraction

1 Introduction

Since the advent of communication be it vocal, written, behavioral, arguments have been there. Argument mining, however, is fairly new, and it is capturing of key logic, evidence, reasons from the corpus. These mined arguments can then be utilized for their effectiveness and influence over the counter arguments. It enables the system to judge the winning argument. Mined arguments can also be used to determine the sentiment, hence enabling detection of racism, cyber bullying, and hatred. This in turn

S. Arora (✉) · A. Singh
Amity University, Noida, Uttar Pradesh 201301, India
e-mail: sakshi.arora@s.amity.edu

A. Singh
e-mail: asingh27@amity.edu

A. Rana
Amity University, Greater Noida, Uttar Pradesh 201308, India
e-mail: ajay_rana@amity.edu

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_28

can be used to filter out the corrupted content from vivid social media platforms such Facebook, Twitter, YouTube, and forums. Mined arguments also have significance in field of law, any kind of debates such as electoral debates, parliamentary debates, political debates, educational debates in determining the outcome, and identification of side which has an upper hand. There is a huge corpus available in form of essays, scriptures, social media comments/posts, case arguments that is used for mining arguments.

In this paper, we are mainly focusing on reviewing argument mining machine learning techniques for written text. Here, we have analyzed various aspects of argument mining and have categorically mentioned these. There exist many survey papers around the field of argument mining, but this paper is an attempt to measure various techniques, classifications against each other, by categorizing various aspects of argument mining. This gives a good starting point for reader for the field of argument mining, giving him/her a good holistic summary.

Toulmin broke the argument into six ingredients, namely Claim; Grounds; Warrant; Qualifier; Rebuttal; and Backing [1]. This formed the basis of argument detection. There are two major components of argument, which are used in argument mining.

1.1 Evidence or Premise

In simple terms, evidence or premise in a sentence is the base of the argument. That is, the case on which argument is formed. It decides the direction and intensity with which the decision can be taken. Context dependent evidence has been classified as study, expert, and anecdotal [2]. Study refers to summary after crunching data/numbers. Expert refers to opinion of a domain knowledge expert. Anecdotal refers to time, place, or individual centric approach.

1.2 Claim or Conclusion

Claim or conclusion is the aim or the intent of the argument, that is, the inclination which the argument emphasis on. Claim detection objective can be tackled using SVM [3].

2 Research Method

Aim of this paper is to review and summarize the current state of argument mining. We collected information from various sources to gain a better perspective of the field of argument mining.

Related Work. We also analyzed various related work, that is, various survey papers in field of argument mining. There are papers which captured various aspects of argument mining, for example, Robin Schaefer and Manfred Stede [8] did a survey on Twitter data for argument mining, Lawrence and Reed [7] have extensively worked on classification of various structures of arguments, etc. We have referenced these papers for categorizing some of the aspects of argument mining.

Sources of information, which were pondered include following databases:

- Scopus
- Elsevier
- Web of Science
- Springer
- ACM
- IEEE.

These databases relate most of the relevant journals, conferences, and workshop proceedings for argument mining. We also searched for argument mining as a keyword, to fetch out relevant research papers. There were lot of papers appearing repeatedly, which were filtered out later. Post filtering out duplicates, we analyzed the papers for the relevance of review being conducted. Once we got the relevant set of papers:

- We attempted to identify various aspects of argument mining.
- Post that we attempted to categorize the shortlisted set of papers to the identified aspects of argument mining.

3 Sources of Argument Mining

There are various sources of data that have been employed as input for argument mining over the years. These data sources mostly constitute of written scriptures and form the primary source of argument mining. The research papers using various argument sources to mine arguments have been summarized in Table 1.

3.1 Social Media (like Twitter)

Social media is place where people vent out their feelings. It has become even more relevant in the world because of pandemic situation. Nowadays, social media sites like Twitter, Facebook, Instagram, LinkedIn, etc., are the medium of networking. Reason for mentioning Twitter as a first data source is because of ease of access and richness of content. Twitter provides data free, via APIs for researchers to research. Hence, it is one of the most explored source of argument mining. This source has been reviewed extensively by Robin Schaefer and Manfred Stede [8]. Anastasios Lytos, Thomas Lagkas, Panagiotis Sarigiannidis, and Kalina Bontchevaa have adapted to

Table 1 Table summarizing sources used for argument mining

Sources for argument mining		
Source used	Research paper	Remarks
Social media	[8, 9]	Reed [8] being a survey paper explores Twitter as a source
Wikipedia	[6, 10]	Lytos et al. [9] makes use of multi social media sources
Scientific articles	[11–17]	Cocarascu et al. [6] uses the source for relation prediction
Essays	[18–21]	Wu and Weld [10] uses it for information extraction
Newspapers	[22]	Lauscher et al. [11], Song et al. [15], Wambsganss et al. [16], Teufel et al. [17] talks about adding analysis to this source
legal documents	[23–25]	Fisas et al. [12, 13], Accuosto et al. [14] provide an annotated corpus of this source in various domains
Political debates and speeches	[26–28]	Wachsmuth et al. [18] does quality analysis
product reviews	[29–32]	Stab and Gurevych [19] does insufficiency recognition
Blogs	[4, 33]	Stab et al. [20] works on structures
Online discussions	[34, 35]	Wang et al. [21] works on multi-scale model

the social media generated argument structures and proposed a framework for mining arguments [9].

3.2 Wikipedia

Wikipedia is another most explored source of argument mining, and the reason is same free and ease of access. Fei Wu and Daniel S. Weld have shown various techniques of information extraction form Wikipedia [10]. Oana Cocarascu, Elena Cabrio, Serena Villata, and Francesca Toni have defined the baseline for relation prediction, using Wikipedia articles among others, in argument mining [6].

3.3 Scientific Articles

This source is something which researchers have loved exploring since it is something that can be validated with assertions. Anne Lauscher, Goran Glavas, and Simone Paolo Ponzetto added argumentation annotation layer [11] over Dr. Inventor Corpus [12] and [13]. Pablo Accuostoa, Mariana Neves, and Horacio Saggion have provided

annotation corpus using 510 scientific abstracts in field of computational linguistics (CL) and biomedicine [14]. Ningyuan Song, Hanghang Cheng, Huimin Zhou, and Xiaoguang Wang have done a comparative analysis on argument structures on scientific articles [15]. Thiemo Wambsganss, Nikolaos Molyndris, and Matthias Söllner have used scientific articles to create a domain independent argumentation model [16]. Simone Teufel, Advaith Siddharthan, and Colin Batchelor attempt to prove the reliability of argument zoning categories using scientific articles [17].

3.4 Essays

This is another rich source of argument mining, and many researchers have pondered on it. HenningWachsmuth, Khalid Al-Khatib, and Benno Stein worked on determining an essays's argumentative quality [18]. Christian Stab and Iryna Gurevych have provided insufficiency recognition in an argument using argumentative essays [19], earlier they also worked on annotation of argument structures on persuasive essays [20]. Hao Wang, Zhen Huang, Yong Dou, and Yu Hong have proposed a multi-scale model for argument mining using essays as input source [21].

3.5 Newspaper

With the digitization of newspaper articles, it becomes a very rich source of content. As newspapers include articles relating to almost everything which is going around, it has a unique un-biasness for domain. Hence, it can be best employed to test domain independent argument mining models. Liat Ein-Dor, Eyal Shnarch, Lena Dankin, Alon Halfon, Benjamin Sznajder, Ariel Gera, Carlos Alzate, Martin Gleize, Leshem Choshen, Yufang Hou, Yonatan Bilu, Ranit Aharonov, and Noam Slonim built a newspaper-based massive corpus [22].

3.6 Legal Documents

As legal cases are all about arguments and counter arguments, argument mining has found a lot of ground to cover in this domain. Adam Wyner, Raquel Mochales-Palau, Marie Francine Moens, and David Milward have proposed approaches to mine arguments in legal cases [23]. Huihui Xu, Jaromir Savelka, and Kevin D. Ashley have worked on summarizing of legal documents [24]. Prakash Poudyal, Jaromir Savelka, Aagje Ieven, Marie Francine Moens, Teresa Goncalves, and Paulo Quaresma have published an annotated corpus of the European Court of Human Rights decisions [25].

3.7 Political Debates and Speeches

Political debates and speeches are all about claims and counter claims and evidence that support it. This is the reason of argument mining interest in this source of information. Marco Lippi and Paolo Torroni worked on mining of claims from political debates and speeches [26]. Shohreh Haddadan, Elena Cabrio, and Serena Villata have built a political debate corpus and utility for researchers [27]. Off recent Mare Koit analyzed discussions held in Parliament of Estonia using annotated corpus [28].

3.8 Product Reviews

Reviews used to be about rating and sentiment. But with lot of fake reviews floating around, argument mining came to save the day with its evidence and claim and persuasiveness computations. Haijing Liu, Yang Gao, Pin Lv, Mengxue Li, Shiqiang Geng, Minglan Li, and Hao Wang have proposed to predict helpfulness of review based on argument features [29], and they also published a review corpus of hotels for argument mining [30]. Henning Wachsmuth, Martin Trenkmann, Benno Stein, Gregor Engels, and Tsvetomira Palarkarska have used local sentiments relating them to sentiment score for argument analysis [31]. Marco Passony, Marco Lippiz, Giuseppe Serray, and Carlo Tassoy have predicted usefulness of reviews, by using trained models of argument mining [32].

3.9 Blogs

Blogs form an integral part for one expressing views, opinions, experiences over Internet. Pierpaolo Basile, Valerio Basile, Elena Cabrio, and Serena Villataa analyzed arguments in comments of Italian news blogs to predict whether a comment supported or opposed the article [33]. Theodosis Goudas, Christos Louizos, Georgia Petasis, and Vangelis Karkaletsis proposed a two-step argument extraction process using blogs [4].

3.10 Online Discussions

As the name suggests, discussions happening online. And where there is a discussion, it is all about plausible scenarios and arguments and counter arguments. Makiko Ida, Gaku Morio, Kosui Iwasa, Tomoyuki Tatsumi, Takaki Yasui, Katsuhide Fujita proposed an automated estimator for argument structure identification and an automated agent to further improve the argument by gathering more evidence [34].

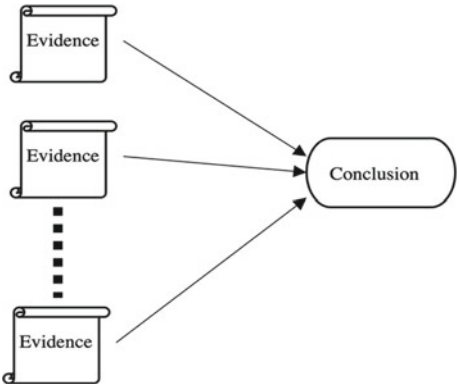


Fig. 1 Figure showing convergent argument structure

Tuhin Chakrabarty, Christopher Hidey, Smaranda Muresan, Kathleen Mckeown, and Alyssa Hwang proposed argument mining computational model that is a fusion of macro and micro-level argumentation models using online discussions [35]. This type of argument has many premises or evidence which on its own lead to one conclusion. So basically, conclusion or claim is fortified independently by each evidence.

4 Structures of Argument

To mine arguments, one would need to understand structure of an argument. Arguments can have various structures. Each structure has its unique properties and needs to be analyzed accordingly. Argument structures' nomenclature has been discussed and detailed out by Lawrence and Reed [7].

4.1 Convergent

This type of argument has many premises or evidence which on its own lead to one conclusion. So basically, conclusion or claim is fortified independently by each evidence. Figure 1 is pictorially representing the convergent argument structure.

4.2 Linked

This type of argument has many premises or evidence working together to lead to one conclusion. So basically, conclusion or claim is inferred together considering all the evidence. Figure 2 is pictorially representing the linked argument structure.

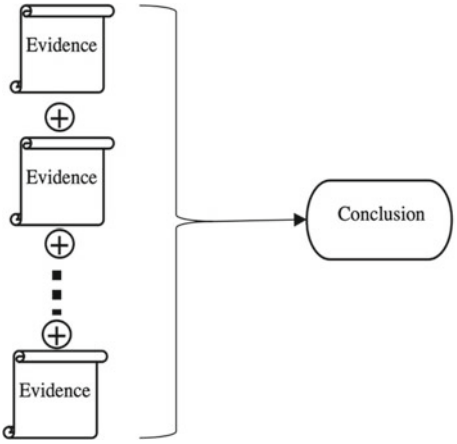


Fig. 2 Figure showing linked argument structure

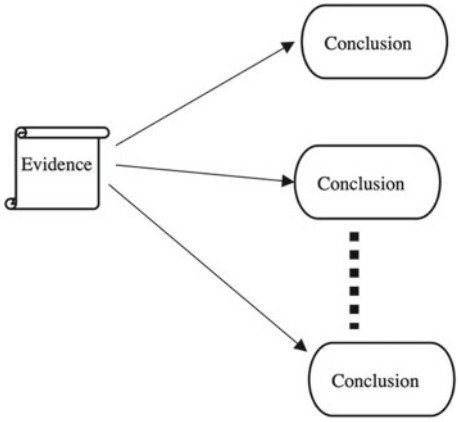


Fig. 3 Figure showing divergent argument structure

4.3 Divergent

This type of argument has one premise or evidence which leads to many claims or conclusions. So basically, multiple claims or conclusions can be inferred from given evidence. Figure 3 is pictorially representing the divergent argument structure.

4.4 Sequential

This type of argument has building blocks as premise or evidence, which when put together sequentially leads to conclusion. So basically, conclusion or claim is reached traversing evidence one after another in order. Figure 4 is pictorially representing the sequential argument structure.

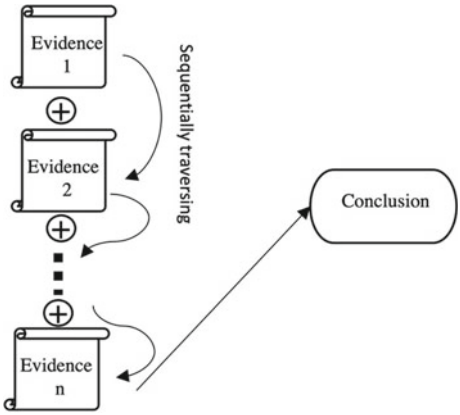


Fig. 4 Figure showing sequential argument structure

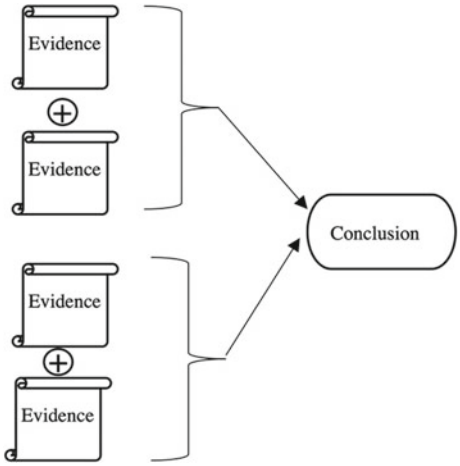


Fig. 5 Figure showing one of the examples of hybrid argument structure

4.5 Hybrid

This type of argument is of complex nature. It can include any argument structure in any combination. These are most difficult to be analyzed and mined. Figure 5 is pictorially representing one of the examples of hybrid argument structure.

4.6 Rebutting

This is very interesting type of argument structure, here evidence conflicts the claim or conclusion. That is, evidence is inferring to the opposite of claim being made. Figure 6 is pictorially representing the rebutting argument structure.

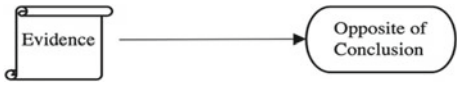


Fig. 6 Figure showing rebutting argument structure

Fig. 7 Figure showing undercutting argument structure

4.7 Undercutting

This type of argument has its evidence or premise infers absence of claim rather than it being present. So, evidence basically nullifies the claim or conclusion. Figure 7 is pictorially representing the rebutting argument structure.

5 Frameworks for Argument Mining

There are couple of frameworks which are utilized together for argument mining. The research papers belonging to these have been summarized in Table 2.

5.1 Argument Extraction

Argument extraction is the first move toward argument mining. One of the approaches to do this is a two-step one, first step is to distinguish sentences which have arguments, and which do not. Second step is to identify portion of the sentence having evidence using conditional random fields [4]. Another way is to use neural events conditional generation tasks [5].

Table 2 Table summarizing frameworks for argument mining

Frameworks for argument mining		
Framework used	Research Paper	Remarks
Argument extraction	[4, 5]	Both papers use a different technique for argument extraction. One is using conditional random fields, whereas other uses neural events
Relation prediction	[6]	This paper describes various models to predict relation, namely concat, mix, autoencoder, and attention base

5.2 Relation Prediction

Relationship prediction in between sentences and arguments is something which is required to arrive at claim or conclusion. There are various neural network approaches to achieve this. Four of them can be concat model, mix model, autoencoder, and attention-based model [6].

6 Features for Argument Extraction and Relation Prediction

Various techniques have been employed for argument extraction and relation prediction over the years. These have a huge overlap with natural language processing techniques as it is a foundation on which argument mining stands. The following are the more prominent techniques.

6.1 Word Level

This is generally relating the count of a word that appears in a document and importance of the word for the document. That is, world level feature extraction is frequency based. One may also have a threshold which can be used to filter out less frequent words. Kristen Johnson and Dan Goldwasser computed agreement and disagreement using word frequency [36]. Additionally, world level features also include bi-gram, trigram, n-gram; here, word appearing in a sequence is treated as a unit.

6.2 Contextual Polarity

This is stitching the precedence with the upcoming data, that is, relating to the scenario presented as a whole. Theresa Wilson, Janyce Wiebe, and Paul Hoffmann used contextual polarity for sentiment analysis on a phrase level [37].

6.3 Dependency Parsing

This is correlation of words with respect to a sentence, to get a sense of its grammatical structure. Ying Li, Zhenghua Li, Min Zhang, Rui Wang, Sheng Li, and Luo Si used a combination of self-attention mechanism and dependency parsing [38]. Yuxiao Ye

and Simone Teufel applied dependency parsing rather than relation extraction for argument mining [39].

6.4 Sentence Splitting

This is forming a tree structure for a sentence and then traversing it. Christian Stab and Iryna Gurevych used sentence splitting for argument mining in essays [19]. Fei Wu and Daniel S. Weld used sentence splitting for Wikipedia information extraction [10]. Michael Fromm, Evgeniy Faerman, Max Berrendorf, Siddharth Bhargava, Ruoxia Qi, Yao Zhang, Lukas Dennert, Sophia Selle, Yang Mao, and Thomas Seidl used sentence splitting keeping idioms and abbreviations in mind to prevent splitting of sentence in between such words [40].

6.5 Negation

This can be thought of in two ways, one as a "know not", that is, lack of knowledge or insufficiency. Other as Prolog's negation as a failure operator [41, 42]. Pierpaolo Basile and Valerio Basile used negative word frequency in their pursuit of extracting message information [33]. Theresa Wilson, Janyce Wiebe, and Paul Hoffmann used negation for classification of polarity [37]. Kristen Johnson and Dan Goldwasser used negation for determining if it is not an agreement [36].

6.6 Feature Cutoff

When doing natural language processing, there can be many features that present themselves. One needs to distinguish between more relevant or more important features and drop the less relevant one. This becomes a little less relevant with the evolution of deep learning which enables to process all features as is.

6.7 History Features

This is a very interesting mechanism; it is a fusion of dynamic programming and localized context. That is, it keeps track of which were the last few features which occurred during processing. Number of features kept track of generally range from one to four.

7 Conclusion and Future Scope

Here, we have analyzed several research papers relating to argument mining. We have categorically summarized various aspects involved in argument mining starting from argument components, sources of arguments, structure of arguments, frameworks employed for mining arguments, and various features basis which arguments are mined. We also categorized various research papers by mapping them against the categories mentioned above. It is also observed that certain types of categories from one aspect gel well with some certain categories in another aspect. This mapping is showcased in the summary tables there in above sections.

Argument mining has a vast horizon and is still far from reaching its full glory. The methodologies and techniques for argument mining are rapidly evolving and have a significant scope of improvement. Argument mining applications are also far from reaching its true potential. Argument mining can be employed in field of law, medicine, customer care and can save a lot of manpower. Argument mining does have presence in these fields, but currently, the state is very basic. It would be very interesting to see how argument mining will reshape these entire fields.

References

1. Toulmin SE (2003) The uses of argument: updated edition. https://doi.org/10.1017/CBO9780511840005
2. Rinott R, Dankin L, Alzate C, Khapra MM, Aharoni E, Slonim N (2015) Show me your evidence—an automatic method for context dependent evidence detection. Conference on—EMNLP 2015 Proceedings of the 2015 conference on empirical methods in natural language processing, pp 440–450. https://doi.org/10.18653/v1/d15-1050
3. Lippi M, Torroni P (2015) Context-independent claim detection for argument mining. IJCAI international joint conference on artificial intelligence 2015 Janua, pp 185–191.
4. Goudas T, Louizos C, Petasis G, Karkaletsis V (2015) Argument extraction from news, blogs, and the social web. Int J Artif Intell Tools 24. https://doi.org/10.1142/S0218213015400242.
5. Li S, Ji H, Han J (2021) Document-level event argument extraction by conditional generation. In: Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 894–908. https://doi.org/10.18653/v1/2021.naacl-main.69
6. Cocarascu O, Cabrio E, Villata S, Toni F (2020) Dataset independent baselines for relation prediction in argument mining. Front Artif Intell Appl 326:45–52. https://doi.org/10.3233/FAIA200490
7. Lawrence J, Reed C (2019) Argument mining: a survey. Comput Linguist 45:765–818. https://doi.org/10.1162/COLIa00364
8. Reed M (2016) A study of social network effects on the stock market. J Behav Financ 17:342–351. https://doi.org/10.1080/15427560.2016.1238371
9. Lytos A, Lagkas T, Sarigiannidis P, Bontcheva K (2019) The evolution of argumentation mining: from models to social media and emerging tools. Inf Process Manag 56. https://doi.org/10.1016/j.ipm.2019.102055
10. Wu F, Weld DS (2010) Open information extraction using Wikipedia. ACL 2010 Proceedings of the 48th annual meeting of the association for computational linguistics, pp 118–127

11. Lauscher A, Glavaš G, Ponzetto SP (2019) An argument-annotated corpus of scientific publications, pp 40–46. https://doi.org/10.18653/v1/w18-5206
12. Fisas B, Ronzano F, Saggion H (2020) On the discoursive structure of computer graphics research papers. In: LAW 2015—9th Proceedings of the 9th linguistic annotation workshop NAACL 2015, pp 42–51. https://doi.org/10.3115/v1/w15-1605
13. Fisas B, Ronzano F, Saggion H (2016) A multi-layered annotated corpus of scientific papers. In: International Conference on Language Resources and Evaluation 2016, pp 3081–3088
14. Accuosto P, Neves M, Saggion H (2021) Argumentation mining in scientific literature: From computational linguistics to biomedicine. CEUR Workshop Proc 2847:20–36
15. Song N, Cheng H, Zhou H, Wang X (2019) Argument structure mining in scientific articles: a comparative analysis. In: 2019 ACM/IEEE joint conference on digital libraries (JCDL), pp 339–340. https://doi.org/10.1109/JCDL.2019.00060
16. Wambsganss T, Molyndris N, Söllner M (2020) Unlocking transfer learning in argumentation mining: a domain-independent modelling approach. 15th International Conference on Wirtschaftsinformatik 2020, Developments, Oppor. Challenges Digit. WIRTSCHAFTSINFORMATIK 2020. https://doi.org/10.30844/wi_2020_c9
17. Teufel S, Siddharthan A, Batchelor C (2009) Towards discipline-independent argumentative zoning: evidence from chemistry and computational linguistics. In: EMNLP Proceedings of the 2009 conference on empirical methods in natural language processing 2009 Aug. A Meet. SIGDAT, a Spec. Interes. Gr. ACL, Held Conjunction with ACL-IJCNLP 2009, pp 1493–1502
18. Wachsmuth H, Al-Khatib K, Stein B (2016) Using argument mining to assess the argumentation quality of essays. COLING 2016—Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers 2016 Dec, pp 1680–1691
19. Stab C, Gurevych I (2017) Recognizing insufficiently supported arguments in argumentative essays. In: Proceedings of the 15th Conference of the European chapter of the association for computational linguistics. EACL 2017, vol 1, pp 980–990. https://doi.org/10.18653/v1/e17-1092.
20. Stab C, Kirschner C, Eckle-Kohler J, Gurevych I (2014) Argumentation mining in persuasive essays and scientific articles from the discourse structure perspective. CEUR Workshop Proceedings, 1341
21. Wang H, Huang Z, Dou Y, Hong Y (2020) Argumentation mining on essays at multi scales. In: Proceedings of the 28th international conference on computational linguistics. International committee on computational linguistics, Stroudsburg, PA, USA, pp 5480–5493. https://doi.org/10.18653/v1/2020.coling-main.478
22. Ein-Dor L, Shnarch E, Dankin L, Halfon A, Sznajder B, Gera A, Alzate C, Gleize M, Choshen L, Hou Y, Bilu Y, Aharonov R, Slonim N (2020) Corpus wide argument mining-a working solution. In: Proceedings of the AAAI Conference on Artificial Intelligence 2020 Apr 3, pp 7683–7691. https://doi.org/10.1609/aaai.v34i05.6270
23. Wyner A, Mochales-Palau R, Moens MF, Milward D (2010) Approaches to text mining arguments from legal cases. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 6036 LNAI, 60–79. https://doi.org/10.1007/978-3-642-12837-0_4
24. Xu H, Šavelka J, Ashley KD (2020) Using argument mining for legal text summarization. Front Artif Intell Appl 334:184–193. https://doi.org/10.3233/FAIA200862
25. Poudyal P, Savelka J, Ieven A, Moens MF, Goncalves T, Quaresma P (2020) ECHR: legal corpus for argument mining. In: Proceedings of the 7th workshop on argument mining 2020, pp 67–75
26. Lippi M, Torroni P (2016) Argument mining from speech: Detecting claims in political debates. In: Proceedings of the AAAI conference on artificial intelligence 2016, pp 2979–2985
27. Haddadan S, Cabrio E, Villata S (2020) Yes, we can! Mining arguments in 50 years of US presidential campaign debates. In: ACL 2019-57th Annual Meeting of the Association for Computational Linguistics 2019, pp 4684–4690. https://doi.org/10.18653/v1/p19-1463
28. Koit M (2021) How are the members of a parliament arguing? Analysis of an argument corpus. In: ICAART 2021—Proceedings of 13th international conferenceon agents artificial intelligence, vol 2, pp 1046–1053. https://doi.org/10.5220/0010314910461053

29. Liu H, Gao Y, Lv P, Li M, Geng S, Li M, Wang H (2017) Using argument-based features to predict and analyse review helpfulness. EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc, pp 1358–1363. https://doi.org/10.18653/v1/d17-1142
30. Duan X, Liao M, Zhao X, Wu W, Lv P (2019) A hotel review corpus for argument mining. https://doi.org/10.1007/978-981-13-7983-3_29
31. Wachsmuth H, Trenkmann M, Stein B, Engels G, Palakarska T (2014) A review corpus for argumentation analysis. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 8404 LNCS, pp 115–127. https://doi.org/10.1007/978-3-642-54903-8_10
32. Passon M, Lippi M, Serra G, Tasso C (2019) Predicting the usefulness of amazon reviews using off-the-shelf argumentation mining, pp 35–39. https://doi.org/10.18653/v1/w18-5205
33. Basile P, Basile V, Cabrio E, Villata S (2016) Argument mining on Italian news blogs. CEUR Workshop Proceedings, pp 1749. https://doi.org/10.4000/books.aaccademia.1706
34. Ida M, Tatsumi T, Morio G, Yasui T, Iwasa K, Fujita K (2019) Can you give me a reason?: Argument-inducing online forum by argument mining. https://doi.org/10.1145/3308558.3314127
35. Chakrabarty T, Hidey C, Muresan S, McKeown K, Hwang A (2020) AmperSand: argument mining for persuasive online discussions. EMNLP-IJCNLP 2019—2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf, pp 2933–2943. https://doi.org/10.18653/v1/d19-1291
36. Johnson K, Goldwasser D (2016) Identifying stance by analyzing political discourse on twitter, pp 66–75. https://doi.org/10.18653/v1/w16-5609
37. Jain TI, Nemade D (2010) Recognizing contextual polarity in phrase-level sentiment analysis. Int J Comput Appl 7:12–21. https://doi.org/10.5120/1160-1453
38. Li Y, Li Z, Zhang M, Wang R, Li S, Si L (2019) Self-attentive biaffine dependency parsing. IJCAI Int. Jt. Conf. Artif. Intell. 2019-Augus, pp 5067–5073. https://doi.org/10.24963/ijcai.2019/704.
39. Ye Y, Teufel S (2021) End-to-end argument mining as biaffine dependency parsing. In: EACL 2021—proceedings of the 16th conference of the European chapter of the association for computational linguistics: main volume 2021, pp 669–678
40. Fromm M, Faerman E, Berrendorf M, Bhargava S, Qi R, Zhang Y, Dennert L, Selle S, Mao Y, Seidl T (2020) Argument mining driven analysis of peer-reviews. https://doi.org/10.5281/zenodo.4314390
41. Green NL (2017) Implementing argumentation schemes as logic programs. CEUR Workshop Proc 1876:1–7
42. Green NL (2018) Towards mining scientific discourse using argumentation schemes. Argument Comput 9:121–135. https://doi.org/10.3233/AAC-180038

Optimizing Text: Representation of Similarity Using Machine Learning

Piyush Yadav, Siddharth Rajput, and Archana Singh

Abstract Strategies, which express high-dimensional information in fewer measurements while keeping a valuable measure of the data content of the first information, are helpful in information preparing and investigation. This paper investigates an impossible use of dimensionality decrease methods, applying procedures amazingly appropriate to representation of information in a few measurements to an assortment of message records in English, which contain an informational index that is not ordinarily dependent upon 2D plotting and spatial examination. The strategies utilized are Semidefinite Embedding (SDE) and Minimum Volume Embedding (MVE), which are both dependent on the overall procedure of Kernel Principal Components Analysis (KPCA). A preprocessed portrayal of text dependent on word recurrence (dismissing word request) is utilized as contribution to these calculations. This kind of portrayal has been demonstrated in other work to be viable at text order with piece strategies, and in a test with few related reports, gives great perception results with a straightforward RBF part, however very little improvement with the more remarkable SDE and MVE methods.

Keywords Dimensionality reduction · Machine learning · Optimization

1 Introduction

Text-based information is among the most pervasive types of information that exist today, and mechanized method for preparing it, performing tasks, for example, looking, classification, interpretation, and amendment, have had incredible achievement and are seeing consistent enhancements. However, the sheer measure of text information implies that even after text is consequently looked, ordered, interpreted, and revised, exploring it may not be simple. To assist with dealing this issue, representation procedures can possibly be applied to give an associate to human route. The thought behind perception for this object is to show archives which are identified

P. Yadav (✉) · S. Rajput · A. Singh
Amity School of Engineering and Technology, Amity University, Uttar Pradesh, Noida, India
e-mail: 01piyush99@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_29

with one another as focuses bunched together in a few dimensional spaces. Visual portrayals like this could have utilized in specific kinds of examination, for instance, in assisting with figuring out enormous chronicles of authentic reports. They could likewise be utilized on the Internet, as backups to indexed lists and the mechanized suggestions ordinarily given on book and film sites. At last, they may have use in the testing and advancement of other regular language handling calculations, giving an instrument to sum up a lot of information, and to assist with assessing test results.

Different visualization techniques have been applied in the past to the issue of record perception, here and there utilizing impromptu strategies. For instance, one framework [2] utilizes a criticism technique, situating focuses addressing records in a space, and iteratively moving the reports in space nearer together or farther separated, contingent upon likeness scores. Another framework [3] utilizes a perception strategy called SAMMON planning to process a comparable kind of low-dimensional portrayal. This paper presents the utilization of two more up to date representation procedures, Semidefinite Embedding (SDE) [10], and Minimum Volume Embedding (MVE) [8] to archive perception and shows the consequences of applying these methods to a little informational collection.

Visual representations like this could have used in certain types of research, for example, in helping to sort through large archives of historical documents. They could also be used on the Internet, as accompaniments to search results and the automated recommendations commonly given on book and movie websites. Finally, they might have used in the testing and development of other natural language processing algorithms, providing a tool to summarize large amount of data, and to help evaluate experimental results. Different visualization techniques have been applied in the past to the problem of document visualization, sometimes using ad-hoc methods. For example, one system uses a feedback method, positioning points representing documents in a space, and iteratively moving the documents in space closer together or farther apart, depending on similarity scores. Another system uses a visualization technique called SAMMON mapping to compute a similar type of low-dimensional representation [19].

1.1 PCA Dimensionality Reduction Techniques

The dimensionality reduction techniques used in this paper can be understood in the context of Principal Component Analysis (PCA). PCA is a widely used technique for dimensionality reduction. In its ideal scenario, a Gaussian distributed (ellipsoid shaped) cloud of M points, $\vec{xj}$, is distributed around some mean point $\vec{\mu}$. Computing the covariance matrix of the points by

$$\sum = \frac{1}{M} \sum_{j=1}^{M} (\vec{xj} - \vec{\mu})(\vec{xj} - \vec{\mu}) \tag{1}$$

The longest ellipsoid pivot, pointing along the heading in the cloud where the information focuses have the biggest change, is portrayed by the eigenvector with the biggest relating eigenvalue. Tomahawks pointing in headings where information has logically less change are portrayed by eigenvectors that have dynamically lower eigenvalues. Projecting information focuses onto tomahawks that have the direction of the eigenvectors, and that all meet at the mean situation of the relative multitude of focuses, and afterward taking the situation along every pivot as another arrangement of directions for each point, gives a substitute portrayal of the information from which it is feasible to recreate the first places of the focuses. With the eigenvectors of C expressed as columns of a matrix $V = (v_{1\ \ldots}\ v_d)$, and the mean-centered, D-dimensional data points given as:

Rows of matrix

$$X = \begin{pmatrix} x_1 - \overline{x} \\ \vdots \\ x_M - \overline{x} \end{pmatrix} \tag{2}$$

The alternate representation is given by $Y = \mathrm{XV}$, where the transformed coordinates of each point can be found in rows of the matrix Y. The direction of the changed information focuses for tomahawks with low eigenvalues will be, naturally, near 0, since differences along these tomahawks are low, and on the grounds that the tomahawks go through the mean of the information. Along these lines, a low-dimensional portrayal of the information can be found by just dropping these directions, and just keeping the directions of the substitute portrayal comparing to tomahawks with high eigenvalues. On the off chance that the eigenvalues of the dropped organizes are exceptionally low, it will in any case be feasible to build great approximations of the first focuses by subbing 0 for those qualities while switching the change. PCA gives the most ideal low-dimensional portrayal of a high-dimensional informational index when the change between the two informational collections must be a straight change [20]. Piece PCA sums up ordinary PCA to permit non-straight changes between the low and high-dimensional portrayals, giving better execution much of the time.

1.2 Kernel PCA

Kernel PCA [6] works, conceptually, by allowing the high-dimensional original data points, $\vec{xj}$, to be mapped to an even higher dimensional intermediate form by an arbitrary function $\varphi(\vec{xj})$, before finding the final low-dimensional representation through the normal process of PCA. To make this idea both computationally feasible and even more general, Kernel PCA recasts the PCA formulas for finding the alternate coordinate representation, so they are expressed in terms of dot products $\varphi\ (\vec{xj})\cdot\varphi\ (\vec{xk})$, and individual $\varphi\ (\vec{xj})$ terms never appear alone. The use of dot products is

important because for many useful φ mappings into high- (or infinitely-) dimensional spaces, it is possible to compute the dot product between two points from the lower dimensional $\vec{xj}$ and $\vec{xk}$ values without ever actually computing the high-dimensional mappings. Functions which compute these types of dot products are called kernels, and are denoted:

$$k(\vec{xj}, \vec{xk}) = \varphi(\vec{xj}) \cdot \varphi(\vec{xk}) \tag{3}$$

The most common way of performing Kernel PCA is not substantially more troublesome than the method involved with performing ordinary straight PCA, in spite of the fact that it does not loan itself as effectively to mathematical instinct. The initial phase in performing Kernel PCA is to process a piece network, K, whose components are given by the accompanying for all information focuses:

$$K_{\mathrm{ij}} = k\left(\vec{xj}, \vec{xk}\right) \tag{4}$$

The second step is to "center" the kernel matrix. This step can seem abstract, but is really just the equivalent of subtracting out the mean $\vec{\mu} \sum_{i=1}^{M} \varphi\left(\vec{xi}\right)$ of the intermediate data points, so the dot products are computed as $(\varphi(\vec{xj}) - \vec{\mu}) \cdot (\varphi(\vec{xk}) - \vec{\mu})$. The efficient way of centering the kernel matrix, without having to compute the high-dimensional mappings, is to compute:

$$\tilde{K}_{\mathrm{ij}} = K_{\mathrm{ij}} - \frac{1}{M} \sum_{m=1}^{M} K_{\mathrm{mj}} - \frac{1}{M} \sum_{n=1}^{M} K_{\mathrm{in}} + \frac{1}{N^2} \sum_{m,n=1}^{M} K_{\mathrm{mn}} \tag{5}$$

Note that for a linear kernel, the centered $\tilde{K}$ matrix is just the $M \times M$ Gram matrix of the centered data points. These contrasts with the computation of a $D \times D$ covariance matrix during the first step in Linear PCA, and can mean data storage requirements will be lower or higher in Kernel PCA depending on whether the number of points in the dataset are lower or higher than the number of dimensions. The next step in Kernel PCA is to find the eigenvalues and eigenvectors of the $\tilde{K}$ matrix,

$$\tilde{K}\, \vec{vj} = \lambda_j\, \vec{vj} \tag{6}$$

With the eigenvectors normalized so $\lambda_j(\vec{vj} \cdot \vec{vj}) = 1$ for all j. As a final step, the transformed representation, Y, of the original data points may be found by:

$$Y = V\Lambda \tag{7}$$

where the transformed representation of each point forms a row in Y, where $V = (\vec{vl} \ldots \vec{vd})$ is a matrix of the normalized eigenvectors as columns, and where Λ is a matrix with the square root of eigenvalues $\sqrt{\lambda 1} \ldots \sqrt{\lambda M}$ along its diagonal [7, 9]. Low-dimensional representations can be found by sorting the eigenvalues and eigenvectors in descending order of the eigenvalues, and skipping computation of the rightmost columns of matrix Y. In addition to compute low-dimensional representations of the original data points with Kernel PCA, it is also possible to compute representations of arbitrary points [6], although it is not necessary to do so in the application described in this paper.

1.3 Semidefinite Embedding

Part PCA gives an incredible representation procedure; however, it leaves open the subject of what kind of piece capacity ought to be utilized with it. The portion work is the thing that characterizes the planning of the first high-dimensional information into the middle higher dimensional space where PCA takes place. For data that is Gaussian distributed, in an elliptical cloud with low variance in some directions, and high variance in others, the linear kernel $k(\vec{xj}, \vec{xk}) = \vec{xj} \cdot \vec{xk}$ may work very well. For data occupying more complicated shapes, polynomial $\left(k(\vec{xj}, \vec{xk}) = (\vec{xj} \cdot \vec{xk} + 1)^p\right)$ and RBF kernels

$$(k(\vec{x}_j, \vec{x}_k) = \left(k(\vec{x_j}, \vec{x_k}) = e^{-\left\|\vec{x}_j, \vec{x}_k\right\|^2 / 2\sigma^2}\right) \tag{8}$$

may provide better results. There are additionally numerous area explicit bits, intended to chip away at picture, text, or discourse information, for instance. One limit of numerous parts with the end goal of perception is that they frequently give poor measures of fondness between far off information focuses which are not firmly identified with one another, in any event, when they do give valuable measures for focuses that are nearer together. For instance, it is feasible to envision a text bit that gives a valuable proportion of closeness between related records that have a similar theme and design, however, give less significant scores between reports that are more unique. This can be an issue any time a shut structure part is utilized on a high-dimensional, however profoundly obliged informational indexes like English text, or normal pictures. Information focuses on these kinds of informational collections will in general happen on contorted, low-dimensional manifolds inside the high-dimensional vector space used to depict them. Powerful perceptions for these informational collections need to show the general situation of information focuses along the manifolds where they happen, rather than their situations in the bigger vector space, and the changes done by shut structure pieces will be unable to precisely catch the state of the complex when giving the affinities for focuses

which are not near one another. Semidefinite Embedding (SDE) [10] endeavors to advance the present circumstance by learning a Kernel framework which adequately levels out the complex. It depends on typical spot item calculation to discover affinities between focuses which are firmly identified with one another and afterward it extends or unfurls the complex by endeavoring to expand the distance between all information focuses which are not connected together. As in put, the SDE calculation takes an availability grid, η, indicating the focuses are dared to lie near one another on the complex, and a Gram network, G, portraying the partiality between focuses that are straightforwardly associated, or are associated through a neighbor. Semidefinite programming, a raised enhancement strategy, is utilized to an ideal lattice, which is connected as the Kernel framework in Kernel PCA to yield the ideal perception. The semidefinite programming issue is portrayed by a target capacity to be augmented, and a rundown of limitations. The primary requirement is that the yielded piece grid, K, is positive semidefinite ($K > 0$). This constraint is natural because, all kernel matrices (matrices whose elements are expressed as dot products of high-dimensional mappings, as in $K_{ij} = \varphi(\vec{xj}) \cdot \varphi(\vec{xk})$) are positive semidefinite. The second constraint is $\Sigma_{ij} K_{ij} = 0$, which limits the output to the set of "centered" kernel matrices which give meaningful results with Kernel PCA. The third constraint is expressed as $K_{ii} + K_{jj} - K_{ij} - K_{ji} = G_{ii} + G_{jj} - G_{ij} - G_{ji}$ and must hold for all points i and j which are neighbors of each other or of a common point. This third requirement is the thing that safeguards the distances and neighborhood connections between sets of associated focuses. The target capacity to be augmented is tr (K). Amplifying this worth boosts the distance between every unconstrained point and is the thing that makes the complex be pulled separated, unfurling, and leveling it. Since SDE accepts a discretionary Gram framework as information, any part capacity can be utilized to register the neighborhood affinities that SDE jam in the last inserting. To figure the network lattice input, k-closest neighbors, or some other bunching calculation might be utilized.

1.4 Minimum Volume Embedding

One shortcoming of SDE is that the target work it augments, tr (K), builds the distance between unassociated focuses in each measurement, rather than simply the a few measurements being envisioned. Since we just at any point see the distances between focuses in the measurements related with the most elevated eigenvalues of the bit network, this is not the target work we genuinely look to augment. Thusly, there are situations where SDE would not make an ideal showing reshaping the complex for perception, like when the availability chart has a center and spokes structure. Least Volume Embedding (MVE) [8] can give further developed outcomes by expanding distances between focuses in the main d noticeable measurements, while at the same time, limiting the distances between focuses in any remaining measurements. This compares to expanding the worth of

$$\sum_{i=1}^{d} \lambda i - \sum_{i=d+1}^{N} \lambda i \tag{9}$$

which is a modification of the objective function used in SDE, recalling that $\mathrm{tr}(K) = \sum_{i=1}^{N} \lambda i$, Semidefinite programming combined with an iterative alternative minimization scheme makes MVE more computationally expensive to implement than SDE, but with better results for certain types of visualizations.

2 Literature Review

For a long time, issues happened include determination has been inspected by the measurements and AI people group. As of late, it has gotten a lot of fixations since study in information mining is excited. In associations for AI and information mining, 220 experiences have addressed problems with element determination, notably in regulated and 221 unassisted models [5]. In the managed setting, the meaning of a component can be assessed by its relationship with the class mark: Fisher score (FS), Relief and Relief [8], Fast Correlation-Based Filter (FCBF) [9], and Spectrum deterioration (SPEC) [10].

Because of the shortfall of class marks which would direct the pertinent information search, the solo element determination is considered as a considerably more troublesome issue. The solo Variance score (VS), Laplacian score was discussed in [11], SPEC [10], Hilbert–Schmidt Independence Criterion (HSIC) [12].

The picking strategy is a fundamental part of component choice. Ghostly component choice figures out how to pick highlights delivering the developments of the chart energized from a bunch of pairwise case similitude [10]. Ghostly component determination can take various structures which can be recorded as: detachability, information reliance, unwavering quality, execution of learning model which utilized in the covering model, and so forth.

2.1 Component Extraction

From past study to accessible the mark data, the element extraction methods can be classified as managed FE or solo FE. One of the examples of managed highlight extraction techniques is Fisher Linear Discriminant (FLD), which can separate the ideal discriminant vectors when class names are free. Additionally, unaided element extraction techniques, the famous Principal Component Analysis (PCA) endeavors to reservation the extensive covariance construction of information when class names are not free. Extra strategies can be start in the writing business with include extraction, (LLE: Locally Linear Embedding), (KPCA: Kernel PCA), (LE: Laplacian Eigen map) [15], and (LPP: Locality Preserving Projection) [13, 14].

3 Methodology and Discussion

1. Raw data was preprocessed, and furnished datasets were utilized.
   a. Deleting columns which contained any missing qualities from dataset.
   b. Introducing the methodology of "missing qualities proportion" calculation.
   c. Listing all out qualities.
   d. DE normalizing information (performed highlight scaling).
2. The dataset was further divided into 66% training and 33% as testing dataset.
3. Selection of model must be applied. The classification algorithm used was Neural Network, Random Forest, and Support Vector Machine to classify the data. The accuracy of the model was compared, and best was picked.

4 Experiments

4.1 Text Representation

A moderately basic message portrayal, which has been displayed to give great outcomes for record grouping with bit strategies [4], was utilized in this venture, and gave great outcomes for report perception with Kernel PCA methods [17, 18]. Each archive was addressed as a vector of word counts. During beginning parsing, stop-words like "and", "or", and "the" were taken out utilizing an English stop-word list [5], and a stemming calculation [1] was utilized to plan the connected words with various postfixes (for instance, "registers", "figuring", and "PC"), into an authoritative word-stem structure, so their includes would be accumulated in the subsequent component vectors. The subsequent word count vectors were standardized to aggregate to one to consider more significant correlations of archives with various lengths. The word counts were then weighted by opposite report recurrence (IDF) values registered by

$$\mathrm{IDF}(w_i) = \log\left(\frac{n}{\mathrm{DF}(w_i)}\right) \tag{10}$$

For each word w_i where n was the all-out number of reports in the set, and DF (w_i) was number of archives the word showed up in (the record recurrence). Weighting by IDF helps give words which just happen in certain reports, which can be acceptable pointers of topic, more impact over the perception results. This vector portrayal was utilized with a RBF part to figure record affinities [16, 20]. The RBF bit was picked because it has been displayed to function admirably with comparable message portrayals for message order. A k-closest neighbors' calculation was utilized to register the availability grids passed as info the SDE and MVE executions.

4.2 Dataset

The informational index utilized was an assortment of issue reports from the 2008 official competitors' mission sites. Each report was marked for its theme (climate, medical care, international strategy, and so forth), and the competitor who sees it communicated (Clinton, Giuliani, and so on). Having an informational index with two unmistakable names for each point was relied upon to make representation seriously fascinating, and to make it simpler to search for designs in yield.

5 Results

The consequences of utilizing Kernel PCA with a RBF piece and Kernel PCA with SDE and MVE bits are displayed in Fig. 1. The perceptions created by these strategies were all practically the same, with the SDE and MVE representations being totally in recognizable. In the diagrams, records are unmistakably grouped together by point. It was more enthusiastically to track down any reasonable relationship between the situation of records in the chart and the applicants' sites they started from. Changing the sigma boundary utilized in the RBF portion did not appear to recognizably affect representation, except if the qualities were very enormous or tiny, in which cases the perception stopped to give significant outcomes. Changing the quantity of neighbors found by the k-closest neighbor calculation additionally did not have a lot of impact on outcomes whenever they were set above around 20% of the all-out number of archives (Figs. 2, 3, and 4).

6 Conclusions

The Kernel PCA techniques utilized in this venture had the option to produce significant perceptions of text reports utilizing an exceptionally straightforward documentation which depended exclusively on individual word counts and did not consider word request or archive structure. It is fascinating to check whether perception could be improved with a more refined portrayal that considered these things, and to check whether the complex leveling SDE and MVE methods would give preferable outcomes over the basic RBF piece, given more perplexing and obliged portrayals. It would likewise be smarter to test these representation methods on bigger informational indexes than the one utilized in this analysis, and to test with informational collections containing various kinds of message, for instance, paper articles, reference book articles, pages, to perceive how well the perception can bunch related records written in various styles and arrangements.

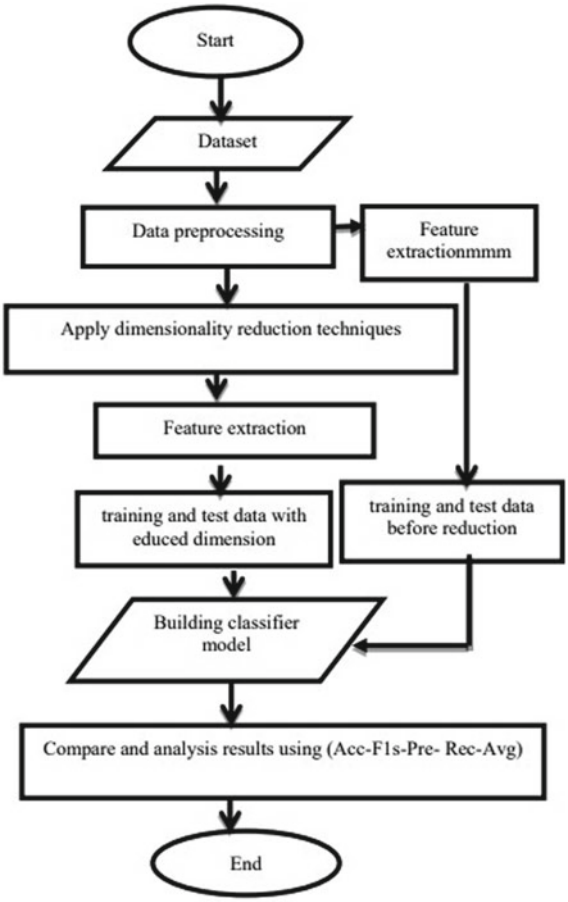


Fig. 1 Methodology

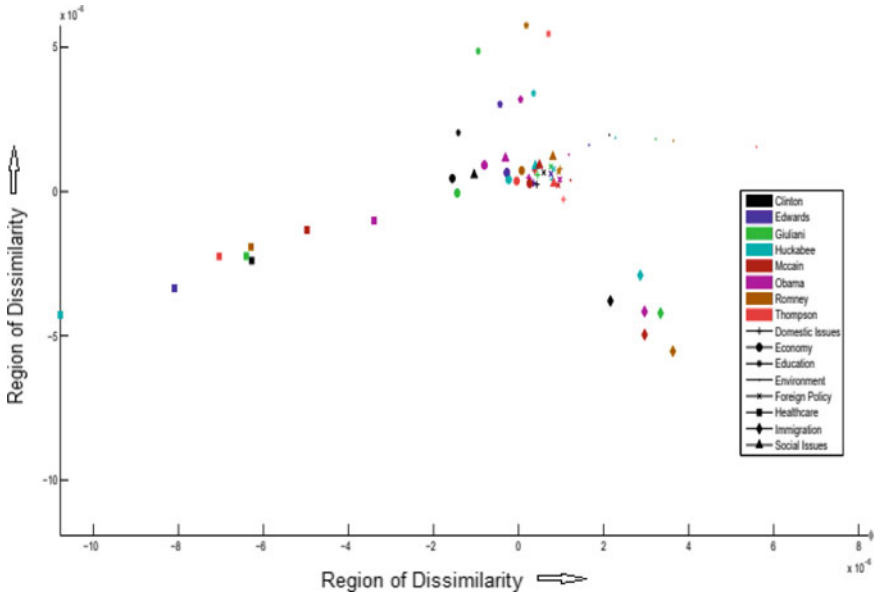


Fig. 2 Visualization results using an RBF kernel with Kernel PCA

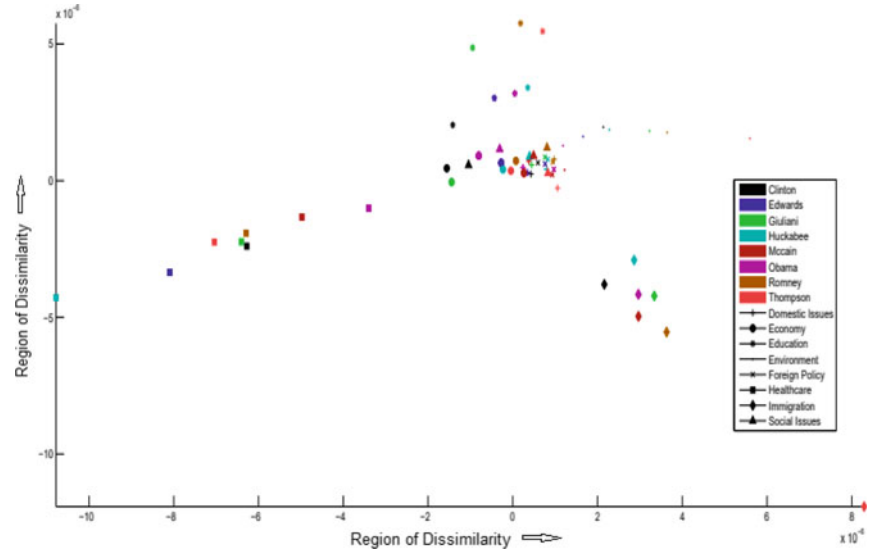


Fig. 3 Visualization results using an RBF kernel with Semidefinite Embedding

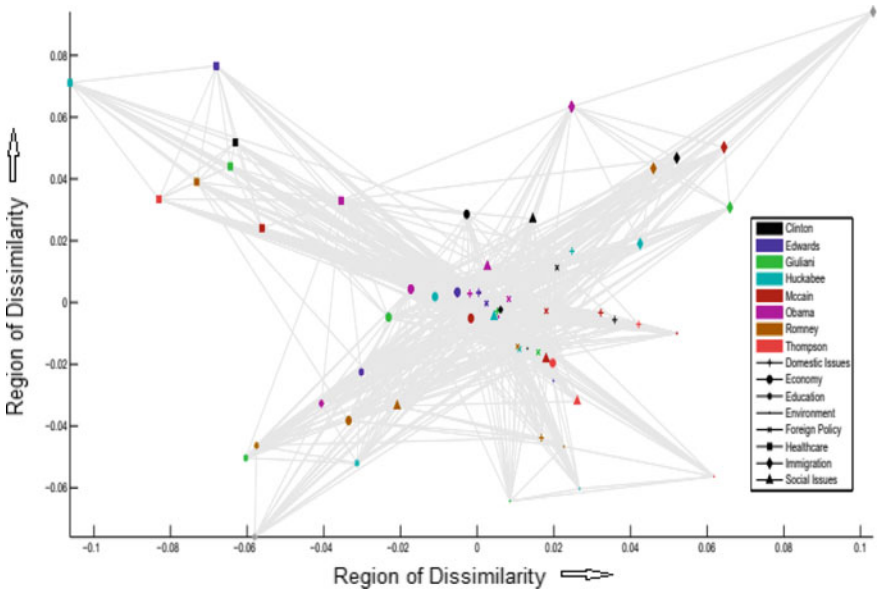


Fig. 4 Visualization results using an RBF kernel with Minimum Volume Embedding

References

1. Yang Y, Pedersen J (1997) A comparative study on feature selection in text categorization: In: International conference on machine learning
2. Mohammed M, Al-Khateeb B, Rashid AN, Ibrahim DA, Ghani MKA, Mostafa SA (2018) Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. Comput Electric Eng 70:871–882
3. Allan J, Leouski A, Swan R (1997) Interactive cluster visualization for information retrieval: technical report IR-116, Center for Intelligent Information Retrieval, University of Massachusetts Amherst Massachusetts
4. Huang X, Wu L, Ye Y (2019) A review on dimensionality reduction techniques. Int J Patt Recogn Artific Intell 33(10)
5. Carey M, Heesch DC, Ruger SM (2003) Info navigator- a visualization tool for document searching and browsing. Distributed Multimedia Systems
6. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features: In Proceedings of ECML-98, 10th European conference on machine learning, Springer, Berlin Heidelberg, pp 137–142
7. Lesort T, Díaz-Rodríguez N, Goudou J-F, Filliat D (2018) State representation learning for control-An overview. Neural Networks Elsevier 108:379–392
8. Scholkopf B, Smola AJ, Muller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput, 1299–1319
9. Kirby M, Sirovich L (1998) Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12:103–108
10. Lewis D, Ringuette M (1994) A comparison of two learning algorithms for text categorization. In: Third annual symposium on document analysis and information retrieval, pp 81–93
11. Shaw B, Jebara T (2007) Minimum volume embedding: In Eleventh international conference on artificial intelligence and statistics, pp 460–467

12. Yang Y, Wilbur W (1996) Using corpus statistics to remove redundant words in text categorization. J Am Soc Inf Sci 47:357–369
13. Weinberger KQ, Sha F, Saul LK (2004) Learning a kernel matrix for nonlinear dimensionality reduction. In: Proceedings of the twenty first international conference on machine learning (ICML-04), Banff Canada, pp 839–846
14. Chormunge S, Jena S (2018) Correlation-based feature selection with clustering for high dimensional data. J Electric Syst Inf Technol 5(3):542–549
15. He X, Niyogi P (2004) Locality preserving projections. In: Proceedings NIPS
16. Belkin M, Niyogi P (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings NIPS
17. Deniz O, Pedraza A, Vallez N, Salido J, Bueno G (2020) Robustness to adversarial examples can be improved with overfitting. Int J Mach Learn Cybernetics, 1–10
18. Zheng A, Casari A (2018) Feature engineering for machine learning. In: Principles and techniques for data scientists. O"Reilly Media, Newton Massachusetts USA
19. Juvonen A, Sipola T, Hämäläinen T (2015) Online anomaly detection using dimensionality reduction techniques for http log analysis. Comput Netw 91:46–56
20. Lewis D, Schapire RE, Callan JP, Papka R (1996) Training algorithms for linear text classifiers. In: Proceedings of the 19th annual international ACM SIGIR conference on research and development in information retrieval, 298–306

Experimental Analysis on Aging Evaluation of Transformer Oil via Data Collection

Harkamal Deep Singh and Jashandeep Singh

Abstract Transformer aging is the most significant feature, and it remains demand for utilities to deal with a larger count of aging transformers. This leads to the requirement for scrutinizing the working conditions of the transformer. One of the most extensively deployed techniques is oil testing and sampling. Databases of oil test reports were thus marked as a greater resource of information for enabling the assessment of transformer aging and for managing assets. Here, this work concerns collecting the data from 20 diverse power transformers from PSEB situated at diverse sites in Punjab. Moreover, the data like locations, constructed authorities, phases, aging, serial no., HV/LV in KV, manufacturing year, installation date, and MVA rating for 20 transformers are specified in this work. With the collected data, further analysis can be made on the aging of insulation oil in transformers.

Keywords Transformer aging · Insulation oil · Deep learning · Machine learning · DP values

Nomenclature

Abbreviations	Descriptions
ANNs	Artificial Neural Networks
AI	Artificial Intelligence
BP	Back Propagation
BDV	Breakdown Voltage
DL	Deep Learning
DDF	Dielectric Dissipation Factor

H. D. Singh (✉)
Punjab Technical University, Kapurthala, India
e-mail: harkamaldeepsingh@gmail.com

J. Singh
Bhai Gurdas Institute of Engineering and Technology, Main Patiala Road, Sangrur, Punjab 148001, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_30

DCNN	Deep Convolutional Neural Network
DBN	Deep Belief Network
DP	Degree Of Polymerization
DNN	Deep Neural Network
FCM	Fuzzy C Means
FDS	Frequency Domain Spectroscopy
IFT	Inter Facial Tensions
LMM	Levenberg–Marquard and Momentum
ML	Machine Learning
MC	Moisture Content
NIS	National And International Standards
PSEB	Punjab State Electricity Board
PT	Power Transformer

1 Introduction

Nowadays, the worldwide rising concern is the aging of power transformer as the requirements for power systems, and power quality turns out to be significant for increasing the life span [1–3]. Transformer insulation using oil-cellulose gets aged irreparably, and the regeneration of cellulose insulation was not feasible. The noteworthy aspects that lessen the life span of the transformers were the associated reduction of insulation age and mechanical strength. Furans are said to be the by-products of aging transformer paper that tend to dissolve in transformer oils [4, 5]. The features are almost related to the age property of the insulating paper. Therefore, the probable insulation hazards are identified by evaluating the age of insulating paper, and in addition, it offers the essential methods for maintaining power equipments [6–8].

However, the investigational time is reduced by carrying out analysis at a temperature beyond 100 °C [9, 10]. Proficient insulation is essential for the reliability of the safe function and power supply of the power equipments. The aging nature is approximated with larger DP values regarding lesser 2-FAL in transformer field oil [11–13]. These values could not recognize the aging nature and evaluation condition of hot-spot insulation. The features, which reduce the insulation characteristics, are the MC and aging [14–17]. Therefore, the diagnosis of aging status remains a motivating research area.

The current technology has recognized the non-destructive assessment in transformer insulation [18, 19]. In the previous decades, the majority of the analysis portrayed the proficient constraints using dielectric spectroscopy for revealing the aging on FDS [20, 21]. It is much complex to discriminate the effects of 2 factors, and therefore, the computed MC and accurateness of the aging state are damaged.

The noteworthy contribution of this work is as follows.

- To carry out an experimental data collection process from diverse 20 transformers, which are located at various locations in Punjab for evaluating the aging of oil in PT.

Contribution:

- To exploiting the gathered data on the aging of insulation oil in PTs.
- The effect of DL and ML on the aging of PT was also discussed in this work.

The work is arranged as: Section 2 makes reviews on the concerned topic. The description of the experimental data collection process for the aging evaluation of PT insulation oil is elaborated in Sect. 3. Explanation of varied parameters is described in Sect. 4. The effect of deep learning and machine learning on aging assessment is illustrated in Sect. 5, and the work is concluded in 4.

2 Literature Review

2.1 Related Works

In 2019, Li et al. [22] developed a self-learning approach for computing the DP of insulating paper based on various oil aging constraints for transformers. The suggested method included linear regression, FCM, and data processing. Further, the resultants exposed that evaluating the DP values based on numerous aging constraints was more accurate than evaluating with a particular constraint. The developed technique also promoted the augmentation of more precise life valuation schemes for PTs based on oil age constraints.

In 2021, Zhang et al. [23] introduced a non-destructive approach by deploying multivariate chemical indicators for computing the leftover availability of cellulosic paper insulation. The chosen chemical indicators closely related to the insulation paper aging were examined. Moreover, by employing the objective and subjective weighting scheme, the evaluated data of all indicators were completely derived to accomplish estimation weights.

In 2019, Mingze et al. [24] have established a technique for quantitative valuation of insulation aging based on the Davidson-Cole method. Accordingly, the constraints of the Davidson-Cole scheme were derived by fitting the model. The investigational outcomes have exposed that the developed constraints were directly related to the aging period.

In 2021, Carmela et al. [25] conducted an investigational analysis on mechanical and tensile properties of 2 types of paper, crepe and plain kraft papers. These papers were typically employed in power transformers for insulation, in varied aged states. These data were further processed for several purposes. They enhanced the current knowledge on degradation procedure and mechanical effect of cellulosic insulation and gave different values for reference that were computed over others.

In 2021, Liu et al. [26] have examined the aging valuation model based on furfural marker for transformer insulation in diverse pressboard/oil ratios. At first, pressboard/oil insulation systems with various pressboard/oil mass proportions were performed for whole changes in the oil. Consequently, the customized furfural-DP method was expanded to diverse pressboard/oil mass proportions under different oil changes. The developed method has promoted the precision of the aging valuation under varied oil change conditions and pressboard/oil mass ratios.

3 Description of Experimental Data Collection Process for Aging Evaluation of PT Insulation Oil

3.1 Details of PTs

For analyzing the age of PT insulation oil, this work carries out experimentation on twenty diverse PTs from PSEB positioned at different locations in Punjab. The specification of PT from which the experimentation is done is stated in Tables 1 and 2, in that order for every 10 transformers. The data such as locations, constructed authorities, serial no., phases, aging, manufacturing year, installation date, MVA rating, and HV/LV in KV are specified in the tables. In addition, the mineral oil was selected depending upon IS: 335–1993 (2005). The above said constraints were evaluated using various NIS [27].

4 Explanation on Varied Parameters

4.1 BDV

BDV [27], a noteworthy constraint aids in estimating the oil state precisely. The soil constituent and MC in oil are portrayed by the BDV metric. When BDV is higher, then it points out clean and dry oil. Temperature, density, and certain features cause complexity in evaluating and modeling the conduction and breakdown schemes. For measuring the oil's dielectric strengths, an oil breakdown experimentation set is deployed. Following all breakdowns, the oil is mixed to avoid the electrode from carbon particles and to avoid the creation of an air bubble. The graph illustration of BDV revealing the aging effect is exposed in Fig. 1. It symbolizes that BDV gradually decreases and endures the nonlinear association with age. Moreover, it also reduces the MC, gets the contamination as a product on oxidizing the oil, and may enhance the density and size of created free particles.

Table 1 Details of power transformers no. 1–10

Transformer No	1	2	3	4	5	6	7	8	9	10
Constructed By	BBL	NUCON	BBL	ECE	NUCON	BBL	IMP	NUCON	TAL	BBL
Serial No	5425/17	64,503	5317/12	84,799	64,505	4656/2	PT.7004	6480	S00000258	5390/10
Location	Bathinda, Phul Dist	Bathinda, Balianwali Dist	Bathinda, Nathana Dist	Bathinda, Nathana Dist	Bathinda, BhuchoKalan Dist	Mohali Phase 7	Kurali	Bathinda, Rampura Dist	Bathinda, Dhan Singh Khana Dist	Kurali
MVA Rating	20	20	20	20	20	20	20	20	20	20
Phases	3	3	3	3	3	3	3	3	3	3
HV/LV in KV	66/11	66/11	66/11	66/11	66/11	66/11	66/11	66/11	66/11	66/11
Aging (Days)	1657	1369	1474	5764	1270	5471	2394	919	1940	1820
Date of Installation	21–11-2014	24–08-2015	13–08-2015	14–11-2003	08–12-2015	23–04-2004	27–10-2012	24–11-2016	26–03-2014	24–05-2014
Year of Manufacturing	2014	2015	2012	2003	2015	2004	2012	2016	2014	2014

Table 2 Details of power tranformers no. 11–20

Transformer No	11	12	13	14	15	16	17	18	19	20
Constructed By	ECE	ECE	T.A	T.A	CTR	ECE	ECE	ECE	TA	ECE
Serial no	8536	85,950	2,000,030	TA2000093	1,220,642	86,274	86,370	84,807	TA2000025	86,424
Location	Bathinda, Phul Dist	Bathinda, Rampura Dist	Bathinda, Rampura Dist	Ropar	Kurali	Bathinda, Balianwali Dist	Mohali Phase-8b	Mohali Phase-8b	Ropar	Mohali Phase 7
MVA rating	20	20	20	20	20	20	20	20	20	20
Phases	3	3	3	3	3	3	3	3	3	3
HV/LV in KV	66/11	66/11	66/11	132/11	66/11	66/11	66/11	66/11	132/11	66/11
Aging (Days)	5564	4330	3354	2638	2394	3360	2491	5376	3083	2806
Date of installation	11–03-2004	24–07-2007	26–03-2010	26–06-2011	27–10-2012	25–05-2010	20–03-2012	26–04-2004	07–04-2010	10–08-2011
Year of manufacturing	2004	2006	2009	2010	2012	2009	2012	2004	2009	2011

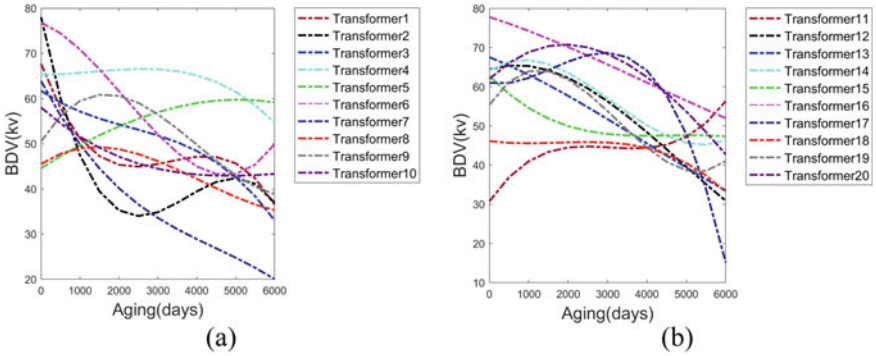


Fig. 1 Graphical illustration revealing the impact of aging on BDV of **a** PTs no.1–10 **b** PTs no. 11–20

4.2 Moisture

The MC[27] in oil is unnecessary since it impacts solid insulation and oil's dielectric properties in transformers. If PT is packed with oil, it causes an impact on insulation property and diminishes its life span. Karl Fischer instrument was deployed for defining the MC, as it is flexible, and accordingly, the resultants are evaluated in ppm. The "Karl Fischer tool, (model MA-101 B, spectra lab)" was utilized for determining the MC. This tool is an adaptable appliance, totally microprocessor oriented and modeled to compute the MC in the oil sample. Figure 2 illustrates the impact of MC on PT aging. In Fig. 2a, PT 9 includes higher MC, and in Fig. 2b, PT 17 includes a high ppm level since the compound stress on PT moisture insulation enhances the aging continually. As stated above, the MC is a life-reduction constraint, and it affects the insulation systems and reduces the overload capacity, decreases the strength, and accelerates the insulating deteriorating procedure.

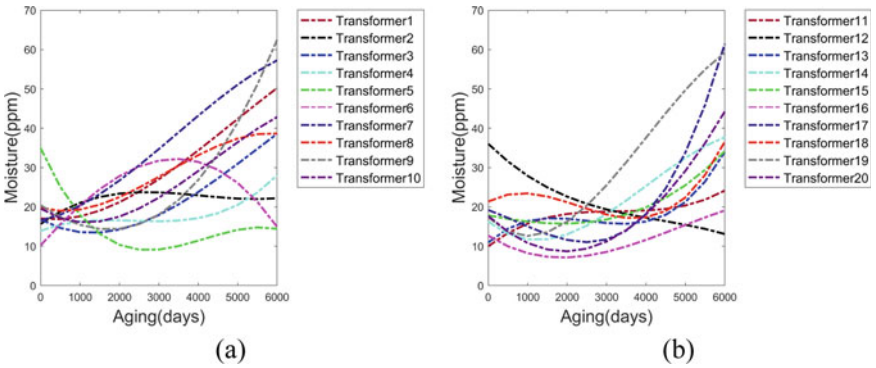


Fig. 2 Graphical illustration revealing the impact of aging on MC of (**a**), **a** PTs no.1–10, **b** PTs no.11–20

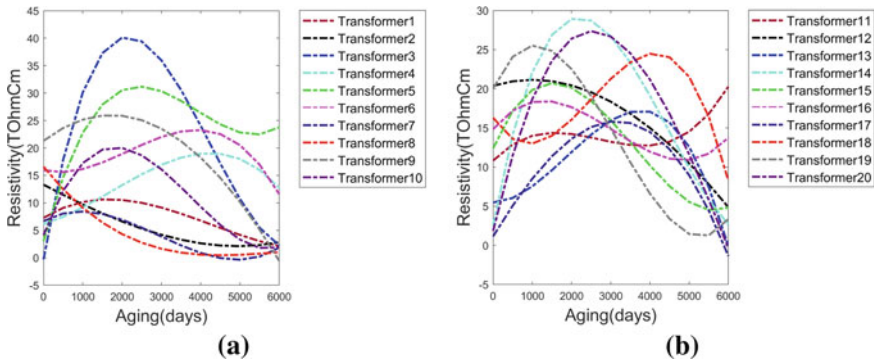


Fig. 3 Graphical illustration revealing the impact of aging on the resistivity of **a** PTs no.1–10 **b** PTs no.11–20

4.3 Resistivity

The oil resistivity has to be higher as feasible for revealing the most excellent performances. Figure 3 illustrates the resistivity effect on aging for PTs. In Fig. 3a, the resistivity of PT 3 is higher and for PT 7 is lesser. Likewise in Fig. 3b, PT 19 includes lesser resistivity while PT 14 includes higher resistivity. The higher resistivity reproduces the ion-forming particles and free ions in lesser content.

4.4 DDF

DDF is also called as tan delta with a value of 0.01 may include excellent oil quality. The effect of age on DDF [27] is symbolized in Fig. 4. Figure 4a illustrates DDF of PT 9 is lesser, and PT 10 is higher. In Fig. 4b, the DDF of PT 20 encompasses more, and PT 19 encompasses less. If the tan delta value is higher, then a huge amount of contaminants exists. It is related to resistivity. The power dissipation in a dielectric loss in a sporadic region is shown in Eq. (1), wherein it denotes that power loss is directly relative to tan delta loss, and it is said to be a loss factor.

$$\text{Pow} = \omega(\tan\delta)\text{Vo}^2 \tag{1}$$

4.5 Interfacial Tension

This constraint is essential for dividing the surface of oil available in oil–water interfaces, which was exploited for calculating the contamination degree and purifying

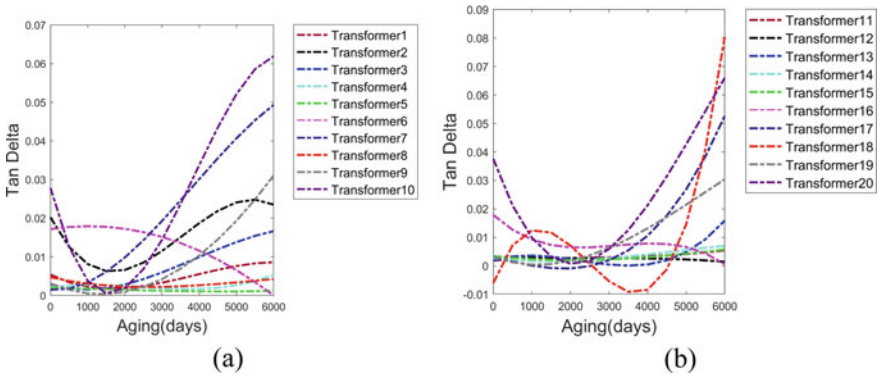


Fig. 4 Graphical illustration revealing the impact of aging on DDF of **a** PTs no.1–10 **b** PTs no.11–20

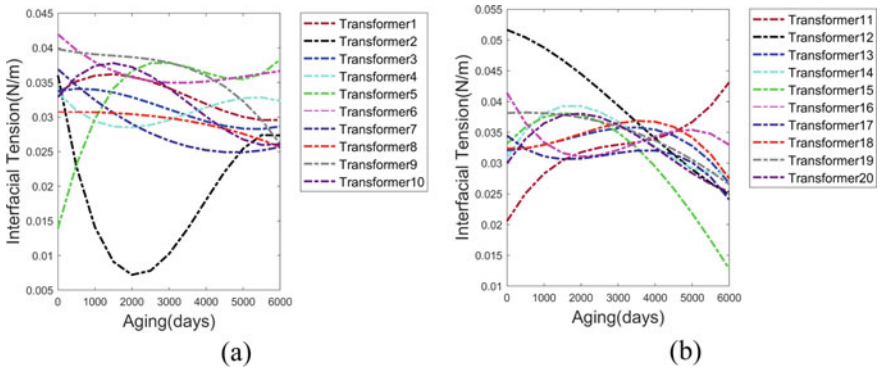


Fig. 5 Graphical illustration revealing the impact of aging on interfacial tensions of **a** PTs no.1–10 **b** PTs no.11–20

the oil. IFT [27] is used for defining the subsistence of polar pollutants and oil decomposed products. If oxidation gets improved, then there is a decrease in IFT since high resemblances exist among oil and water molecules. Figure 5 demonstrates the effect of IFT aging. In Fig. 5a, PT 2 exhibits lower IFT, and PT 9 provides higher IFT. Figure 5b demonstrates lower IFT for PT 15 and higher for PT 12.

4.6 Flashpoint

It is desirable that oil should encompass intense flash points [27]. Owing to the joint stress effect, the flash point diminishes with age. In Fig. 6a, PT 10 includes a lower flash point. PT 12 includes a lower flash point in Fig. 6b.

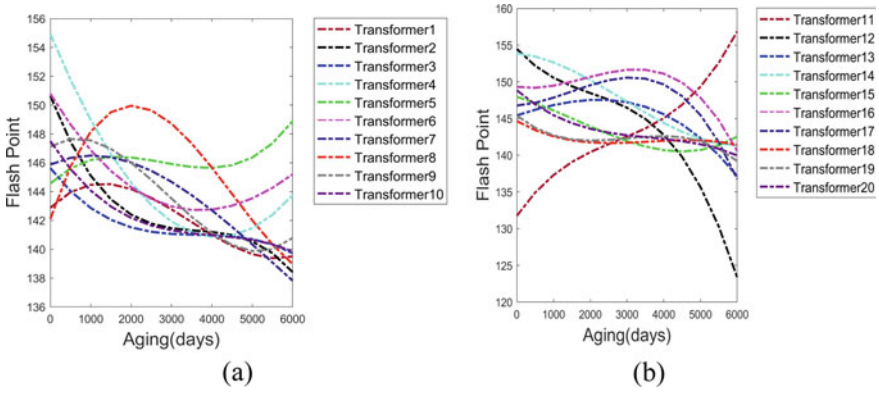


Fig. 6 Graphical illustration revealing the impact of aging on flashpoints of **a** PTs no.1–10 **b** PTs no.11–20

5 Effect of Deep Learning and Machine Learning on Aging Assessment

The accurate assessment of insulation oil aging is necessary for PT since a misconception might lead to transformer spoil, thus causing workers to deadly disasters, massive financial losses, shut down of network, and the exclusion of networks from service [1, 26]. Therefore, intellectual schemes and their appliances are essential to be investigated for storing human experiences and for reacting wisely and automatically to assist inexpert researchers in taking correct decisions. In addition, this intellect could be included in a general online system for scrutinizing systems. Several analyses exist in the literature regarding the appliance of intellectual schemes for the diagnosis and monitoring of PTs. In various analyses, intellectual schemes were deployed for interpreting suspended gas analysis data or for predicting PT health index states. Nevertheless, only, certain works focus on the issue of PT oil quality evaluation via DL and ML schemes. Analysts have examined the performances of ANNs, particularly BP scheme strained using fuzzy [28] logic and LMM scheme for evaluating the oil quality of PT [7, 9]. In addition, preprocessing of data was a significant stage in the AI process, which could offer further information for improved prediction. As a result, the analysts deployed the raw data for developing prediction schemes. The quantity of data exploited for substantiation was restricted to fewer cases that were inadequate for simplification [2, 10]. Moreover, the main shortcoming of ANN-oriented schemes is that they necessitate a larger dataset for appropriate training to assure a reliable analysis. In addition, fuzzy logic does not handle the computed data directly, and the extraction of their rules might be complex. In recent times, with a rise in the capability of computers for processing data, varied DL-based schemes like DCNN, DNN, DBN, and so on were introduced for assessing the conditions of power equipment (Fig. 7).

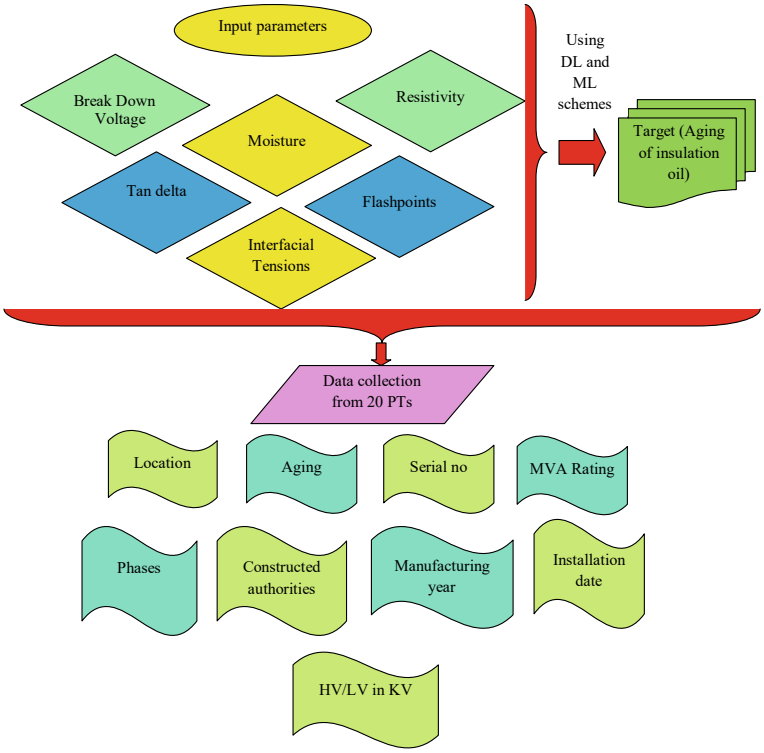


Fig. 7 Block representation of experimental data collection for proposed aging evaluation of oil in PT

6 Conclusion

This work concerns gathering the data from 20 varied PTs from PSEB situated at diverse locations in Punjab. Furthermore, the data such as locations, constructed authorities, phases, aging, serial no., HV/LV in KV, manufacturing year, installation date, and MVA rating for 20 PTs were specified in this work. By exploiting the gathered data, further analysis could be performed on the aging of insulation oil in PTs. In addition, the effect of DL and ML on the aging of PT was also discussed in this work. Here, this work concerns collecting the data from only 20 power transformers from PSEB situated at diverse sites in Punjab, but there is a need of collecting more samples for adopting the deep learning models. In future work, by combining sampling data obtained from actual transformers, the capacity of the evaluation database will be expanded. The performance evaluation of gathered data will be conducted in the future.

References

1. Sun W, Yang L, Cheng Z (2019) Improved method for aging assessment of winding hot-spot insulation of transformer based on the 2-FAL concentration in oil. Int J Electric Power Energy Syst 112:191–198
2. Matharage SY, Liu Q, Wang ZD (2016) Aging assessment of kraft paper insulation through methanol in oil measurement. Trans Dielectrics Electric Insulation 23(3):1589–1596
3. Gouda OE, Dein AZE (2019) Prediction of aged transformer oil and paper insulation. J Electric Power Components Syst 47(4–5)
4. Yang L, Zou T, Deng B, Zhang H, Mo Y, Peng P (2019) Assessment of oil-paper insulation aging using frequency domain spectroscopy and moisture equilibrium curves. IEEE Access 7:45670–45678
5. Liu J, Fan X, Zheng H, Zhang Y, Zhang C, Lai B, Wang J, Ren G, Zhang E (2019) Aging condition assessment of transformer oil-immersed cellulosic insulation based upon the average activation energy method. Cellulose 26(6):3891–3908
6. Gao J, Yang L, Wang X, Liu Y, Lv Y, Zheng H (2016) Condition diagnosis of transformer oil-paper insulation using dielectric response fingerprint characteristics. Trans Dielectrics Electric Insulation 23(2):1207–1218
7. Marín-Serrano A, Balderas-López JA, Calva PA, Aranda-Pérez A (2019) Thermo-optical properties as complementary parameters for damage assessment of mineral oils aged under controlled conditions used in power transformers. Thermochim Acta 676:33–38
8. Lin Y, Wei C, Tao F, Li J (2019) Aging assessment of oil-paper insulation of power equipment with furfural analysis based on furfural generation and partitioning. IEEE Trans Power Delivery 34(4):1626–1633
9. Khan I, Abid MA, Ullah K et al (2021) Multi-aging effects on vegetable based oils for transformer insulation in HV systems. J Electr Eng Technol 16:2709–2720. https://doi.org/10.1007/s42835-021-00795-5
10. Liu J, Lv J, Zhang M, Jia H, Zhuang W, Chen X (2020) Research on Life prediction of oil-paper insulation based on dielectric response. Electric Power Components Syst 48(3):281–290. https://doi.org/10.1080/15325008.2020.1758839
11. Lin Y, Liao R, Tao F, Wei C (2019) Effects of moisture on furfural partitioning in oil-paper insulation system and aging assessment of power transformers. Electric Power Components Syst 47(1–2):192–199. https://doi.org/10.1080/15325008.2019.1565135
12. Thabet A, Allam M, Shaaban SA (2019) Assessment of individual and multiple nanoparticles on electrical insulation of power transformers nanofluids. Electric Power Components Syst 47(4–5):420–430. https://doi.org/10.1080/15325008.2019.1609624
13. Maharana M, Baruah N, Nayak SK, Meher N, Iyer PK (2019) Condition assessment of aged ester-based nanofluid through physicochemical and spectroscopic measurement. IEEE Trans Instrument Measure 68(12):4853–4863. https://doi.org/10.1109/TIM.2019.2900883
14. Fan X, Yang S, Benhui L, Liu J, Zhang Y, Wang Z (2020) Normalization for FDS of transformer insulation considering the synergistic effect generated by temperature and moisture. IEEE Access 8:202013–202021. https://doi.org/10.1109/ACCESS.2020.3036418
15. Jacob ND, Kordi B, Sherif SS (2020) Assessment of power transformer paper ageing using wavelet texture analysis of microscopy images. IEEE Trans Dielectr Electr Insul 27(6):1898–1905. https://doi.org/10.1109/TDEI.2020.009086
16. Song R, Chen W, Yang D, Shi H, Zhang R, Wang Z (2021) Aging assessment of oil–paper insulation based on visional recognition of the dimensional expanded Raman spectra. IEEE Trans Instrument Measure 70:1–10, Art no. 6007110. https://doi.org/10.1109/TIM.2021.3075525
17. Matharage SY, Liu Q, Wang ZD, Wilson G, Krause C (2018) Aging assessment of synthetic ester impregnated thermally non-upgraded kraft paper through chemical markers in oil. IEEE Trans Dielectrics Electric Insulation 25(2):507–515. https://doi.org/10.1109/TDEI.2018.006833

18. Rao UM, Jarial R (2019) Measurement of transformer solid insulation degradation using dilatometry and X-ray diffraction analysis. Measurement 131:701–705. https://doi.org/10.1016/j.measurement.2018.09.024
19. Samarasinghe WMSC, Kumara JRSS, Fernando MARM, Gunawardena AUAW (2017) Aging assessment of transformer pressboard insulation by microstrip ring resonator at GHz frequencies. Trans Dielectrics Electric Insulation 24(3):1923–1930
20. Moodley N, Gaunt CT (2017) Low energy degradation triangle for power transformer health assessment. IEEE Trans Dielectr Electr Insul 24(1):639–646. https://doi.org/10.1109/TDEI.2016.006042
21. Mishra D, Haque N, Baral A, Chakravorti S (2017) Assessment of interfacial charge accumulation in oil-paper interface in transformer insulation from polarization-depolarization current measurements. IEEE Trans Dielectr Electr Insul 24(3):1665–1673. https://doi.org/10.1109/TDEI.2017.006525
22. Li S, Ge Z, Abu-Siada A, Yang L, Li S, Wakimoto K (2019) A new technique to estimate the degree of polymerization of insulation paper using multiple aging parameters of transformer oil. IEEE Access 7:157471–157479. https://doi.org/10.1109/ACCESS.2019.2949580
23. Zhang E, Zheng H, Zhang C et al (2021) Aging state assessment of transformer cellulosic paper insulation using multivariate chemical indicators. Cellulose 28:2445–2460. https://doi.org/10.1007/s10570-021-03683-3
24. Zhang M, Liu J, Yin M, Jia H, Lv J (2019) Assessment on oil-paper insulation aging of transformer based on dielectric response model electric power. Components Syst 47(13):1145–1155. https://doi.org/10.1080/15325008.2019.1663454
25. Oria C, Carrascal I, Ortiz A (2021) Experimental dataset on the tensile and compressive mechanical properties of plain Kraft and crepe papers used as insulation in power transformers after ageing in mineral oil. Data in Brief 36 (Cover date: June 2021):107031
26. Liu J, Zhang H, Geng C, Fan X, Zhang Y (2021) Aging assessment model of transformer insulation based on furfural indicator under different oil/pressboard ratios and oil change. IEEE Trans Dielectr Electr Insul 28(3):1061–1069. https://doi.org/10.1109/TDEI.2021.009530
27. Singh J, Sood YR, Verma P (2012) The influence of service aging on transformer insulating oil parameters. IEEE Trans Dielectr Electr Insul 19(2):421–426
28. Malhotra J, Bakal J (2018) Grey wolf optimization based clustering of hybrid fingerprint for efficient de-duplication. Multiagent Grid Syst 14(2):145–160

A Systematic Survey of Automatic Text Summarization Using Deep Learning Techniques

Madhuri Yadav and Rahul Katarya

Abstract Today, with the advancement of web technology and the accessibility of the Internet, one can now share their opinion almost everywhere. The Internet user is now not limited to consume the information; instead, they are now information producer. This evolution has resulted in an enormous quantity of data produced over social media and other platforms. This overloaded information makes it complicated for the user to extract and summarize this content. Thus, text summarization is required to generate a shorter version of information to handle the complexity of data. This need leads to exposure to various summarization techniques. This paper gives an overview to the summarization process and studies recent advancements in deep learning-based summarization. Dataset and summary evaluation methodologies are discussed in the area of text summarization. Finally, we conclude the study with various observations, challenges, and future scope in automatic text summarization. This paper is a contribution in providing help to researchers to better understand summarization and find new opportunities in the field of summarization.

Keywords Abstractive summarization · Automatic text summarization · Extractive summarization · Deep learning · Natural language processing

1 Introduction

The web is becoming a repository of data as every user is generating data, resulting in a significant boost in the amount of information available over the Internet. Because of the ease of access to the Internet, there has been tremendous growth in Internet users in the past ten years. Another reason for data generation over the web is due to social media platforms. Due to this overflow of data, a lot of effort is wasted to extract useful information. Not everybody has sufficient time to read all the information available to find the desired information. One solution to this task is to summarize the text or data which is to be read. But manual summarization also becomes difficult and

M. Yadav (✉) · R. Katarya
Department of Computer Science Engineering, Delhi Technological University, New Delhi, India
e-mail: madhuriyadav.me@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_31

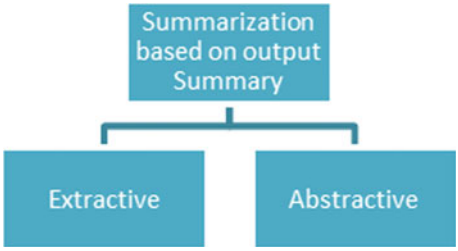


Fig. 1 Summary classification based on type of output summary

time-consuming when the text is large. Automatic summarization is the task which produces a distilled version of the text where important information is captured from the document. Automatic text summarization is an activity where extraction of significant information is done from one or more sources to produce a condensed version for particular tasks [1]. Basically, text summarization is obtaining the salient features of the input document and producing an output document. Automatic text summarization is divided on the basis of type of output summary: extractive and abstractive (Fig. 1).

The primary focus of extractive summarization is on finding out the salient paragraph and significant sentences that altogether represent the precise summary of the document [2]. The extractive summarization technique selects essential sentences, paragraphs, etc., and concatenates them to make a smaller and distilled version of the source document. The level of importance of sentences is determined on the basis of various statistical and linguistic features of sentences. The abstractive summarization approach, contrarily, understands the text and retells the text in lesser sentences as compared to the source text. For an abstractive summary generation, linguistics methods are used to understand and redefine the text [3].

The major contributions of this work are:

1. This work gives a clear understanding of the summarization process to the reader.
2. The latest work in deep learning is discussed which gives an insight into current progress in the summarization field.
3. Challenges of summarization are discussed for the modern web, which can give a direction to the reader to explore the area of summarization.

2 Automatic Text Summarization

The web is filled with information, and to handle and understand this gigantic data, we need some methods that generate a summary and tell us the important information. Generating a good summary needs a good understanding of the qualities of a good summary [4, 5]. The attributes of a good summary are given in Fig. 2.

Fig. 2 Qualities of a Good Summary

Fig. 3 Basic summarization process

The Summarization Process

The basic architecture of the summarization process can be given as follows in Fig. 3. It contains pre-processing, summarization, and post-processing [6].

i. Pre-processing: This is an important task that generates a structured representation of the source document by applying various linguistic techniques such as sentence segmentation, tokenization, stop word removal, part-of-speech tagging, and stemming [7].
ii. Summarization: In this step, the existing text summarization techniques are applied for generating the summary of the source document. The summarization approaches are discussed in 2.1.
iii. Post-processing: The generated summary may have some structural issues; to make the summary more fluent and structured, we must reorder the sentences. Such tasks are done post-processing [7].

Based on the type of output summary, the ATS approaches are divided as extractive and abstractive summarization.

2.1 Extractive Summarization

While summarizing a document, our main focus is to generate the summary that describes the overall conclusion of the document. For this task, extractive summarization selects those sentences or paragraphs that properly describe the document's importance in precise form. Because of the approach of directly choosing the informative sentences, the method of extractive summarization is easier in comparison with the other techniques. The importance of the sentences is dependent upon the linguistic features [2]. Extractive text summarization can be divided into the following independent tasks [8].

2.2 Abstractive Summarization

Produces a summary that includes the sentences and phrases that do not belong to the input text but holds a similar meaning as to the original text. Abstractive text summarization tries to locate the significant and relevant features of the text and produces a summary by use of phrases/words that may/may not be in the original text [9].

2.3 Evaluation Techniques

Despite the fact that data should be compressed from the huge data available, it is also essential to evaluate the generated summary. The summary evaluation before the automatic evaluation was typically manual. The summaries were manually judged on some specific measures.

Informativeness
It refers to the measure of information gain from the summary generated. To measure informativeness, various methods are proposed. The authors [10] divided these measures into questionnaire-based and overlap-based matrices: Questionnaire-based summary evaluation is based on a questionnaire set that is designed from the original documents. Overlap-based matrices for summary evaluation are based on the similarity of the generated summary to the reference summary.

There are three common criteria for proficiency evaluation of a summarization system [10]: Recall, Precision, given by Eqs. 1 and 2, respectively:

$$R = \frac{|S \cup C|}{|S|} \tag{1}$$

$$P = \frac{|S \cap C|}{|C|} \tag{2}$$

where S and C are the collection of sentences/terms in reference summary and candidate summary, respectively. Another metric that is used for summary evaluation is ROUGE(Recall-Oriented Understudy for Gisting Evaluation) was a summarization task in the Document Understanding Conference (DUC) [11]. ROUGE is based on overlap measures in comparison with the reference summary. The ROUGE matrices have several variants ROUGE-1, ROUGE-2, ROUGE-N, ROUGE-L, ROUGE-S, etc.

Table 1 List of the popular datasets

Dataset	Description	Research done
DUC[1]	The data available here consist of all the previous DUC containing documents and summaries (manually created, automatically created baselines, submitted summaries by participating groups), evaluation results, etc.	[12–14]
TAC[2]	Test data, summary, and evaluation results	[15–17]
Opinosis dataset[3]	This is a public dataset that contains sentences extracted from user reviews on 51 topics collected from Tripadvisor, Edmunds.com, and Amazon.com [18]	[18, 19]
Customer reviews[4]	A public dataset contains customer reviews of the product from Amazon.com, including information about the reviewer, review texts, ratings, etc. This dataset is best suited for aspect-based summarization [20]	[20, 21]
FIRE[5]	Previous year's news data for Indian languages such as Hindi. Bengali, Marathi, and Odia are available	[22]
CNN/Daily mail	It is a dataset collection of around 300 k English news article	[23]

2.4 Datasets

There are plenty of datasets available; some of the most popular are mentioned in Table 1.

3 Deep Learning-Based Automatic Text Summarization

Advancement in deep learning methods has been beneficial for artificial intelligence and NLP. Here, we study recent development in deep learning techniques in the summarization field. We divide this section into two subsections: extractive and abstractive text summarization using deep learning.

[1] https://duc.nist.gov/data.html.

[2] https://tac.nist.gov/data/index.html.

[3] http://kavita-ganesan.com/opinosis-opinion-dataset/#.XnCMuKgzZPY.

[4] https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html.

[5] http://fire.irsi.res.in/fire/static/data.

3.1 Deep Learning-Based Extractive Summarization

Extractive summarization is basically a method where we try the sentence with some technique and then choose the most significant sentences to make the output sentence. In [24], the authors did a single document summarization using a hierarchical sentence encoder and attention-based extractor. The sentence extraction is done using supervised training. The authors used a word extractor to minimize the redundancy in the output. The dataset used was DUC2002 and Daily Mail. The sentence representation was done using a hierarchical document reader using convolutional neural network, and the selection of sentences was done by the attention-based extractor. The authors of [25] proposed a method called PriorSum, which is a multi-document summarizer which is a neural network-based technique that captures the important aspect with considering the context. Sentence ranking is based on improved CNN by obtaining context-independent features. For this work, DUC 2001, DUC 2002, and DUC 2004 were used. [23] Introduced extractive summarization as a problem of sequence classification. Here, the text is traversed sequentially to check whether the sentence should be included or not based on previous decisions made for sentences. Here, the authors use RNN-based classifier. The dataset used is DUC2002and Daily Mail.

Singh et al. [26] Introduced a bilingual (Hindi and English) multi-document summarization approach. Here, unsupervised deep learning was used. They used Restricted Boltzmann Machine. The pre-processing steps include segmentation, tokenization, stop word removal, part-of-speech tagging, and feature selection. These were done using TF-IST. [27] In this work, authors used unsupervised deep learning for single document query-focused summarization. Here, the authors used deep autoencoder for the summarization process. The model was divided into two tasks pre-training and fine-tuning. A noisy autoencoder was used for sentence representation and ranking. The sentence selection was done by applying cosine similarity on the concept vector which was generated using the encoder.

3.2 Deep Learning-Based Abstractive Summarization

In [28], the author developed an abstractive summarization approach. This work is about using the deep learning concept to generate the summary. Three encoders were used: a Bag-of-words encoder, CNN, and an attention network for decoding purposes. The author used a neural network language model as a decoder. For word generator, probability distribution was used. The experimentation was done on the Gigaword and DUC datasets. For the optimizer, the stochastic gradient descent was used.

Yao et al. [29] The author developed a deep learning-based summarization approach. Here, instead of using one encoder the authors tried to improve the process by adding a dual encoder. The secondary encoder is other than the regular encoder

models in sentence score based on the history. In the primary encoder, the bi-direction GRU-based RNN was used. For secondary, the unidirectional GRU-based RNN was used. This output is fed into an attention-based decoder. The experiments were performed on CNN/Daily Mail and DUC 2004. Authors of [30] also focused on CNN encoder and RNN decoder. For optimization purposes, SGT was used as used in [28]. In [31], introduced a framework which focused on encoding the keywords to the key information representation, used CNN-Daily Mail dataset encoder was a bidirectional LSTM and decoding was done using LSTM. Authors in [32] proposed a framework fact aware summarization using reinforcement learning to enhance the factual correctness of the summary. In work [33], authors developed an LSTM-based abstractive summarization that focused on finding the key features. For this, the authors divided the process into two phases: where the first one finds the important features and the second phase generated using DL techniques.

Here, we tried to discuss the most recent work done in summarization using deep learning. Despite being this, much research in the area of summarization this field still faces some challenges that are discussed in the next section.

4 Challenges in Text Summarization

Processing the text has been always a challenging task but with emerging times the level of challenges has also increased. Now, the people on social media are talking in their own languages other than English. Processing such text can be a difficult task. Also, the text over the web is not grammatically correct, spellings are wrong, and uses of slang are very common. Processing such text can be very challenging. Producing an abstractive summary is also very challenging as it focuses on conciseness and informativeness both at the same time. In the next section, we discuss some suggestive future work in this area so that these challenges can be tackled.

5 Conclusion and Future Work

Automatic text summarization is a very important task in natural language processing and for the current situation, i.e., web is overloaded with information, summarization becomes very important. In this survey, we discussed the entire process of summarization and approaches of summary generation. This has been concluded that the area of extractive summarization is explored and researched more than abstractive summarization. In this work, we have discussed deep learning-based research in the area of summarization. Supervised methods of extractive and abstractive summarization are explored well, but still challenges are being faced. In the future, we intend to develop a good unsupervised deep learning-based algorithm for text summarization, especially for abstractive summarization.

References

1. Belguith L, Ellouze M, Maaloul M, Jaoua M, Jaoua F, Blache P (2014) Automatic summarization. Natl Lang Process Semitic Lang, 371–408
2. Moratanch N, Chitrakala S (2017) A survey on extractive text summarization. In: 2017 international conference on computer, communication and signal processing (ICCCSP)
3. Dalal V, Malik L (2013) A survey of extractive and abstractive text summarization techniques. In: 2013 6th international conference on emerging trends in engineering and technology
4. Shareghi E, Hassanabadi L (2008) Text summarization with harmony search algorithm-based sentence extraction. In: Proceedings of the 5th international conference on soft computing as transdisciplinary science and technology—CSTST '08
5. Goldstien J, Mittal V, Carbonell J, Kantrowitz M (2000) Multi-document summarization by sentence extraction. In: Proceedings of the 2000 NAACL-ANLPWorkshop on automatic summarization
6. El-Kassas W, Salama C, Rafea A, Mohamed H (2021) automatic text summarization: a comprehensive survey. Expert Syst Appl 165:113679
7. Gupta V, Lehal G (2010) A survey of text summarization extractive techniques. J Emerg Technol Web Intell 2
8. Nenkova A, McKeown K (2012) A survey of text summarization techniques. Mining Text Data 43–76
9. Raphal N, Duwarah H, Daniel P (2018) Survey on abstractive text summarization. In: 2018 international conference on communication and signal processing (ICCSP)
10. Ermakova L, Cossu J, Mothe J (2019) A survey on evaluation of summarization methods. Inf Process Manage 56:1794–1814
11. Lin C (2004) ROUGE: a package for automatic evaluation of summaries. Text Summarization Branches Out 74–81
12. Luo W, Zhuang F, He Q, Shi Z (2013) Exploiting relevance, coverage, and novelty for query-focused multi-document summarization. Knowl-Based Syst 46:33–42
13. Canhasi E, Kononenko I (2014) Weighted archetypal analysis of the multi-element graph for query-focused multi-document summarization. Expert Syst Appl 41:535–543
14. Al-Sabahi K, Zhang Z, Long J, Alwesabi K (2018) An enhanced latent semantic analysis approach for arabic document summarization. Arab J Sci Eng 43:8079–8094
15. Jin F, Huang M, Zhu X (2009) A query-specific opinion summarization system. In: 2009 8th IEEE international conference on cognitive informatics
16. Campr M, Jezek K (2012) Comparative summarization via latent semantic analysis. Latest Trends Inf Technol
17. Jezek K, Campr M (2013) Comparative summarization via latent dirichlet allocation. Dateso 2013:80–86
18. Ganesan K, Zhai C, Han J (2010) Opinosis: graph-based approach to abstractive summarization of highly redundant opinions. In: Proceedings of the 23rd international conference on computational linguistics
19. Bhargava R, Sharma Y, Sharma G (2016) ATSSI: abstractive text summarization using sentiment infusion. Procedia Comput Sci 89:404–411
20. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04
21. Hu M, Liu B (2004) Mining opinion features in customer reviews. In: Proceedings of the 19th national conference on artificial intelligence (AAAI'04)
22. Verma P, Pal S, Om H (2019) A comparative analysis on Hindi and English extractive text summarization. ACM Trans Asian Low-Resource Lang Inf Process 18:1–39
23. Nallapati R, Zhai F, Zhou B (2017) SummaRuNNer: a recurrent neural network based sequence model for extractive summarization of documents. In: AAAI'17: proceedings of the thirty-first AAAI conference on artificial intelligence, 3075–3081

24. Cheng J, Lapata M (2016) Neural summarization by extracting sentences and words. In: Proceedings of the 54th annual meeting of the association for computational linguistics, vol 1: Long Papers
25. Cao Z, Wei F, Li S, Li W, Zhou M, Wang H (2015) Learning summary prior representation for extractive summarization. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing, vol 2: Short Papers
26. Singh S, Kumar A, Mangal A, Singhal S (2016) Bilingual automatic text summarization using unsupervised deep learning. In: 2016 international conference on electrical, electronics, and optimization techniques (ICEEOT)
27. Yousefi-Azar M, Hamey L (2017) Text summarization using unsupervised deep learning. Expert Syst Appl 68:93–105
28. Rush A, Chopra S, Weston J (2015) A neural attention model for abstractive sentence summarization. In: Proceedings of the 2015 conference on empirical methods in natural language processing
29. Yao K, Zhang L, Du D, Luo T, Tao L, Wu Y (2020) Dual encoding for abstractive text summarization. IEEE Trans Cybernetics 50:985–996
30. Chopra S, Auli M, Rush A (2016) Abstractive sentence summarization with attentive recurrent neural networks. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies
31. Li C, Xu W, Li S, Gao S (2018) Guiding generation for abstractive text summarization based on key information guide network. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies
32. Zhang M, Zhou G, Yu W, Liu W (2021) FAR-ASS: fact-aware reinforced abstractive sentence summarization. Inf Process Manage 58:102478
33. Song S, Huang H, Ruan T (2018) Abstractive text summarization using LSTM-CNN based deep learning. Multimedia Tools Appl 78:857–875

An IoT-Based Smart Garbage Collection and Monitoring System

Taranpreet Singh Ruprah, Pramod Dharmadhikari, and Krunal Pawar

Abstract The populated region gives rise to household waste, trash, garbage, dust, etc. The statutory bodies take many measures to keep up with the neatness of the city. Cleaning the trash is the significantly identified with medical problems in accordance with spread of diseases through unhygienic smell in surrounding. The fitness effects of gases depend upon the quantity inhalation. Exposure can produce irritation of the nose and throat and cause lack of urge for food and headache. Can reason eye inflammation, coughing, and loss of odor, vomiting, nausea, headache are short symptoms that aim to risky disease. We planned a model say "IoT-based smart garbage collection and monitoring system," which demonstrates straight forwardly that the garbage bin is filled to a threshold-specific level by the trash and cleaning or purging them is matter of prompt concern as higher priority. The moisture, gas, weight are important parameters considered for measurement. This forestalls the lumping of trash in the side of the road dustbin which winds up giving foul smell and sickness to everyone.

Keywords Sensor dustbin · IoT-based dustbin · Garbage collection · Smart dustbin

1 Introduction

Now a day's, principle difficulty for contamination is rubbish overflow. It makes unhygienic situation for individuals and makes horrible stench across the environmental elements this leads in spreading a few risky infections and human sickness [5]. To stay far away from all such circumstances, the proposed work is to plot an IoT sensor-based dustbin. In proposed framework, precise dustbins are situated all through the town or the campus; the ones dustbins are given a sensor which lets in following the extent and weight of the trash dustbins, and annotable identity might

T. S. Ruprah (✉) · P. Dharmadhikari · K. Pawar
K. E. Society's Rajarambapu Institute of Technology Rajaramnagar, Sangli, Maharashtra, India
e-mail: taranpreet.singh@ritindia.edu

P. Dharmadhikari
e-mail: pramod.dharmadhikari@ritindia.edu

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_32

be allowed every dustbin in the town in order that it isn't always tough to differentiate which trash dustbin is whole. District takes many measures to maintain up with the tidiness of the metropolis. One of that is putting in place dustbins in standard distance for the comfort of public to dispose of factors. Cleaning this trash is significant capacity of area which is straightforwardly identified with medical problems. The proposed work is to advise that the dustbin is filled to a selected stage by means of the trash and cleaning at high priority has to be done. The multiple dustbins are positioned for the duration of the metropolis or the campus; these dustbins are furnished with a sensor which facilitates in monitoring the level and weight of the rubbish packing containers, and a unique identity will be furnished for each dustbin in the city in order that it is simple to pick out which garbage bin is complete. The exceptional sensors like weight sensor, moisture sensor and air quality sensor, and many others can be used to recognize weight and toxic level of trash in dustbin. The records procured through the sensors and dispatched to android software to reveal the vicinity and extraordinary records of dustbin. While the extent and weight of the bin reaches the threshold limit, the device will transmit the analyzing along with the precise identity provided. The status of the bin is accessed through the concerned authorities from their location with the assist of net, and an immediate movement might be taken to replace overflowing boxes with the empty bins. The paper is having literature survey, proposed model, experimental setup, results and conclusions ahead.

2 Literature Survey

Vijaya et al. [1] have clarified concerning how the IoT module is utilized to govern waste and records the level of dustbin. Moreover, the paper gives statistics concerning how the data might be informed to organization. The paper gives facts approximately unique flagging hardware. The equipment used in the dustbin comprises of two sections; one is the recipient/transmitter and sensor. Kumar et al. [2] have clarified approximately IoT primarily based intelligent waste clean management framework which assessments weight of the dustbin utilizing weight sensor. The weight sensor will give the information about weight of dustbin after the trash is delivered. A heap mobile is a type of transducer, explicitly a strength transducer. They convert the electricity like pressure, strain, pressing factor, or pressure into an electrical sign that may be estimated and normalized. Because the electricity carried out to the heap mobile expands, the electric signal adjustments exceedingly. The most famous sorts of burden cell applied are water driven, pneumatic, and strain degree. Bhardwaj et al. [3] have disclosed concerning how to make use of sensor to detect hazardous gases. The gas sensor identifies the toxic gases present within the dustbin. Likewise, the dustbin gives the statistics about harmfulness to the authorized party. The sensor used for identifying an extensive scope of gases, to get her with NH_3, NO, benzene, smoke, and CO_2. This is satisfactory to use in place of work or processing plant. Sensor has excessive affect ability to ammonia, sulfide, and benzene steam, moreover sensitive to smoke and other hurtful gases. Singh et al. [4] have disclosed concerning the way

to give more prominent availability to trash collection. Likewise, the paper offers information about suitable making arrangements for removal of trash. It gives the techniques to removal of waste. The paper offers information about the Internet application and android application. Jabbar et al. [5] related all its home equipment with the IoT primarily based sensor to reveal and manage the equipment with the help of sensor like smoke sensor, movement sensor, darkness sensor, and temperature sensor. Lulla et al. [6] use couples of sensors to come across illegal activity on the property of the proprietor, to be able to report them of the occurrence of an illegal person. The device gives a notice to the person that has intentionally or unintentionally entered the belongings. Patel et al. [7] automatically detect and discover garbage in actual-world pictures in addition to video. They propose new approach with different deep learning algorithms. The YOLOv5M gives the good results for the proposed approach by achieving a mean average precision (mAP@0.5) value of 0.613. Medehal et al. [8] show the powerful tools of IoT to completely automate the process of garbage monitoring using ultrasonic sensors and Node MCU and provide an optimal route for garbage collection using the cluster first-route second ML algorithm. Anitha [9] proposes an idea of a sensor put in the top of the bin to notifying the percentage of waste present at that time. At the point when the garbage will reach to the supreme degree, a notification may be dispatched to the company, and then the process to drain the bin started. It will helps in keeping the city clean. Akshita and Amita [10] propose an idea of a sensor used to sense and send a message at the duration 1500 ms. Message alert is passed to the municipal corporation with the bin location when the bin is near to overflow, or the percentage of gas has reached the threshold limit. The system makes use of AVR family microcontroller, LEDs, a SIM800L GSM/GPRS Module and was powdered by a 12 V transformer. Kumar et al. proposed [11] a smart alert gadget for garbage clearance by way of giving an alert sign to the Internet server for fast cleansing of dustbin with right verification based totally on level of garbage filling. This gadget is aided with Arduino uno and ultrasonic sensor to provide alert to server to identify level of waste filled inside the dustbin. Nirde et al. [12] used alert mechanism through SMS based on GSM module; an alert is generated on dustbin has reached its most stage; further, department can send waste collector automobile to collect garbage. Raju et al. [13] proposed a device. This device is powered by a solar panel present a smart garbage management solution based totally on given that intelligence to waste dirt packing containers, using an IOT with the help of ZigBee, with sensors and modules. Anagha et al. [14] suggested as soon as the trash has passed the set threshold limit value, the record is dispatched to every other node which is Raspberry Pi, and this node prints the respective residence number on to the screen with the assist of socket programming. Bento et al. [15] proposed solution with genetic algorithms to optimize the translation of the waste collection routes, deliberating parameters related to traffic, and trash bin positions. The gadget for optimizing waste collection consists essentially of two functions: The first function is liable for acquiring the facts from the map, named map parsing, and the second one characteristic is the genetic set of rules itself, named route optimization. Sharma et al. [16] proposed the automatic waste segregating and self-sanitizing dustbin with the help of basic IoT sensors. Annie et al. [17] use GSM, GPRS, and sensors to check

the garbage bin level; they used various color LED for showing the bin is full, empty, normal. Abhishek et al. [18] use voice signal to control the dustbin; they used voice modulation with android to send a command message to the robot's microcontroller. Sonali et al. [19] use classification technique of ML to check the correctness of sending alert messages for order to manage the waste.

3 Proposed Model

The proposed model is one of the solution for multiple problems like empty vehicle return, unnecessary rounds of our vehicle, and waste of fuel. The proposed model is having three main components of the system such as dustbin, authorized person, and collector vehicle. When the trash inside the dustbin reaches to certain threshold level, an android application will generate notification. After clicking on the notification, worker will know all the information regarding the dustbin. After that garbage collector vehicle will collect the garbage. The ultrasonic sensor (HC-SR04) senses the level of the trash, and the moisture sensor will check the garbage is wet or dry. The MQ135 is the gas sensor which has the excessive affect ability to ammonia, sulfide, and beans steam, likewise delicate to smoke and different hurtful gases. Each garbage bin is having GPRS sensor for location tracing. After the trash reaches to threshold level, message was passed to the authorized person along with the location of the garbage bin; the authorized person checks the information and reports the garbage collector officer to collect the garbage from that mentioned location. Every garbage bin was allocated with unique id, and that unique id was stored in the database which having all the information regarding that garbage bin, like in a day how many times that garbage bin was come for garbage removal (Figs. 1 and 2).

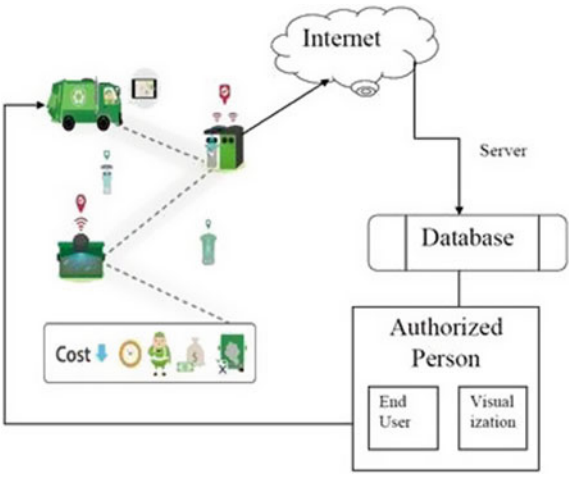


Fig. 1 Working of proposed model

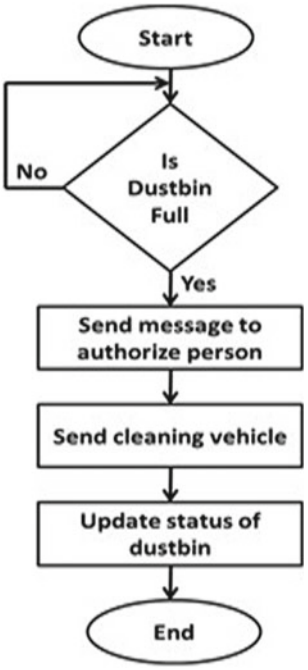


Fig. 2 Flowchart of proposed model

4 Experimental Setup and Results

See Fig. 3.

The experimental setup requires following components.

Raspberry Pi 3B—The Raspberry Pi having a 64-bit processor, double-band remote LAN, Bluetooth, faster Ethernet, the double-band remote LAN accompanies specific consistence accreditation, allowing the burden up to be deliberate into finished outcomes with basically reduced, remote LAN consistence checking out, running on both expense and time to promote it.

Ultrasonic Sensor (HC-SR04)—in modern applications, ultrasonic sensors are portrayed by their unwavering quality and outstanding versatility. Ultrasonic sensors can be utilized to address even the most intricate errands involving object location or level estimation with millimeter accuracy, on the grounds that their estimating method works dependably. No other estimating technique can be effectively put to use on a wide scale and in such countless various applications. The gadgets are very strong, making them suitable for even the hardest conditions.

Air Quality Sensor—the air quality sensor MQ135 has excessive affectability to ammonia, sulfide, and benzene steam, likewise delicate to smoke and different hurtful gases. An air high-quality sensor used for identifying a huge scope of gases, along with NH3, NO, CO2, smoke, and benzene. This is ideal for use in workplace or

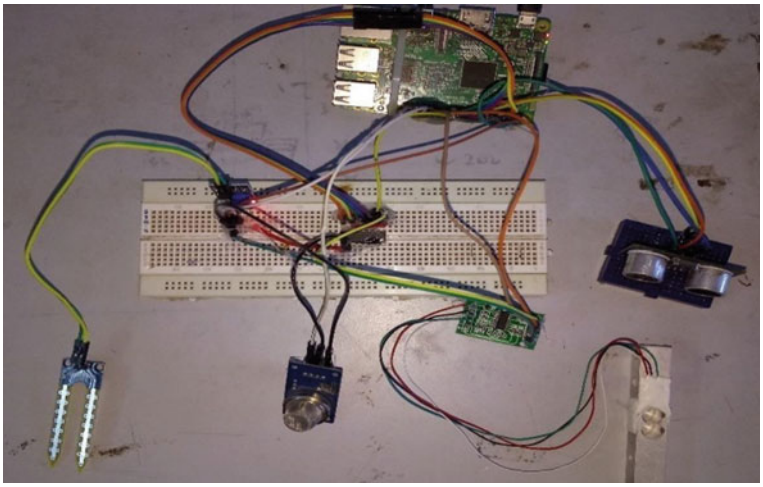


Fig. 3 Experimental setup

processing plant. It is with minimum cost and in particular affordable for air excellent looking at application.

Moisture Sensor—the 2 probes of moisture sensor allow the contemporary to bypass through the garbage after which it receives the resistance fee to measure the moisture fee. Dry rubbish conducts power poorly. So, while there will be less water, then rubbish will behave low strength because of this that there could be more resistance. Therefore, the moisture stage may decrease. While there is extra water, the rubbish will conduct greater power this means that that there might be much less resistance. Therefore, the moisture stage will be better.

Results

Figure 4 indicates readings taken by weight sensor for five dustbins. The weight measurement is compared with 50% threshold value. The dustbins whose weight going to be above threshold level will be identified for garbage collection; pickup vehicle will be send to respective location.

Figure 5 indicates readings taken by moisture sensor for five dustbins. The moisture measurement is compared with 40% threshold value. The dustbins whose moisture going to be above threshold level will be identified for garbage collection; pickup vehicle will be send to respective location.

5 Conclusions

The implementation of IoT based dustbin is totally sensor based dustbin offers a solution for environmental situation in a town. Implementation of smart dustbin is

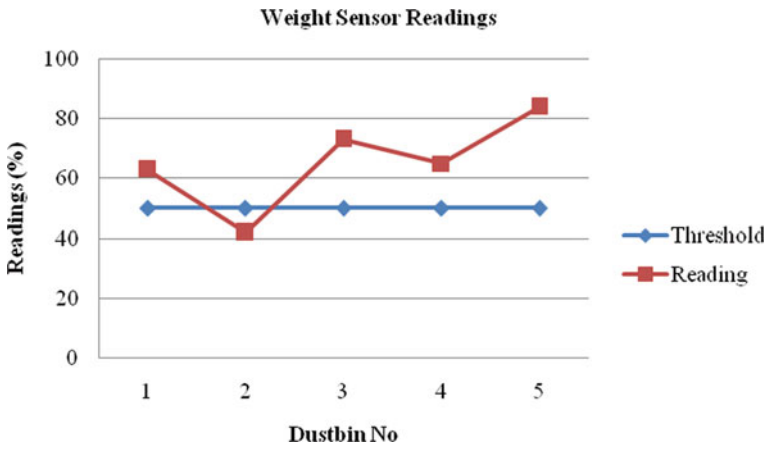


Fig. 4 Weight sensor readings

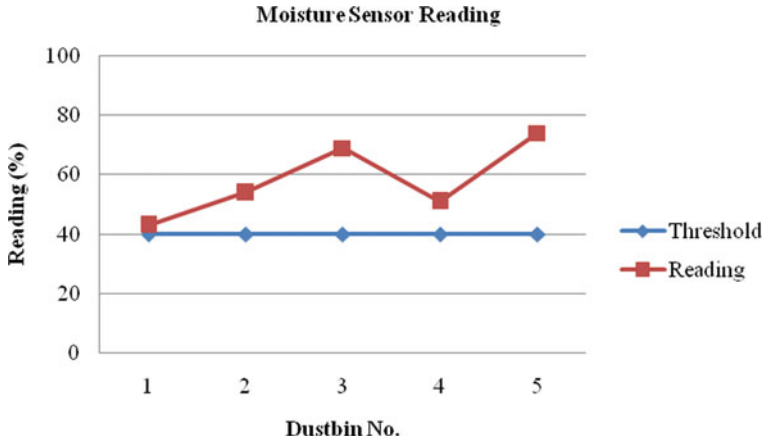


Fig. 5 Moisture sensor readings

accomplished the use of diverse sensors with Raspberry Pi. Prototype device assures to deliver information of every sensor to internet. While the dustbin will acquire to threshold, a notification can be deliver to android utility. After clicking on the notification, android software will open, and it will display the vicinity of stuffed dustbin in Google map. The main purpose of this task is to lessen human assets and efforts along the enhancement of a smart city imaginative and prescient. We have regularly seen garbage spilling over from dustbins directly to streets and this turned into a difficulty that required instant attention.

References

1. Wijaya A, Zain U, Niswar M (2017) Design smart bin for smart waste management. In: 5th international conference on instrumentation control
2. Kumar V, Kumaran S, Kumar K, Mahapati M (2017) Smart garbage monitoring and clearance system using IOT. IEEE International Conference
3. Bharadwaj B, Kumudha M, Gowri C (2017) Automation of smart waste management. In: 2nd International Conference
4. Singh B, Kaur M (2016) Smart dustbin for smart cities. IJICCE 4(2)
5. Jabbar WA, Kian TK, Ramli RM, Zubir SN, Zamrizaman NSM, Balfaqih M, Shepelev V, Alharbi S Design and fabrication of smart home with internet of things enabled automation system. IEEE Access 7:2169–3536
6. Lulla G, Kumar A, Pole G, Deshmukh G (2021) IoT based smart security and surveillance system. Int Conf Emerg Smart Comput Inf
7. Patel D, Patel F, Patel S, Patel N, Shah D, Patel V (2021) Garbage detection using advanced object detection techniques. Int Conf Artific Intell Smart Syst
8. Medehal A, Annaluru A, Bandyopadhyay S, Chandar TS (2020) Automated smart garbage monitoring system with optimal route generation for collection. In: Proceedings of IEEE international smart cities conference, 1–7
9. Anitha A (2017) Garbage monitoring system using IoT. In: Proceedings of IOP conference series: materials science and engineering
10. Goel A, Goel A (2021) Smart garbage monitoring system: book chapter of smart and sustainable intelligent systems chapter 24. Wiley Online Library, Published
11. Sathish Kumar N, Vijayalakshmi B, Jenifer Prarthana R, Shankar A (2016) IOT based smart garbage alert system using arduino UNO. In: Proceedings of the international conference IEEE Region 10 conference IEEE, 1028–1034
12. Nirde K, Mulay PS, Chaskar UM (2017) IoT based solid waste management system for smart city. In: Proceeding of International Conference on Intelligent Computing and Control Systems, IEEE, 666–669
13. Aswin Raaju V, Mappilllai Meeran J, Sasidharan M, Premkumar K (2019) IoT based smart garbage monitoring system using ZIGBEE. In: Proceeding of international conference on systems computation automation and networking, IEEE
14. Praveen A, Radhika R, Rammohan MU, Sidharth D, Ambat S, Anjali T (2020) IoT based Smart Bin A Swachh- Bharat initiative" Contribution title. In: Proceedings of the international conference on electronics and sustainable communication systems, IEEE, 83–786
15. Melo AB, Oliveira AM, de Souza DS (2017) Optimization of garbage collection using genetic algorithm. In: 14th International conference on mobile ad hoc and sensor systems, IEEE
16. Sharma P, Kumar P, Nigam R, Singh K (2020) Automatic waste segregating and self sanitizing dustbin. In: 2nd international conference on advances in computing, communication control and networking
17. Annie Lincy F, Sasikala T (2021) Smart dustbin management using IOT and Blynk application. In: 2021 5th international conference on trends in electronics and informatics
18. Ayush A, Kumar A, Jha A, Sarkar N, Moharana SC, Das H (2019) Voice controlled automatic dustbin with garbage level sensing. Int Conf Intell Comput Control Syst
19. Dubey S, Singh MK, Singh P, Aggarwal S (2020) Waste management of residential society using machine learning and IoT approach. Int Conf Emerg Smart Comput Inf

Estimation of Accuracy in Gender Identification Based on Voice Signals Using Different Classifiers

Abhishek Singhal and Devendra Kumar Sharma

Abstract This paper presents the estimation of accuracy of different classifiers which are used for the identification of gender with the help of voice signals. The voice signal has rich information about the speaker. The analysis of the voice signals is very vital for accurate and fast identification of gender. In this paper, the Mel Frequency Cepstral Coefficients (MFCCs) are used as an extracted feature of the voice signals of the speakers. MFCCs are the most convenient and reliable feature which sets to configure the gender identification system. Recurrent Neural Network—Bidirectional Long Short-Term Memory (RNN-BiLSTM), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) Gaussian Mixture Model (GMM) and K-Nearest Neighbor (KNN) are utilized as classifiers in this proposed work. In this article, the RNN-BiLSTM classifier has single-layer architecture, while SVM and LDA have a kfold value of 5. The highest accuracy for gender identification is found as 88.66%. The result of the simulations shows that the accuracy of the RNN is always found at a higher value as compared to SVM, LDA, GMM and KNN.

Keywords Classifier · Support vector machine · Linear discriminant analysis · Recurrent neural network—bidirectional long short-term memory · Mel frequency cepstral coefficients · Accuracy

1 Introduction

As per the dictionary of Webster, "Speech" is the "communication or expression of throughout in speaker words." The voice signals have carried the information of the message and the characteristics of the speaker during the conversation. The nonlinguistic attributes of the signals are utilized to identify the gender of the speaker. The variation in the features of the voice signal depends upon the plasticity of the voice signal generator and several internal and external environmental parameters. The voice signal has also carried the information of the speaker, which are related to

A. Singhal (✉) · D. K. Sharma
Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Delhi, NCR Campus, Ghaziabad, UP, India
e-mail: abhisheksinghal.srm@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_33

the affective factors, social factors, physical structure of the voice generator, age and gender. By analyzing the voice signals, the characteristics are also identified even then the speaker is hidden or during the conversation on telephone.

Speaker has several valuable characteristics which can be utilized in several applications. Gender is the unique characteristic of the speaker. This characteristic of the speaker has several special applications such as gender-based surveys through telephones conversation, automatic speaker recognition, voice recognition [1], innovative workplace and robot-human interaction [2]. The identification of the gender from the voice signals is also utilized in criminal cases as available evidence in the recorded form. The characteristics of the voice signal are unique for the specific gender or speaker [1–3]. In consideration of the security issues which affect the whole world, the gender classification shows excellent attention toward the researcher. A gender recognition system is a mandatory requirement in the current rapid development of the computerized environment [4]. In the present scenario, the analysis of the voice signals is very useful for disabled people, telephony military systems, and health care systems. Due to these applications, the processing of the speech signals is the hotcake for the research.

This paper discusses an approach for gender identification through voice with the help of several different classifiers and a feature, namely, SVM classifier, LDA classifier, RNN-BiLSTM classifier, GMM classifier, KNN classifier and Mel Frequency Cepstral Coefficients (MFCC) with 12 coefficients. The algorithms of the SVM, RNN-BiLSTM and LDA are developed for the analysis of the voice signals. The SVM and LDA algorithms have the kfold value of 5. A new optimization technique named "ADAM" is used in the RNN-BiLTSM algorithm to achieve the goal of the proposed work. Among the five classifiers, RNN-BiLSTM shows the highest accuracy percentage with 88.67% with single-layer architecture. The remaining part of the paper is organized as follows. Section 2 explains the work which has been reported in the literature. The material and methods are described in Sect. 3. Section 4 shows the results and discussion of the proposed work. Section 5 concludes the paper.

2 Literature Review

The MFCC was developed and utilized to identify gender in 2012. After the development of MFCC, several modifications have been taken place to improve the performance of the system. MFCC has also been examined for the identification of gender in different domains [5]. SVM is generally used for binary classification to identify gender. It shows the best accuracy in the class separation technique. An article reported that the Gaussian Radial Basis function SVM has the best accuracy compared to other kernels of SVM [2]. The gender identity of the speaker was addressed with age. Hidden Markov Model (HMM), Gaussian Mixture Model (GMM) and SVM classifiers are also reported in the literature to identify the gender of the speakers. Interpretation of the performance of the system is challenging due to variation in the data sets, features coefficients, protocols, etc. [6].

The information of the gender can be utilized to develop the gender-dependent modules. To identify the gender and age, the features of the voice signals were decided for all samples database. To create the super vector, the hybrid model was developed to train the system in the training phase. This method was advantageous in criminal cases. The characteristics of the voice signals are also varied according to the spoken text and the native language. Lexical features and linguistics features were utilized to identify the age and gender of the speaker [7].

3 Materials and Methods

3.1 About the Database

Voice signals are available in the.mp3 format. Samples of the voice signals are downloaded from the available voice website: http://voice.mozilla.org. The duration of the voice samples is 2–7 s. The voice signals are recorded at 44 kHz. The samples for gender identification groups are distributed as follows: 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200. The number of samples for the male and female are equals in each group. Every group of the voice samples has 80% samples for the training and 20% samples are reserved for the testing purpose.

3.2 The Architecture of the Gender Recognition System

The architecture of the gender recognition system is divided into two phases: the training phase and the testing phase, as shown in Fig. 1. Both phases have a common step which is feature extraction. This step is also known as the front end of the system. In the process of feature extraction, the voice signal is converted into numerical signifiers. A series of numerical signifiers are also known as feature vectors. The valuable information and parameters of the voice samples are extracted in this process.

In the training phase, 80% of the voice samples in each database group are used to extract the feature MFCCs. The system is trained with the help of extracted feature vectors to segregate the gender as male and female. In the testing phase, the extraction of the features is similar to the training phase. With the help of classification techniques, the system shows the result in percentage accuracy for the identification of genders.

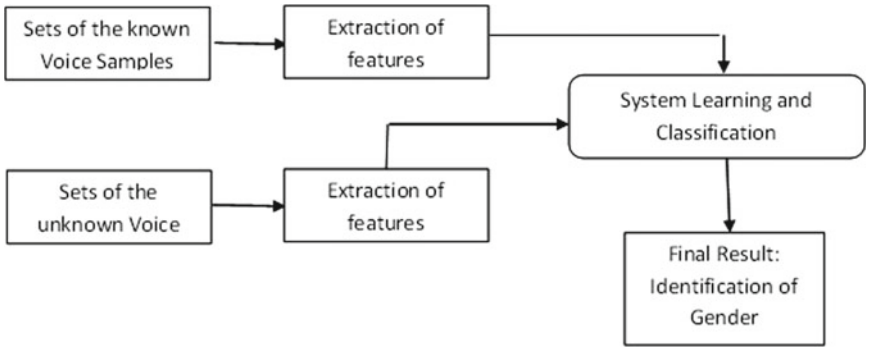


Fig. 1 Model for voice signal processing

3.3 Extraction of the MFFCs

The process for data reduction is known as feature extraction. The reduced data is beneficial to discriminate the gender of the speakers. The selection of the feature vector is essential for the performance of the system because the feature vector affects the behaviors of the training and classification process of the system. MFCCs are the unique features for the identification of gender. The MFCCs are extracted from the voice signals by using the following steps [1, 6, 8]:

- Pre-emphasis is the first step. In this step, the energy level of the high-frequency signals is increased.
- In this step, the framing of the voice signals is performed.
- To create a window shape for the feature extraction, Hamming window is used.
- To find the magnitude and frequency response of every frame of the voice samples, the Fast Fourier Transform (FFT) is utilized.
- Due to the wide range of the FFT spectrum, the mel scale filter bank is used. The output of this step is the log mel spectrum.
- In this step, the log mel spectrums are converted into a time-domain with the help of Discrete Cosine Transform (DCT). The result of this step is known as MFCC. The collection of MFCC is called a feature vector.

3.4 Classification Algorithms

The process of classification is similar to the process of supervised learning. In the classification, the features of the output voice signals are mapped with features of the input voice signals. There are many types of classification techniques that can classify the genders such as SVM, RNN and LDA.

3.5 Support Vector Machine (SVM)

SVM is a very powerful algorithm to identify gender with the help of voice signals. The main aim of the SVM is to fix the hyperplane according to the features that differentiate the genders. With the help of the hyperplane, the SVM can execute the binary classification [9]. The data points, which are lying near the hyperplane, are known as support vectors. The separation of the support vector is complicated from the other data points which are available near the hyperplane. The classification of the unknown samples is decided by the margin value. The margin is the perpendicular line from the hyperplane [2, 7, 10].

3.6 Recurrent Neural Network (RNN)

Based on the working of the human brain, a non-linear classifier is developed, known as an artificial neural network (ANN). In the ANN, the weights and biases are continuously adjusted throughout the training phase. This process is going on until the changes in the values of the consequences and bias will become negligible [11–13]. The basic or initial ANN has only three layers: input layer, one hidden layer and output layer. As per the application, the hidden layer can be increased. RNN is also a form of ANN with internal loops. Several operational models are available to implement the RNN such as Vector to Vector mapping, Sequence to Vector mapping, Vector to Sequence mapping and Sequence to Sequence mapping [5, 14].

3.7 Linear Discriminant Analysis (LDA)

LDA is generally used to segregate between two labels or many labels. If the classes are two then labels are classified linearly with the help of one hyperplane. On the other hand, several hyperplanes are required to segregate the classes in multiple discrimination. The hyperplane is developed according to the following principles: (i) the distance between the two labels is maximum and (ii) the variation of the values of the features should be minimum in both labels [15].

4 Results and Discussion

The voice signals have several characteristics and information about the speaker such as gender, age, emotion, health. In several applications, the gender of the speaker has an important and essential parameter. So, with the help of voice analysis, the gender of the speaker can be easily identified. The MFCCs feature is used to determine

the accuracy of the system for gender identification. The percentage accuracy is calculated with the help of a confusion matrix.

Figure 2 shows that the accuracy of the SVM classifier with a kfold value of 5. Initially, for the 20 samples, i.e., ten samples for males and ten samples for females, the system has maximum accuracy. If the number of samples has increased, the accuracy decreases because the data vectors also increase near the hyperplane. So, during a large number of the dataset, the percentage accuracy for identification of the gender will decrease.

For the same dataset, the percentage accuracy for the RNN-BiLSTM is varied as shown in Fig. 3. For the data sets 40, 80, 120 and 140, the system steers the worse accuracy. The highest accuracy is 88.66%. During the execution of the algorithm, an epoch is fixed as ten and the hidden layer is one.

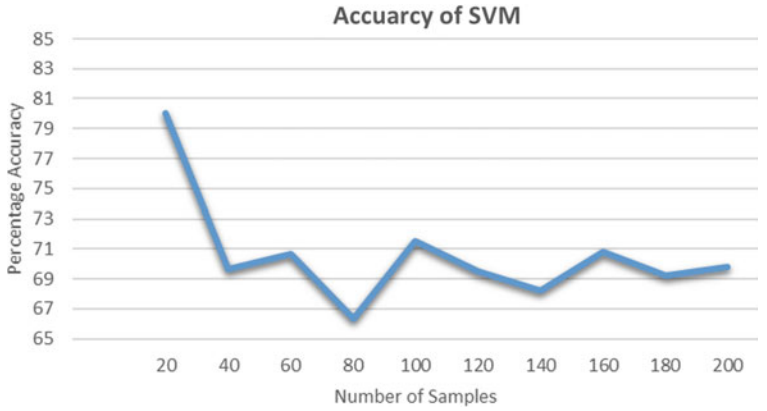


Fig. 2 Percentage accuracy for the SVM

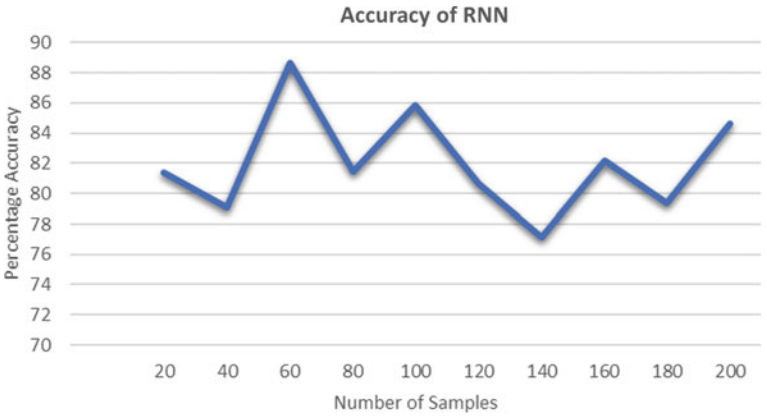


Fig. 3 Percentage accuracy for the RNN

The accuracy of the system is consistently decreased if the number of voice samples increases, as shown in Fig. 4. The system shows maximum accuracy at the 20 samples because the support vectors are less near the hyperplane. If the number of samples is increased, the accuracy decreases.

As per Fig. 5, the accuracy of the RNN-BiLSTM is the highest accuracy value in comparison to the other classifiers. RNN-BiLSTM accuracy is continuously high for all datasets of the voice samples. On the other hand, the accuracy of the SVM classifier is approximately similar to the accuracy of LDA. The accuracy of SVM, as well as the LDA classifier, shows the worse result.

Figure 6 represents the average accuracy of the different classification systems. The average accuracy for the RNN-BiLSTM classifier is 82.03%. On the other hand,

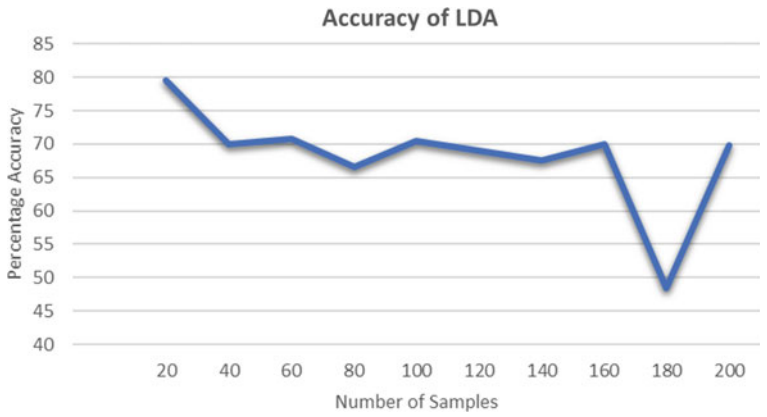


Fig. 4 Percentage accuracy for the LDA

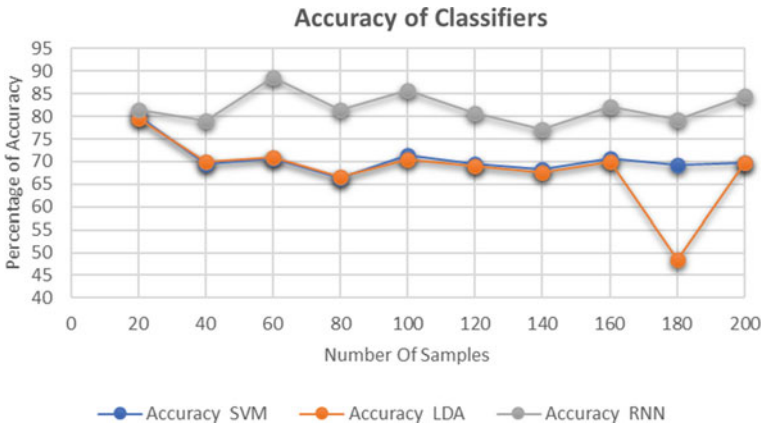


Fig. 5 Percentage accuracy comparison for the different classifiers

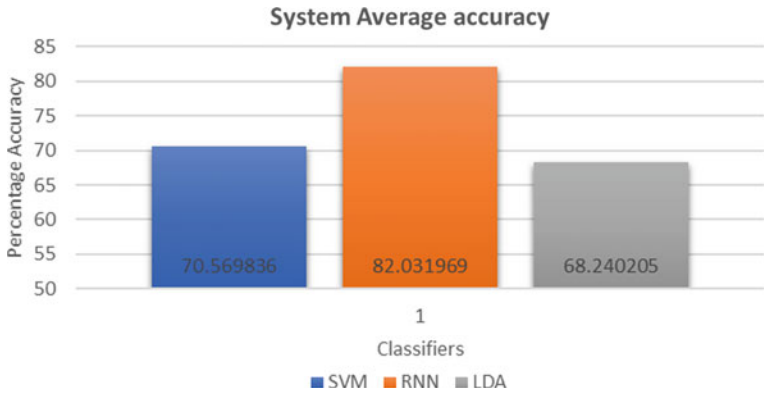


Fig. 6 Percentage average accuracy comparison for the different classifiers

the average accuracy of the SVM classifier is 70.56%. LDA has the minimum system accuracy in identifying the gender, which is equal to 68.24%.

The accuracy of the system also depends on the length of utterance. If the length of the utterance is increased, the accuracy of the system is also increased. In the above experiments, the length of the utterance is short. To extend the comparison of the accuracy of the classifiers, GMM and KNN are also involved. The accuracy for these classifiers is based on the results that are available in the literature. The maximum accuracy for the RNN-BiLSTM algorithm is 88.67%, while SVM has 80.05% accuracy. The accuracy for the LDA is 79.59%. GMM and KNN has 78.2% [6] and 66.66% [16] system accuracy, respectively. After comparing the accuracy for all five classifiers, RNN shows the best accuracy and other classifiers show slightly worse accuracy.

5 Conclusion

The estimation of the accuracy for the identification of gender is the main objective of this article. For this purpose, five classifiers and one feature are considered. The accuracy for SVM, RNN-BiLSTM and LDA was calculated after analysis of the voice samples and compared with the accuracy of the GMM and KNN as available in the literature. The MFCCs are used as a feature of the voice signals. The RNN-BiLSTM has achieved an accuracy of 88.67%, which is higher than the other classifiers. The KNN classifier shows the worst accuracy. The performance of the classification system can be enhanced if the number of epochs is also increased in case of the RNN-BiLSTM classifier. For better accuracy in identifying the gender, the emotion of the speakers is also included because the emotion of the male and females has very wide and different characteristics. The health condition, emotions and age are

also useful to predict the gender of the speakers. The accuracy of the system may also be increased by using the majority rule in the confusion matrix.

References

1. Passricha V, Aggarwal RK (2019) Convolutional support vector machines for speech recognition. Int J Speech Technol 22(3):601–609
2. Ramdinmawii E, Mittal VK (2016) Gender identification from speech signal by examining the speech production characteristics. In: 2016 international conference on signal processing and communication (ICSC), 244–249
3. Patil V, Vineetha R, Vatsa S, Shetty DK, Raju A, Naik N, Malarout N (2020) Artificial neural network for gender determination using mandibular morphometric parameters: a comparative retrospective study. Cogent Eng 7:1–12
4. Faragallah OS (2018) Robust noise MKMFCC–SVM automatic speaker identification. Int J Speech Technol 21(2):185–192
5. Nasr MA, Abd-Elnaby M, El-Fishawy AS, El-Rabaie S, El-Samie AFE (2018) Speaker identification based on normalized pitch frequency and mel frequency cepstral coefficients. Int J Speech Technol 21:921–941
6. Djemili R, Bourouba H, Korba MCA (2012) A speech signal-based gender identification system using four classifiers. Int Conf Multi Comput Syst, 184–187
7. Chaudhari SJ, Kagalkar R (2015) Methodology for gender identification, classification and recognition of human age. Int J Comput Appl Natl Conf Adv Comput, 5–9
8. Ratanpara T, Patel N (2015) Singer identification using perceptual features and cepstral coefficients of an audio signal from Indian video songs. J Audio Speech Music Proc 16:1–12
9. Jena B., Mohanty A. and Mohanty S. K.: Gender Recognition of Speech Signal using KNN and SVM. International Conference on IoT based Control Networks and Intelligent Systems, (2020)548–557
10. Gupta M, Bharti SS, Agarwal S (2016) Support vector machine based gender identification using voiced speech frames. In: 2016 fourth international conference on parallel, distributed and grid computing (PDGC), 737–741
11. Reddy MK, Rao KS (2020) Excitation modelling using epoch features for statistical parametric speech synthesis. Comput Speech Lang 60
12. Liu Y, He L, Liu J, Johnson MT (2019) Introducing phonetic information to speaker embedding for speaker verification. J Audio Speech Music 19
13. Greco A, Saggese A, Vento M, Vigilante V (2020) A convolutional neural network for gender recognition optimizing the accuracy/speed tradeoff. IEEE Access 8:13077–130781
14. Radzikowski K, Nowak R, Wang L, Yoshie O (2019) Dual supervised learning for non-native speech recognition. J Audio Speech Music 3
15. Castaldelloa C, Guberta A, Galvaninb F, Casonatoc A, Padrinid R, Baroloa M, Bezzoa F (2017) A model-based support for diagnosing von willebrand disease. In: 27th European symposium on computer aided process engineering—ESCAPE 27, 2779–2784
16. Abdulsatar AA, Davydov VV, Yushkova VV, Glinushkin AP, Rud VY (2019) Age and gender recognition from speech signals. J Phys: Conf Series, 1–7

Detection of Covid-19 Using CT-Scan Images and Deep Transfer Learning

Potluri Sairaj and Kritika Jain

Abstract Several variants of the severe acute respiratory syndrome (SARS-Cov2) were identified globally in recent times. According to the survey conducted by World Health Organization (WHO) in the year 2021, there were four dominant variants of the virus, which were named alpha, beta, gamma, and delta, respectively. Radiological examinations such as CT scan and chest X-rays are considered to be effective clinical examinations that determine the presence of viruses in the human body. The study of these radiological examinations aids to prevent the spread of the virus and supports in treating the patients effectively. This paper aims to discuss various aspects of covid-19 and describes an automated system that uses the CT-scan images to diagnose covid-19. The proposed method uses deep learning technique along with SoftMax activation function and was found to be effective in determining the novel virus.

Keywords Computer vision and pattern recognition · COVID-19 · Deep learning · Convolution neural network

1 Introduction

Coronavirus pandemic (COVID-19) continues to be a global problem that threatens global health. In an effort to prevent the spread of diseases, early deterrence is one of the first ways to fight this pandemic. While the current gold standard for illness detection is the reverse transcription-polymerase chain reaction (RT-PCR), molecular tests of respiratory tract spectrums are also highly recommended that provide laboratory infection confirmations. However, the fast spread of COVID-19 led to inadequate laboratory kits and created a major issue. Therefore, the use of radiological exams to diagnose illnesses during the COVID-19 pandemic became increasingly popular [1, 2].

P. Sairaj (✉) · K. Jain
Department of Computer Science and Engineering, Jain University, Bangalore, Karnataka, India
e-mail: sairajpotluri@outlook.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_34

While CT scans have proved more effective, growing number of patients and the resulting increase in radiology tests prevent each individual patient from receiving diagnoses to being continually dependent on chest CT scans. Furthermore, a high dependence on CT scans imposed a substantial on X-rays (CXRs), which enabled a more viable alternative for the detection of COVID-19. While CXRs in early stages are considered less sensitive to COVID-19, the progressive development of luminary abnormalities is advantageous [3–5].

Researchers have started employing cutting-edge technology for the automated identification of COVID-19 in patients with the fast worldwide spread of COVID-19. In its earliest phases, the burden of acquiring data from COVID-19 has pushed researchers to construct a model of their own utilizing pretrained networks. But, most of these investigations have employed only few COVID-19 samples in a small dataset. This makes it difficult to generalize the outcomes of these research and does not assure that the reporting output is maintained when these models are assessed on a bigger dataset. Therefore, a large dataset must be used to verify the deep learning technique used for the detection of COVID-19 in 2D scan images. As well as the fact that it is regarded wrong to combine healthy and pneumonia instances when the model tries to neglect the intra-class variance within these two classes, the exactness attained in this method is not an exact measure.

The distinction between viral and bacterial pneumonia and the diagnosis of the most prevalent thoracic illnesses has been demonstrated to include in-depth learning. In addition, the task is to build a COVID-19 algorithm that identifies a patient. This work remains difficult, however, as COVID-19 can share comparable X-rays with other pneumonia kinds. In [9], the authors pointed to the low performance of MobileNet in instances of COVID-19 in comparison to other cases of pneumonia, where only bacterial pneumonia cases were included in training data. We therefore seek, by speeding up detections of COVID-19 clusters generated by a new virus, to separate COVID-19 from viral pneumonia (not bacterial pneumonia). Moreover, a significant imbalance in the COVID-19 vs non-COVID classification has occurred as a result of a difficulty in getting an appropriate number of positive corona virus samples, compared to the number of non-positive samples [6–9].

In this framework, introduction was discussed in Sect. 1. Related work will be discussing in Sect. 2; proposed work will be describing in Sect. 3, and Sect. 4 consists of results for our proposed work, and Sect. 5 will be concluding the paper.

2 Related Work

Background CT chest is a major complement of RT-PCR test used in identification 2019 coronavirus illness. Purpose to examine the diagnostic usefulness and consistency of chest CT in relation to the COVID-19 RT-PCR test. Materials and processes in Wuhan, China, 1014 patients received both CT and RT-PCR testing from 6 January to 6 February 2020. The performance of breast CT in the COVID-19 prognosis was

evaluated using RT-PCR as the reference standard. In addition, the dynamic conversion of RT-PCR (negative to positive, positive to negative) findings was analyzed for patients with numerous RT-PCR assays compared to serial chest CT scanners for those with a gap of 4 days or more between RT-PCR tests [4].

Results of 1014 patients, sixty-one out of 1014, were positive results for RT-PCR (59%), and 888 out of 1014 (88%) received positive CT scans for the chest. Sensitivity of chest CT on the basis of positive data from RT-PCR suggested that COVID-19 is 97% (95% confidence interval 95%, 98%; 580 out of601 individuals). 308 out of 413 (75 percent) exhibited positive chest CT findings in 413 individuals with negative T-PCR results [10].

Reliability and interpretability issues arise in the deployment of clinical decision support systems for medical imaging. Here, we are developing a deep learning diagnostic tool for the screening of patients with prevalent retinal blinding disorders. We use transfers that form a neural network using a fraction of standard approach data. Using this method, we demonstrate a similar performance in the classification of aging-related macular degeneration and diabetic macular edema using optical consistency tomography pictures.

In addition, by emphasizing the areas detected by the neural network, we give a clearer and more interpretable diagnose. We further demonstrate that our AI approach is generally applicable for diagnostic pediatric pneumonia with chest X-ray pictures. In the final analysis, this tool can help speed the diagnosis and reference of these curable diseases, which facilitates earlier treatment and leads to better clinical results [6].

We cover the techniques of accuracy assessment and compare validation and bootstrap with the two most prevalent approaches. Recent test findings on fake data and theoretical cultures in constrained circumstances have revealed that tenfold cross-validation may be superior to a more expensive one-out leave validation for picking an appropriate classifier from a collection of classifiers (model selection). A large-scale experiment—over half a million, a Naive Bayes algorithm—is reported to evaluate the effects on real-world datasets of varied settings on these algorithms. We alter the number of folds for cross-validation and whether or not the folds are stratified; for bootstrap, we change the number of samples for bootstrap. Our results show that the optimum approach to employ to pick the model is tenfold stratified cross-validation for actual words datasets comparable to ours even if computing capacity permits more folding.

In [1], the model proposed is based on XceptionNet. An automated method is suggested for diagnosis of COVID-19 using images of X-ray scans. The author has used computer vision and pattern recognition techniques along with deep learning algorithms to receive high accuracy. The comparative study mentioned in the paper depicts the method provides high accuracy than other methods.

Most notably, in [11], 3D CNN-based FE model with coupled regularization to extract effective hyperspectral authors discussed about imaging spectrum-spatial characteristics. Finally, a virtual sample improved technique is presented to further increase performance. Three hyperspectral collections of data are being developed: Indian Pines, University of Pavia, and Center Kennedy.

Table 1 Related works summary

Author and Year	Proposed	Finding/Outcomes
T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia (2020)	Association of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases	COVID-19 disease testing
D. S. Kermany et al. (2018)	Detection of medical diagnoses and treatable diseases by image-based deep learning	Medical image disease prediction using deep learning
R. Kohavi	A study of cross-validation and bootstrap for accuracy estimation and model selection	Brief study about cross-validation and bootstrap
Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi	Deep feature extraction and classification of hyperspectral images based on convolutional neural networks	Using CNN for classification of hyperspectral images
A. L. Maas, A. Y. Hannun, and A. Y. Ng	Rectifier nonlinearities improve neural network acoustic models	Neural networks acoustic model

Acoustic deep neural network models generate significant improvements in big vocabulary systems for continuous voice identification. Emerging work using hidden rectified linear units (ReLU) shows further advantages in final system performance compared to more often utilized nonlinear sigmoid units. We examine in this work the usage of deep corrective networks as acoustic models for the 300-h conversation identification challenge in the switchboard.

Networks using nonlinear rectifiers result in 2 percent absolute reductions in word mistake rates in relation to their sigmoidal counterparts by just training without pre-training. In order to quantize the difference in how ReLU [9] units encrypt inputs compared to the sigmoid units, we investigate hidden layer representations. In order to further improve the depth rectifier nets, we analyze a variation of the ReLU unit with a gradient that is more optimal (Table 1).

3 Methodology

The procedure to develop our system is described in this section.

- At first, to perform the operation, we need a dataset of the CT-scan images. We have collected covid-positive and covid-negative images from various sources and

have prepared our own dataset for the process. Next, we performed augmentation on the prepared dataset.

- Once the dataset is augmented, we preprocessed it. In the preprocessing phase, we removed noise from the data and split the dataset into training data and test data.
- In the training process, we built our proposed network with the help of Keras and TensorFlow. Here, we used deep learning model, convolutional neural network (CNN) along with transfer learning (MobileNet).
- The proposed transfer learning-based CNN architecture (MobileNet) is based on neural network model. The proposed model is built using convolution layer, pooling layer and dense layer followed by SoftMax and Leaky ReLU layers.
- After data training, we initiated the testing phase. Testing was conducted using the front end of the application we built using flask.
- In the testing phase, image was uploaded, and our model performed the operations, and the result was classified as covid positive or covid negative or pneumonia (Fig. 1).

Convolutional Neural Network

Convolutional neural network is a deep learning algorithm used primarily for visualizing the images and analyzing them. This mainly works on the basis of neural networks. They are also named differently based on the weights shared by them in the architecture of the considered kernels of convolution which are used for scanning the layers which are hidden. This network is used in other applications also which are

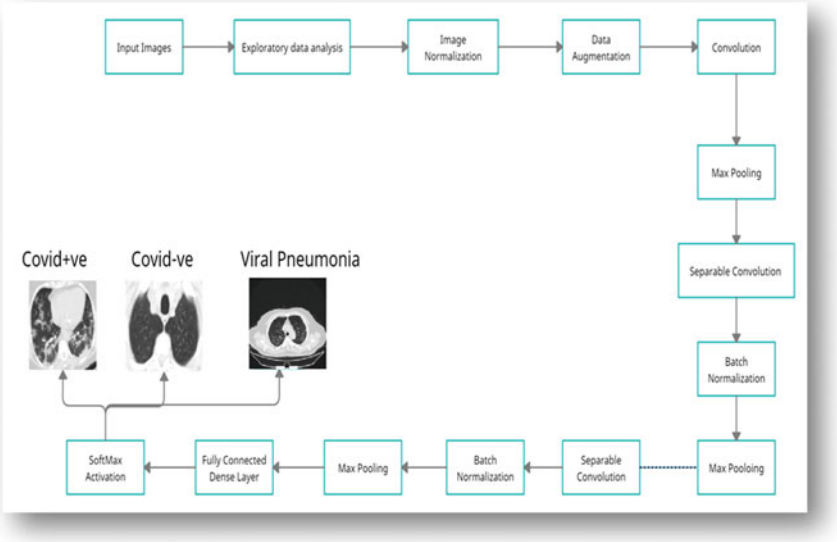


Fig. 1 Proposed method block diagram

used in analyzing images and video. This is used for segmentation, classification, and for analyzing the images in different fields. The name of this convolution is mainly given by the operations that are done using mathematics where this operation is used for the matrix multiplications in the layers of the network.

In neural network, the middle layers of feed forward are named as hidden because activation function is used for input and output masking. In hidden layer, some of the other layers are also included like those used for operation of multiplication, matrix operations, etc. Leaky ReLU is used as an activation function which allows small negative values when the input is less than zero. Along with these, some other layers are also present, namely pooling, fully connected and normalization layers.

Convolutional Layer

A convolution layer should follow the some of the attributes in a neural network. Convolution filters are also called as kernels that which means height and width, which are hyperparameters. The consideration of input and output channels that are to be used. The depth of the filter or kernel in an input channel of the convolution should be same as channels of the depth of the input. Padding size and strides are used in the operation of the convolution. The inputs from the layers of the convolution are considered, and those are passed to the next layers. This is considered as equal to the neuron response in the visual cortex to the stimulus.

Pooling Layer

Local layers or global layers are the pooling layers included in the convolution networks which are used for the computation. This layer is used for reducing the dimension of the considered dataset by summarizing the features thus reducing the number of parameters to learn. Here, two types of pooling are described, namely max layer and average layer. Max is used for calculating the maximum values, and average is used for calculating the average values.

Flattening

The next step was to flatten the pooled featured diagram. This step entails converting the whole pooled feature map matrix into a single column (long vector) which is then passed to the final classification model (fully connected layer).

Dropout

Dropdown can result in a network weight that is higher than average (over-fitting). Therefore, weights were first reduced by the selected dropout rate, before the network is finalized. The network was then be used to forecast as normal. The weight rescaling should be carried out instead after the workout, after each weight update at the end of the mini-batch. Often this is referred to as "reverse abandonment" and does not entail a weight change during exercise. The deep learning libraries Keras and PyTorch all enforce this dropout. Dropout ratio value used in the suggested model was 0.22.

Dense Layer

The layer named dense is one of the layers from a neural network which is deeply connected. It received an input from each neuron which was used in the previous layers. The most common layer used in the model is dense layer.

These layers are used for performing the multiplications of matrix vectors. Back propagation is used for training and updating the values that have been used in the matrix. The output generated from this layer was clubbed with a specific dimension as a vector. The dense layer was primarily used for dimension changing. Using this, some of the other operations were also being performed, namely rotation, translation, and scaling of the vectors.

ReLU Layer
ReLU is called as rectified linear unit, which was used for the function of activation. This is mainly used for giving a small slope for the negative values from the activation instead of making them as zero values. This is used for increasing the properties of nonlinear which are used for making the affectless the fields of the convolution.

Along with this, some other functions are also used that which help in the nonlinearity improvement.

SoftMax
The SoftMax is a function which is used for converting the real values of a vector to the sum to 1. Here, the inputs that were considered were a positive value, a negative value, and a zero value, or it may be also greater than one also.

SoftMax function is used to assign the values in between the range of 0 and 1 to each class in case of a multi-class problem like ours. So, that the probabilities are considered easily. If among the considered inputs there was a small value or a negative one, then this function turned that into a smaller probability. Similarly, if the input was large then it turned the probability to a larger probability.

Fully Connected Layer
Fully connected layers are the layers in which all inputs from one layer communicate with each triggering neuron of the next layer. They are entirely linked to one network. The last few layers in the most common learning machines are complete layers, which compile the data collected from the previous layers in order to form the final output. It is convolution layer's second most time-consuming layer.

RAdam Optimizer
Adam is a stochastic gradient descent alternative for deep learning model training. It combines the best aspects of the AdaGrad and RMSProp methods to develop an optimization technique for sparse gradients on noisy problems. Adam is straightforward to configure, and the default settings are adequate for most problems (Table 2).

4 Results and Discussions

In this section, we have compared our obtained result with similar studies. After the completion of preprocessing phase and the training of model, we tested the model

Table 2 Distribution of dataset

Image type	Training data	Test data
COVID + VE	590	177
COVID -VE	423	127
Viral pneumonia	357	108
Total	1370	412

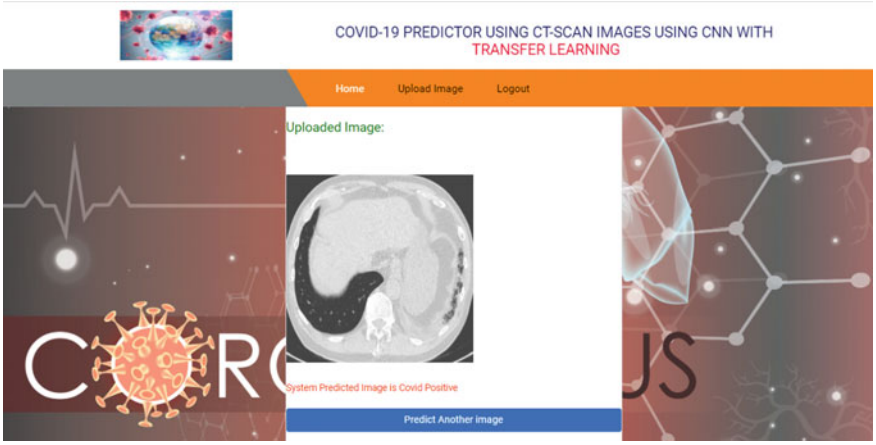


Fig. 2 Output predicted as covid positive

using test set. In the application, we uploaded test images to verify the output (Figs. 2 and 3).

Comparative Study
See Table 3.

5 Conclusion

The study provides the detection of corona virus positive or negative of the given CT images using a deep learning technique. Here, we used transfer learning-based convolution neural network (CNN) algorithm MobileNet. Our proposed architecture CNN, which is a one of the main algorithms of neural networks, is considered for the training. By using our proposed architecture, prediction of the results is accurate and provides a great performance. From our model, we can predict the covid case by uploading CT-scanned images.

Fig. 3 Output predicted as covid negative

Table 3 Comparison between multiple studies

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	*F*l—score (%)
DarkCovidNet [12]	87.12	89.95	85.33	87.36
DeTraC-ResNet18[13]	95.10	93.35	97.98	95.50
Proposed	95.82	95.13	93.21	95.84

References

1. Singh KK, Siddhartha M, Singh A (2020) Diagnosis of coronavirus disease (COVID-19) from chest X-ray images using modified XceptionNet. Romanian J Inf Sci Technol 23(S):S 91-S115
2. Udugama B, Kadhiresan P, Kozlowski HN, Malekjahani A, Osborne M, Li VYC, Chen H, Mubareka S, Gubbay JB, Chan WCW (2020) Diagnosing COVID-19: the disease and tools for detection. ACS Nano 14(4):3822–3835. https://doi.org/10.1021/acsnano.0c02624
3. American College of Radiology (ACR). ACR suggestions for the use of brest roentgen rays and computed tomography (CT) for suspected COVID-19 infection. Accessed 15 Jul 2020. [Online]. Available
4. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L (2020) Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 296(2):E32–E40. https://doi.org/10.1148/radiol.2020200642
5. Dangis A, Gieraerts C, Bruecker YD, Janssen L, Valgaeren H, Obbels D, Gillis M, Ranst MV, Frans J, Demeyere A, Symons R (2020) Accuracy and reproducibility of low-dose submillisievert chest CT for the diagnosis of COVID-19. Radiol Cardiothoracic Imag 2(2), Art. no. e200196. https://doi.org/10.1148/ryct.2020200196
6. Kermany DS et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122–1131. https://doi.org/10.1016/j.cell.2018.02.010
7. Jacobi A, Chung M, Bernheim A, Eber C (2020) Portable chest X-ray in coronavirus disease-19 (COVID-19): a pictorial review. Clin Imag 64:35–42. https://doi.org/10.1016/j.clinimag.2020.04.001

8. Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev early access 21 Apr 2020. https://doi.org/10.1007/s10462-020-09825-6
9. Maas AL, Hannun AY, Ng AY (2013) Rectifier nonlinearities improve neural network acoustic models. Presented at the Workshop Deep Learn. Audio, Speech Lang. Process., Int. Conf. Mach. Learn. (ICML), Atlanta, GA, USA, vol 30
10. Mossa-Basha M, Meltzer CC, Kim DC, Tuite MJ, Kolli KP, Tan BS (2020) Radiology department preparedness for COVID-19: radiology scientific expert review panel. Radiology 296(2):E106–E112. https://doi.org/10.1148/radiol.2020200988
11. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. in Proc. 14th Int. Joint Conf. Artif. Intell. Vol 2, pp 1137–1143.
12. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med, 103792
13. Abbas A, Abdelsamea MM, Gaber MM (2020) Classification of covid-19 in chest X-ray images using detrac deep convolutional neural network. arXiv preprint arXiv:2003, 13815

Controlling Load Frequency in a Hybrid Power System Including Electrical Vehicles

Nagendra Kumar, Priyanka Datta, Kailash Sharma, Rajat Mehrotra, and Jay Singh

Abstract This paper revolves around load frequency control (LFC) model with distributive energy resources and Electric Vehicle (EV) for a hybrid system. The research takes into account hybrid generating systems including wind turbine generators (WTG), photovoltaic (PV), solar thermal power systems (STPS), diesel sources (DEPS), fuel cells (FC), battery storage systems (BESS), ultracapacitors (UCs), and soon. Nowadays, Electrical vehicles (EVs) are presented a new opportunity for controlling and arrangement of power grids. Due to two-way system connection capabilities of converters in electrical vehicles, it considers as a safest backup for frequency support. This paper focused on EVs as one of the controlling loads with other conventional generation sources for frequency control after disturbances in a two-area or multi-area power system. The proportional-integral-derivative (PID) method is used in this paper to model electric vehicles. Optimization of tuning parameters and outcomes of PID controllers in EV and automatic generation control (AGC) systems, the big bang big crunch (BBBC) optimization technique is applied. Several methods, such as step and random load changes, have been used to demonstrate the controllers' efficacy relating to peak time, settling time, peak overshoot, and frequency deviations.

Keywords Load frequency control (LFC) · Tie-line · Electric vehicle · Renewable sources

1 Introduction

After any unbalance between generation and load requirement, load frequency control is accountable for controlling frequency in power systems. The purpose of LFC is to preserve scheduled frequency and tie-line power at normal conditions. The LFC is used at two stages or loops in power systems: primary control and secondary control [1]. The arrangement that take care the secondary frequency control in an

N. Kumar (✉) · P. Datta · K. Sharma · R. Mehrotra · J. Singh
Department of Electrical and Electronics Engineering, G. L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: nagendra.k962@gmail.com; nagendra.kumar@glbitm.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_35

electrical system is defined as AGC. It keeps the generation-load of an area in a specified limit and take care of any unbalance in frequency and power exchange on occurrence of load change [1, 2]. Since power system is a vast network therefore, to have a better control it is divided into control areas. Using tie-lines, one control area is linked to another for exchanging the power and to support one another in case of emergency conditions [3, 4]. Incorporating distributed generation (DG) into existing power networks to address growing power requirements is now a growing trend. Electricity can be produced using a single source system, such as photovoltaic, wind turbine, micro hydro, diesel set, or by merging two or more sources of electricity in a hybrid system [5]. The resultant hybrid power system seeks to deliver efficient and high-quality service to its clients. While these innovations may satisfy the demand for electricity they may affect the safe and efficient operation of entire system at same time. Furthermore, because load demand is highly erratic, maintaining the balance of generation and load is challenging [6]. The device frequency and voltage profile can be impacted if there is an imbalance between generation and load that lasts for a long time. Generally, the active power balance is responsible for preserving the nominal frequency of the system and the reactive power balance preserves the system's flat voltage profile. This equilibrium can be achieved via a control method, i.e., AGC [7]. As a result, AGC's primary role is to maintain system frequency and power between limits via tie-line in the case of a system failure [8]. A small rapid load shifts in either of the areas enables the frequencies of every area to fluctuate, as well as the power in the tie-line, in a linked power system. As a result, an LFC scheme essentially entails a suitable control system for a networked power system. Power system performance is influenced by control structure and fitness function, according to a literature review. I, PI, and PID [9] are important to researchers because of their benefits such as ease of installation, good performance. The usefulness of various control techniques has been demonstrated in a variety of settings, ranging from power systems to certain other systems. Previously, a traditional system based on thermal/hydro or a combination of both was investigated [10, 11]. In hybrid system study, the PI controller is utilized to control the outcome powers of distributed generation systems in order to satisfy power balancing conditions arises because of sudden generation and load variations. The PI controller's gain values are selected by trial and error technique. The Ziegler and Nichols approach in [12] was used to calibrate the traditional PI controller. It is been demonstrated that one conventional optimization strategy is also not recommended when there are a lot of parameters to optimize. In a hybrid system, the frequency is adversely affected by load deviations and changes in wind power, thus the power balance must be maintained. Reference [13] discusses the issue of frequency regulation in power systems with considerable wind penetration. In power systems with high wind system penetration, the issue of frequency control is raised. Some research assessing the effect of energy storage on frequency deviation considering DG system based on renewable resources was published. A small-signal analysis of a hybrid renewable system with energy storage, which including battery or pump storage, has been studied by many authors to give swift response and minimal oscillation. Electric vehicles are now widely used. Adopting demand

response provides us with a viable approach for restoring system frequency imbalance in unusual circumstances. Electrical vehicles (EVs) can engage in frequency management of power networks using this technology, as one of the viable replacements for demand response [14, 15]. In the presence of electric vehicles (EVs) and wind turbines, there has been a proposal for an LFC and multi model anticipatory control strategy. There is more and more research has been done in terms of EVs. PID parameters for the AGC scheme are determined using BBBC in this research because of inexpensive computing costs and high convergence speed over PSO and differential evolution (DE) optimization techniques.

2 Test System

Whenever a load imbalance occurs, the speed of prime mover changes to compensate that effect. The frequency also gets affected from that load imbalance. That change in speed is controlled by the governor of generators. Once the speed change eliminates, the governor speed is made constant. Due to a varying nature of load governor speed is also not fix to a set point, therefore, secondary control mechanism with integrator is used to remove offset and bring the system to its normal value called reset point. Figures 1 and 2 shows the test system with DG and other sources. Figure 3 represents complete test system taken in this paper to simulate the LFC effects. Various DG sources included in this research are described in [5].

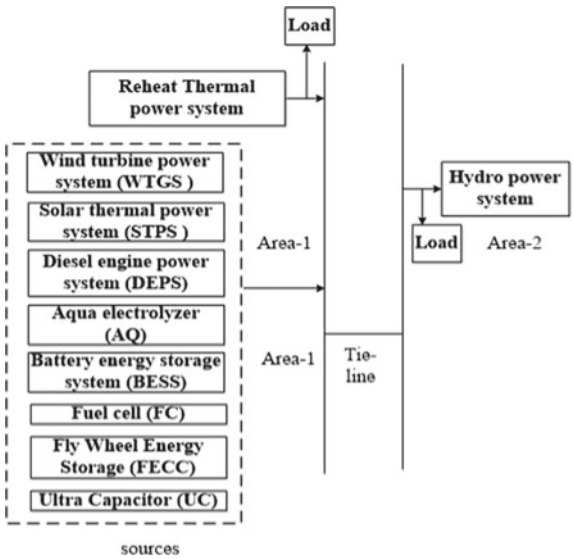


Fig. 1 Hybrid power System

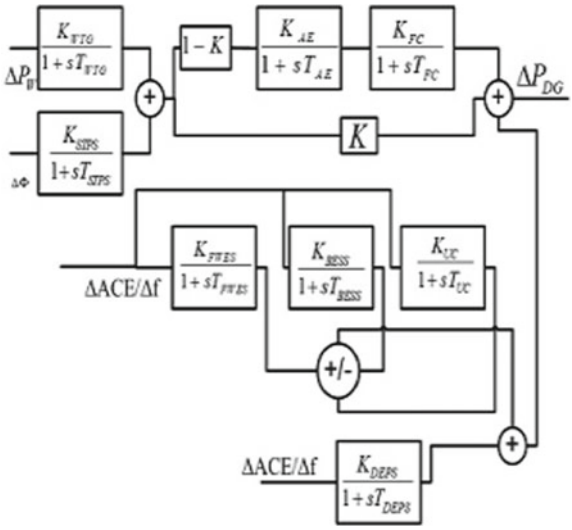


Fig. 2 Block diagram of the DG sources

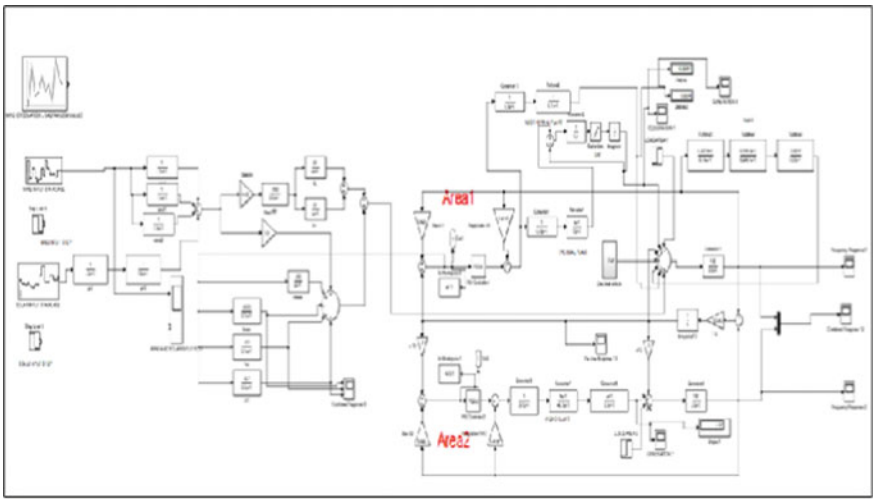


Fig. 3 Hybrid power system with DG sources and Electric Vehicle

2.1 State of Charge (SOC) Mode

Model of EV: Electrical circuit diagram of an EV, where a distribution grid is associated to the bidirectional converter presented in Fig. 1 to characterize the state of charge (Fig. 4).

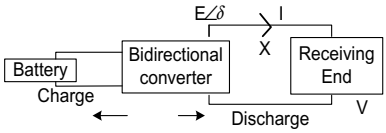


Fig. 4 Electrical diagram of EV

Active and main power delivered to the grid by the battery can be calculated as follows:

$$P = \frac{EV \sin \delta}{X} \tag{1}$$

where E, d, and X are the terminal voltages, V is the receiving end voltage, and I is battery current, angle between E and V, and line reactance among converter and the distribution node, correspondingly. It is presumed that the converter is controlled by the mentioned pulse width modulation (PWM) switching, with the consequence that EV's reactive power remains zero. Only inverter is linked to power system's network via a very tiny inductor (L). For EVs, a first-order lag function is approved for the frequency calculation. For modeling the EV, the first-order transfer function and time constant are as follows:

$$\text{TEV} = L/R \quad \text{and} \quad G = \frac{1}{1 + sT_{\text{EV}}} \tag{2}$$

where R denotes the reactor resistance and I denote the inverter losses. The SOC of EV may be determined using the following formula:

$$\text{SOC} = \frac{\text{Battery capcaity} - \text{output enrgy}}{\text{Battery capcaity}} \tag{3}$$

Lumped model of EVs. In this paper for EVs, a newer version of the lumped model (LM) published is employed as a resource. Since there are numbers of different EVs in each EV station, as a result, an analogous EV model can be considered, which gives each parameter of an EV with varying inverter capacities. Figure 5 depicts an equivalent EV model that will be utilized for LFC in the future. This model attempts to represent the behavior of a single electric vehicle battery. In a controllable condition, this EV model might be used to determine the total charging or discharging power. Parameters, such as gains and time constants, were taken from the ref [5] to simulate numerous sources and a standard system. They are given in Table 1.

2.2 PID Control Scheme

PID is a simple and easy to implement control scheme that is being used regularly in many applications. PID controller is denoted by

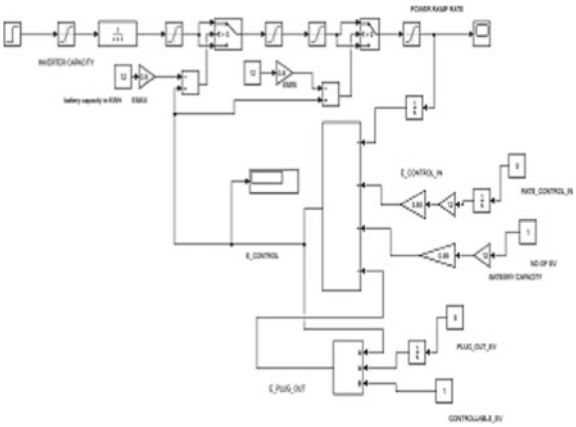


Fig. 5 Block diagram of SOC

Table 1 Hybrid system chosen parameters

Gains	Time constant (s)
$K_{\mathrm{WTG}} = 1$, $K_T = 1$, $K_S = 1$, $K_{\mathrm{AE}} = 0.002$, $K_{\mathrm{DEG}} = 0.0033$, $K_{\mathrm{FC}} = 0.01$, $K_{\mathrm{BESS}} = -0.0033$, $K_{\mathrm{FWES}} =$ -0.01, $\mathrm{K_{UC}} =$ -0.7	$T_{WTG} = 1.5$ s, $T_T = 0.3$ s, $T_S = 1.5$ s $T_{\mathrm{AE}} = 0.5$ s, $T_{\mathrm{FC}} = 4$ s, $T_{\mathrm{DEG}} = 2$ s $T_{\mathrm{BESS}} = 0.1$ s, $T_{\mathrm{FWES}} = 0.1$ s, $T_{\mathrm{UC}} = 0.9$ s
$K_p = 120$, $K_r = 5$, $K_g = \mathrm{K_t} = 1$	$T_p = 20$ s, $T_r = 10$ s, $T_g =$ 0.08 s, $T_t = 0.3$
$R = 0.416$, $B = 0.425$	$T_3 = 48.7$ s, $T_2 = 5$ s, $T_1 = 0.513$ s, $T_{\mathrm{w}} = 1$ s, $T^0 = 0.545$ s

$$G_{\mathrm{PID}}(s) = K_P + K_I/s + K_D s \tag{4}$$

where K_P, K_I, and K_D are PID parameters, for effective performance of PID, its parameters should be obtained optimally. To design PID parameters BBBC algorithm has been used and the detailed information of this algorithm can be referred from [7].

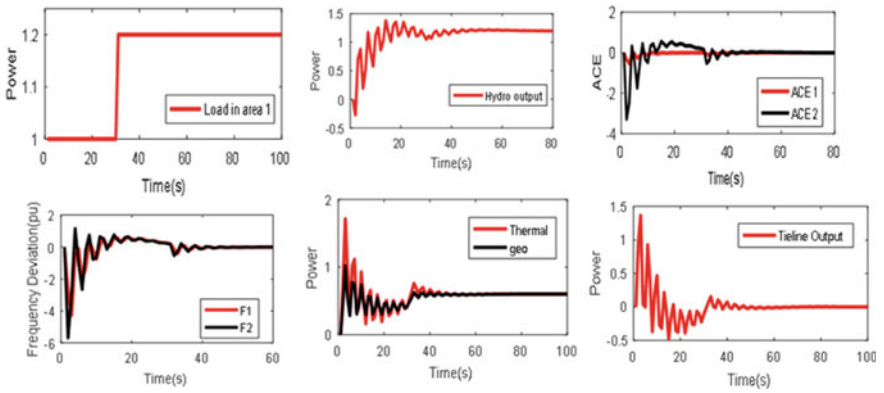


Fig. 6 Various responses of Case 1

3 Simulation Results

3.1 Case I: Only Thermal, Geo, Hydro Power Plants Have Been Considered to Contribute in LFC Scheme

In this case only thermal and Geo-thermal (area 1) and hydro (area 2) power plants are in operation. A step load of 1–1.2 have been taken into account when evaluating the designed control scheme's efficacy. Various generation changes and frequency deviation have been observed to respond to the given load disturbance. A good control approach eliminates the error and settle down the parameters at the desired level by satisfying various specifications. It is seen from the result obtained that the designed control scheme is performing good in terms of different time domain specification (Fig. 6).

3.2 Case II: Wind, Thermal, Geo, Hydro Power Plants Have Been Considered to Contribute in LFC Scheme

In this case wind power plant (area 1) with staircase input has been considered to check LFC effects on load perturbation. The sources have to react against this load perturbation to keep balance of the power. The responses of various sources (Plants) in accordance with the load to overcome the effect and settle down the parameters at desired values are given in Fig. 7.

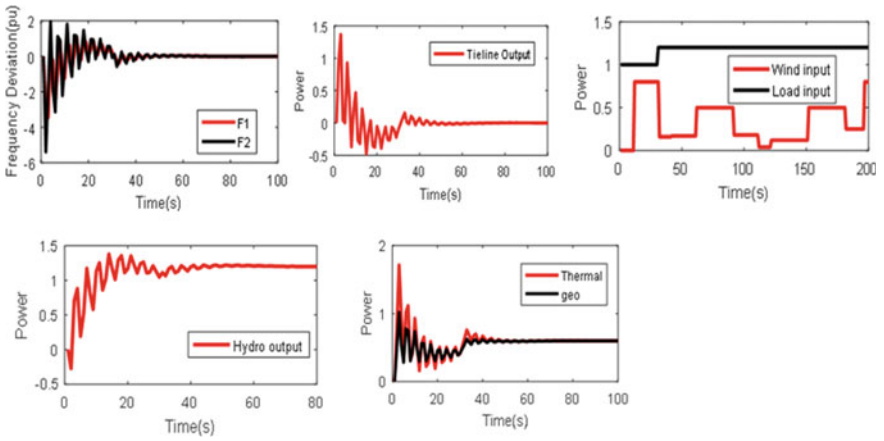


Fig. 7 Various responses of Case 2

3.3 Case III: Wind and Solar Power Plant with Hydro, Geo, Thermal Power Plants

In this case wind, solar, and conventional power plants are in operation and all are excited with step load. Step value of solar [0.38–0.20], wind [0.6–0.4], load [1–1.2], have been taken into account when evaluating the designed control scheme's efficacy and the observations have been given in Fig. 8. It is evident from the Fig. that frequency deviation and area control error vanishes at the steady state using the designed control scheme.

3.4 Case IV: Wind, Solar, and Conventional Power Plant with Random Load

In this case wind and solar input is taken staircase while a random load has been considered to check the effect of load perturbation on the performance of the designed control scheme. An optimal control scheme must respond accurately to eliminate any unwanted effects on the system. It is evident from the result that various responses are settling down to the desired value using the designed control approach. The same has been given in Fig. 9.

In this work, the performance of proposed control method was determined for four different cases. The same can also be seen from the Table 2, which shows exact point of maximum overshoot, maximum undershoot, and settling time for all considered cases.

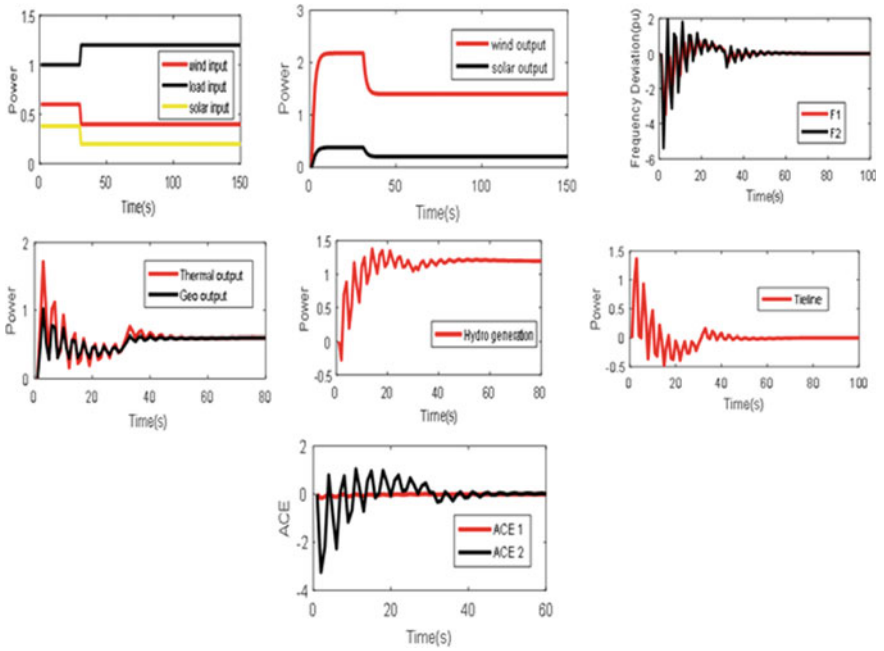


Fig. 8 Various responses of Case 3

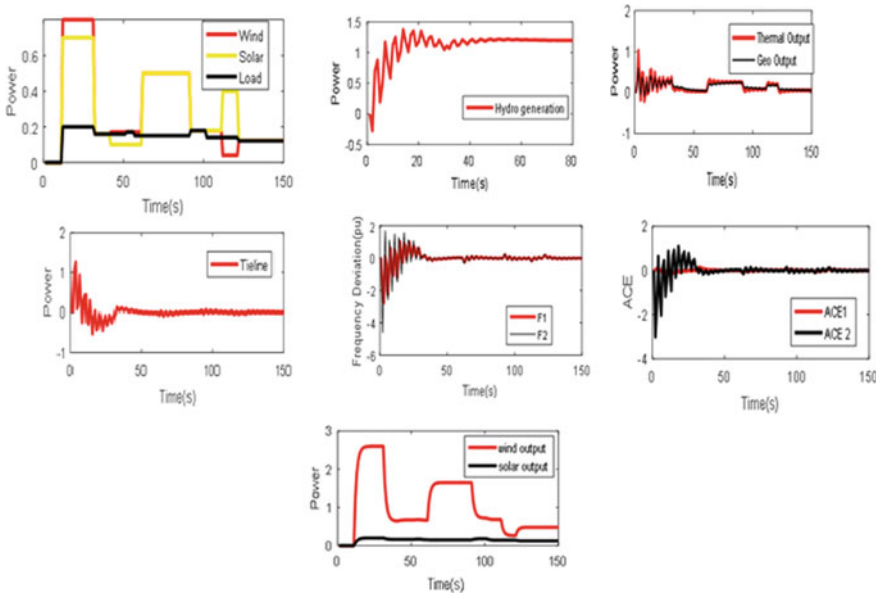


Fig. 9 Various responses of Case 4

Table 2 Different time domain characteristics for frequency deviation for all Cases

Cases		Maximum overshoot	Maximum undershoot	Settling time (s)
Case 1	F1	0.624	−4.2901	100
	F2	1.1751	−5.6983	100
Case 2	F1	1.0852	−3.4631	110
	F2	1.9666	−5.3953	120
Case 3	F1	0.9399	−3.4631	80
	F2	1.9668	−5.3953	95
Case 4	F1	1.0262	−2.7667	150
	F2	1.7034	−4.6189	150

4 Conclusion

Hybrid power systems are efficient energy and power generators. The following systems can be used alone or in combination with traditional sources to deliver more efficient electricity (in villages and isolated locations). Power systems can be installed closer to the point of use/distribution to save money on transmission and line costs. Hybrid power systems are typically thought of as less polluting, clean, and non-hazardous to the environment because they emit less pollutants and produce less waste than conventional power systems. Integration of environmental sources with conventional sources, on the other hand, may have effect on power system stability and also increases the power system's total complexity. As a result, to compensate for the negative effects on frequency/tie-line power, an optimal/robust control strategy should be in place.

References

1. Yesil E, Gzelkaya M, Eksin I (2004) Self-tuning fuzzy PID type load and frequency controller. Energy Convers Manage 45(3):377–390
2. Farahani M, Ganjefar S, Alizadeh M (2012) PID controller adjustment using chaotic optimisation algorithm for multi-area load frequency control. IET-Control Theory Appl 6(13):1984–1992
3. Khodabakhshian A, Hooshmand R (2012) Design of a robust load frequency control using sequential quadratic programming technique. Int J Electric Power Energy Syst 40(1):1–8
4. Sambariya DK, Prasad R (2013) Design of harmony search algorithm based tuned fuzzy logic power system stabilizer. Int Rev Electric Eng (IREE) 8(5):1594–1607
5. Ray P, Mohanty S, Kishor N (2011) Proportional integral controller based small signal analysis of hybrid distributed generation system. Energy Convers Manage 52:1943–1954
6. Kumar N, Tyagi B, Kumar V (2016) Multi area AGC scheme using Imperialist competition algorithm in restructured power system. Appl Soft Comput Elsevier 48:160–168
7. Kumar N, Tyagi B, Kumar V (2015) Optimization of PID parameters using BBBC for multi-area AGC scheme in deregulated power system. Turk J Electr Eng Comput Sci 24:4105–4116

8. Kumar N, Tyagi B, Kumar V (2017) Multi-area deregulated automatic generation control scheme of power system using imperialist competitive algorithm based robust controller. IETE J Res Taylor Francis 64:528–537
9. Tan W (2009) Tuning of PID load frequency controller for power systems. Energy Convers Manage 50(6):1465–1472
10. Debbarma S, Saikia LC, Sinha N (2014) Automatic generation control using two degree of freedom fractional order PID controller. Int J Electric Power Energy Syst 58:120–129
11. Ghosal SP (2004) Optimization of PID gains by particle swarm optimization in fuzzy based automatic generation control. Electric Power Syst Res 72:203–212
12. Elmas C, Yigit T (2007) Genetic Algorithm based on-line tuning of a PI controller for a switched reluctance motor drive. Electr Power Comp Syst 35(6):675–691
13. Li H, Liu S, Ji H, Yang D, Yang C, Zhao HCB, Hu Y, Chen Z (2014) Damping control strategies of inter-area low-frequency oscillation for DFIG-based wind farms integrated into a power system. Int J Electric Power Energy Syst 61:279–287
14. Z F, A R, Lone SA (2021) Load frequency control of multi-source electrical power system integrated with solar-thermal and electric vehicle. Int Trans Electric Energy Syst 31(7):1–20
15. Q Z, Y L, C L, Li CY (2019) Real-time adjustment of load frequency control based on controllable energy of electric vehicles. Trans Inst Measure Control 42(1):42–54

Lung Cancer Detection Using CT Scan Images

Kilari Veera Swamy, Jatavath Priyanka, Duggi Sri Venkata Likhitha, and Sreeja Veeragoni

Abstract Early diagnosis and treatment of Lung cancer, which is a life taking disease, can save many human lives. Doctors use CT scan images to detect the presence of lung cancer in any patient, but it is almost impossible for doctors to check every CT scan and accurately predict the presence of cancer cells. Computer aided detection or diagnosis will be helpful for the accurate identification of cancer cells. In the past, machine learning and image processing techniques have assisted computer aided detection. This work will explore various computer-assisted procedures (methods), identifying the best methodology now in use, describing its benefits and limitations. Hence, a new model is proposed which is more effective. In terms of accuracy, it outperforms existing techniques. In this work, Google Colaboratory is used to execute this work. It provides free GPU and CPU. In order to detect the cancer nodules, segmentation is done based on area. In order to avoid over-segmentation, a marker-controlled watershed algorithm is used. In order to deploy segmentation, the CT scan is first converted into a grayscale image. Then, noise is removed using Gaussian and Median filters followed by segmentation. Further, morphological transformations are done to remove white noises. Features such as area, center, and radius of the cancer nodules are extracted from the experimental results. Proposed technique is outperforming in detection of the cancer and stage of the cancer.

Keywords Gaussian filter · Median filter · Watershed algorithm · Morphological transformations · Lung cancer

1 Introduction

Most of the cancer deaths are caused due to lung cancer. It is hard to ascertain because it appears unexpectedly and only exhibits symptoms afterward. Early identification and treatment, on the other hand, can minimize the likelihood and fatality rate. Because CT imaging will reveal every suspected and unsuspected lung cancer

K. Veera Swamy (✉) · J. Priyanka · D. S. V. Likhitha · S. Veeragoni
Department of Electronics and Communication Engineering, Vasavi College of Engineering, Hyderabad, India
e-mail: k.veeraswamy@staff.vce.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_36

nodule, it is a trustworthy tool for lung cancer screening and diagnosis. However, because the anatomical structure and intensity of the lung in CT scan images vary, there is always the possibility of human error, i.e., radiologists and physicians may commit mistakes that make it harder to identify malignant nodules/cells. Computer Aided Diagnosis has recently emerged as a powerful and promising technique to aid radiologists and physicians in effectively detecting cancer. Many techniques have been created in the past by various researchers [1–3]. However, some systems have to be improved as they have a detection accuracy that is not satisfactory. All the efforts made by various researchers are to tend the accuracy to a 100%. In order to detect and categorize lung cancer, machine learning and image processing techniques were used [4–7]. Stages include preprocessing, segmentation, morphological transformations and feature extractions are used in the proposed technique. The classification is based on presence of nodule or not. Stage of the cancer depends on number of nodules and area of each nodule. Generally, the area of a cancer nodule ranges between 10 and 400 mm^2. Draw the circle if the above condition is satisfied. So, using segmentation [8], all the areas which fall in between 10 and 400 mm^2 will be identified and based on the maximum area spread which will be calculated by the algorithm; stage of cancer will be determined. Watershed algorithm is used for segmentation.

2 Literature Survey

Aggarwal et al. [9] devised a classification system for nodules and normal lung anatomical structure. Statistical, geometrical, and gray level parameters are extracted using this method. The method has an accuracy of 84%, a sensitivity of 97.14%, and a specificity of 53.33%. Simple segmentation techniques were used for classification. Although cancer nodules are being detected, its accuracy is still unacceptable. Jin et al. [10] employed a convolution neural network as classifier in their CAD system in order to detect lung cancer. The method has an accuracy of 84.6%, sensitivity of 82.5%, and specificity of 86.7%. Although the cost of implementation was lowered, the precision remained still insufficient. For segmentation, Sangamithraa and Govindaraju [11] used the K-means unsupervised learning technique. The system has a 90.7% accuracy rate. The system created by Roy et al. [12] used a fuzzy interface system and an active contour model to recognize lung cancer nodules. This technique employs gray transformation to improve visual contrast. Features such as mean, entropy, area, correlation, and major and minor axis length are used to train the classifier. It does not differentiate between healthy and diseased a cancer perfectly, which is a flaw in this system. Ignatious and Joseph [13] used watershed segmentation to implement lung cancer detection system. In order to improve the image quality, Gabor filter is used in the preprocessing stage. This system has 90.1% accuracy, which is higher than the model that uses neural fuzzy model as well as a region-growing approach for segmentation. This model's accuracy is likely to improve to a satisfactory level with only a few adjustments and contributions. SVM-based cancer detection technique is also available in the literature [6]. Gonzalez and

Ponomaryvo [14] presented a methodology for distinguishing between healthy and diseased lung cancer. To compute the Region of Interest (ROI); the system makes use of the Hounsfield Unit (HU) and priori data. It requires prior information about ROI in order to categorize the nodules as healthy or diseased which is a drawback. This model can be used as a reference for our proposed system. The following limitations are observed: (1) Only a few traits have been retrieved for the nodules. (2) Perfect classification as healthy or deceased is missing. (3) Perfect identification of cancer stage is not determined.

3 Proposed Method

Gabor filter [13] did not give satisfactory results in filtering. Hence, in this work Gaussian and median filters in the preprocessing stage are used to remove noise. Further, we will perform a marker-controlled watershed algorithm in order to segment the CT scan. Then, morphological transformations which are erosion and dilation are deployed in order to eliminate white noise. Features such as area, center, and radius are extracted to identify the cancer and stage of the cancer. Block diagram of the proposed method is shown in Fig. 1.

3.1 CT Scan Images

The abbreviation CT stands for Computerized Tomography. It uses computer processing to construct cross-sectional images (slices) of the bones, blood arteries, and soft tissues inside the body from a compilation of X-ray images obtained from various angles around the body. The information offered by CT scan images is more sophisticated than that offered by standard X-rays.

3.2 Preprocessing Stage

To prepare image data for model input, preprocessing is required. We can reduce unwanted distortions and boost some qualities that are important for the application we're working on by preprocessing. In preprocessing stage, Gaussian filter and median filter are used in this work. The high frequency components are eliminated

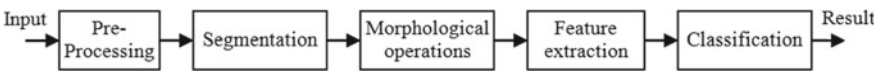


Fig. 1 Block diagram of the proposed method

as it is a low pass filter. They do not overshoot a step function input when reducing rise and fall time. Gaussian filter has the least conceivable group delay. It is widely regarded as the best time domain filter. The Median blur operation works in the same way as the other averaging procedures. The image's focal element is replaced with the median of all the pixels in the kernel region. The edges are processed while the noise is removed in this operation. Median of all pixel values is generated by the median blur filter. Edges and round corners of the image are preserved in this work.

3.3 Watershed Segmentation

In image recognition systems, segmentation [8] is a crucial step as it extracts the things of our interest for other processes like description and identification. Image segmentation is used for categorizing pixels in an image. One of the region-based image segmentation methods is Watershed segmentation. In this segmentation method the entire input image is divided into different regions. Generally, in medical CT scan images, X-rays a lot of noise is present which causes over-segmentation during Watershed segmentation. The input image should be filtered to avoid over-segmentation. The concept is based on the visualization of a three-dimensional image: intensity versus two spatial coordinates.

3.4 Morphological Transformation

Processing an image using morphological transformation is a set of non-linear processes that deal with the structure or architecture of picture characteristics. It requires input data: one of these is the original image, and the other is a kernel that determines the operation's nature. Erosion and Dilation are two primary morphological operations. Mathematically erosion is represented by $A \ominus B$. Mathematically dilation is denoted by $A \oplus B$. A pixel in the original image will be considered 1 if at least one of its pixels under kernel is 1. It increases white region in image.

3.5 Feature Extraction

In this model, features of cancer nodules such as center, radius and area are used. The cancer nodules are in an irregular shape of a circle. We find the number of contours in the image and then the area of each contour. After this, the area of each contour is in the range of specified limit or not. The range is declared from 10 to 400 because the minimum diameter of cancer nodule is 3 mm. Area of cancer nodule is approximately 10 mm square. The maximum diameter of cancer nodule is 30 mm, i.e., area of the cancer nodule is 400 mm square. If the above condition is satisfied,

we find the minimum enclosing circle of contour from which we can get the radius and center. By using the above center and radius we will draw the circle on the image which represent the cancer nodule. In this way we find the features of cancer nodules present in the image. These numerical values are concluded with lot of experiments with the known data set.

4 Experimental Results

Figure 2a is lung CT scan image which is given as input to detect the cancer nodules and identify the stage of cancer. Depending on the size of nodules present in the CT scan image, the stage of cancer will be displayed. Figure 2b shows the Gaussian blurred image where the noise and unimportant details are removed using Gaussian filter. The Gaussian blurred image is then median blurred using median filter. Median filter reduces the variation of intensity from one pixel to another which is shown in Fig. 2c. It also reduces the noise but keeps the edges relatively sharp.

Each pixel value in the image is compared with the threshold limit if pixel intensity below threshold limit then the pixel intensity is assigned to 255 whereas if pixel intensity is above threshold limit then the pixel intensity is assigned to 0. The threshold binary inverted image is shown in Fig. 3a.

Erosion removes the borders of foreground objects which attempt to keep the foreground white and shrinks the image. Figure 3b shows the eroded image. The

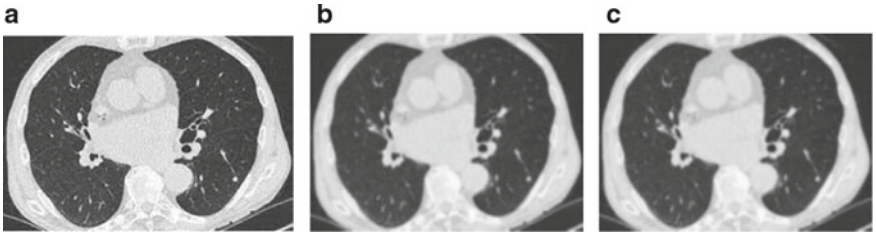


Fig. 2 **a** CT Scan image. **b** Gaussian filtered image. **c** Median filtered image

Fig. 3 **a** Threshold image. **b** Eroded image. **c** Dilated image

eroded image is then dilated. In the process of dilation, the area of objects increases and removes the noise. Figure 3c shows the output image after dilation process.

Watershed segmentation is followed by dilation to build proper dams during segmentation. Marker-based segmentation processes to isolate different objects and identify meaningful information is implemented. Figure 4a shows the watershed segmented image. All the detected cancer nodules are marked on the input CT scan image as shown in Fig. 4b. As there are more than one cancer nodules of area less than 100, it is lung cancer stage 2. The number of abnormal contours in the image are identified. These are referred as cancer cell IDs. Based on the size and number of nodules the stage of cancer will be displayed. Table 1 shows the number of nodules and size of each nodule. Number of nodes detected are 4. Areas of four nodes are 132, 147, 247, and 170 mm, respectively. Hence, cancer stage is 2.

The CT scan shown in Fig. 5a is the input image to our method and Fig. 5b shows the output image. This input image has cancer stage 1 because there is only one cancer nodule of area less than 100.

The CT scan shown in Fig. 5c is the input image to our method and Fig. 5d shows the output image. This CT scan image has lung cancer stage 3 because there are five nodules with area greater than 100 and few nodules area less than 100. The CT scan shown in Fig. 6a is the input image to our method and Fig. 6b shows the output image. This input CT scan image has Lung cancer stage 4 as there are more than five nodules which have area greater than 100.

A total of 50 CT scans were used to test the proposed work out of which 46 were correctly detected. There were 6 CT scans which did not have cancer, 5 CTs having stage 1 cancer, 11 CTs having stage 2 cancer, 15 CTs having stage 3 cancer, and 13 CTs having stage 4 cancer. To calculate the accuracy confusion matrix is used.

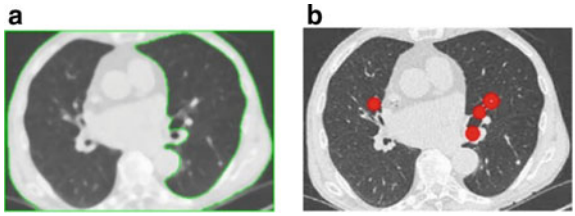


Fig. 4 **a** Watershed segmented image. **b** Cancer detected cells

Table 1 Details of detected nodules and cell range

Cell ID	Cell range (Sq.mm)
7	62
10	132
13	147
15	247
16	170
21	42

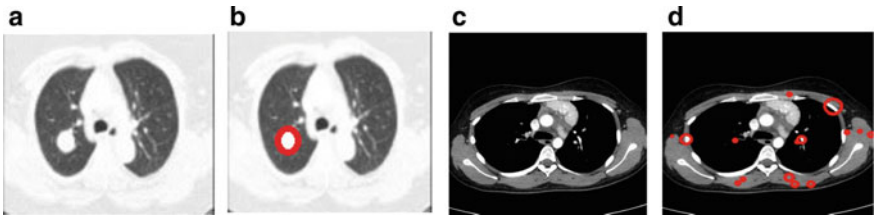


Fig. 5 **a** Input image. **b** Output image. **c** Input image. **d** Output image

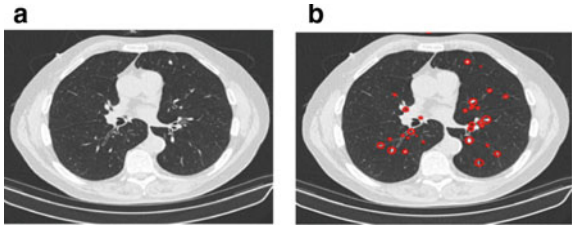


Fig. 6 **a** Input image. **b** Output image

Table 2 Detection of cancer in 50 CT scan images

Cancer details	CT scans available	CT scan detected correctly
No cancer	6	6
Stage 1 cancer	5	4
Stage 2 cancer	11	10
Stage 3 cancer	15	14
Stage 4 cancer	13	12

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives. Actual and prediction are the same class then this number is denoted as TP. Confusion matrix is used to calculate accuracy using the Eq. 1. Accuracy for the 50 images is 92% (Table 2).

5 Conclusion

In the existing models there is no appropriate accuracy outcome and it will not segregate the input CT scan image into the different stages of cancer effectively. So, a new model is proposed which will identify the nodules in the input image with cancer using image segmentation with watershed algorithm. Based on the detected cancer nodules, the input image is classified within the 4 stages of cancer, i.e., Cancer

Stage 1, Stage 2, Stage 3, Stage 4, and No cancer. The accuracy of the proposed model is giving 92% accuracy when it comes to detecting cancer in CT scan images. Proposed method is able to detect no cancer images with 100% accuracy. Further, this method is able identify the stage of the cancer effectively. In future, this technique with machine learning classification on many more images will be implemented to increase the accuracy. X-ray, PET, and MRI images of lungs to find out which type of image shows a better outcome can also be explored. Details of cancer nodules such as eccentricity, intensity, Average intensity, Perimeter are also important. Convolution Neural Networks with kernels can be explored to make more automation and increase the efficiency further. The proposed method can be explored to detect some other types of cancers such as skin cancer, brain tumor, and breast cancer.

Acknowledgements The authors would like to thank the management of the Vasavi College of Engineering, Ibrahimbagh, Hyderabad, India for their constant encouragement, providing laboratory facilities, financial aid, and support to complete this work.

References

1. Chapaliuk B, Zaychenko Y (2018) Deep learning approach in computer-aided detection system for lung cancer. In: IEEE First International conference on system analysis & intelligent computing, pp 1–4
2. Fang T (2018) A novel computer-aided lung cancer detection method based on transfer learning from GoogleNet and median intensity projections. In: IEEE International conference on computer and communication engineering technology, pp 286–290
3. Hoque A, Farabi AKMA, Ahmed F, Islam MZ (2020) Automated detection of lung cancer using CT scan images. In: IEEE Region 10 symposium, pp 1030–1033
4. Koteswara Rao M, Veera Swamy K, Anitha Sheela K (2018) Advanced machine learning discriminant analysis models for face retrieval system. In: 2nd International conference on I-SMAC, pp 609–613
5. Veera Swamy K, Vineetha K (2021) Performance comparison of retinal image enhancement techniques using wavelet transforms. Int J Analog Integr Circuits 1–7
6. Nadkarni NS, Borkar S (2019) Detection of lung cancer in CT images using image processing. In: 3rd International conference on trends in electronics and informatics, pp 863–866
7. Ashwanth B, Veera Swamy K (2020) Medical image fusion using transform techniques. In: 5th International conference on devices, circuits and systems, pp 303–306
8. Neetu Ch, Veera Swamy K (2019) Performance assessment of various thyroid image segmentation techniques with consistency verification. J Adv Res Dyn Control Syst 11(02-Special Issue):1299–1309
9. Aggarwal T, Furqan A, Kalra K (2015) Feature extraction and LDA based classification of lung nodules in chest CT scan images. In: International conference on advances in computing, communications and informatics, pp 1189–1193
10. Jin X, Zhang Y, Jin Q (2016) Pulmonary nodule detection based on CT images using convolution neural network. In: 9th International symposium on computational intelligence and design, pp 202–204
11. Sangamithra PB, Govindaraju S (2016) Lung tumour detection and classification using EK-mean clustering. In: International conference on wireless communications, signal processing and networking, pp 2201–2206

12. Roy TS, Sirohi N, Patle A (2015) Classification of lung image and nodule detection using fuzzy inference system. In: International conference on computing, communication & automation, pp 1204–1207
13. Ignatious S, Joseph R (2015) Computer aided lung cancer detection system. In: Global conference on communication technologies, pp 555–558
14. Rendon Gonzalez E, Ponomaryov V (2016) Automatic lung nodule segmentation and classification in CT images based on SVM. In: International Kharkiv symposium on physics and engineering of microwaves, millimeter and submillimeter waves, pp 1–4

A Wideband Microstrip Line-Based Balun Structure for High Power Amplifier Applications

Manishankar Prasad Gupta, Pradeep Gorre, Sandeep Kumar, and Hanjung Song

Abstract This paper proposes a balun matching technique to achieving a high output power and wide bandwidth. The proposed structure includes microstrip transmission line-based even and odd mode-matching circuits. A three-port unipolar microstrip line is designed to transform the balanced load termination to 50 Ω unbalanced port impedance. The proposed network design is based on real symmetrical four port network with open ended transmission line is inserted between the middle of the structure. To improve the isolation, transmission coefficient parameter and match the 50 Ω termination, a resistive network is inserted between the two balanced ports. The proposed structure is simulated in Keysight Technologies Advanced Design System (ADS), fabrication is done by using 0.51 mm RT Duriod substrate alignments. To verify the design concept, first of all, a wideband microstrip matching technique is designed and characterized at the frequency of L_5 band (1.17 GHz). Then a prototype of microstrip transmission line-based wideband balun matching circuit is designed and fabricated. Analytical design equations have been derived for the even mode as well as odd mode techniques which satisfied the results. The proposed balun could overcome power loss mechanism over traditional transmission line structures and can utilize for high power application.

Keywords Isolation · Microstrip lines · Broadband matching balun · Even and odd mode techniques · UHF-band · L-band and S-band

M. P. Gupta · P. Gorre · S. Kumar (✉)
Department of Electronics and Communication, National Institute of Technology, Surathkal, Karnataka, India
e-mail: fedrer.engg@gmail.com

H. Song
Department of Nanoscience and Engineering, Centre of Nano-Manufacturing, Inje University, Gimhae, South Korea
e-mail: hjsong@inje.ac.kr

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_37

1 Introduction

A balun was originally called balanced to unbalanced, but nowadays, it is called a balancing unit. Balun is an electrical device that does conversion between balanced and unbalanced signals, i.e., Balun is a radio device for converting from a balanced line to an unbalanced line and is usually used at high frequencies. A balun can take many forms and it may include a device that also transforms impedances. Transformer balun can also be used to connect lines of different impedances by using magnetic coupling. Generally, a balun consists of two wires namely, primary and secondary wire and a toroid core. The current in the primary wire generates magnetic fields in the core, which in turn includes an electrical field in the secondary wire. But, auto-transformer-based balun has only one coil or two coils that have electrical connections [1–3]. Balun made with auto-transformer winding are also renamed as voltage balun, but these types of baluns produce balanced output voltages, but it is not necessarily balanced current. A transmission line-based balun is a simple form of transformer. Sometimes, transmission line-based balun is also called current balun because it ensures equal currents in both sides of its output, but not necessarily equal voltage. So, to remove this difficulty, magnetic coupling is combined with a transmission line (also called electromagnetic coupling) with a very small ferrite core, but transmission line-based balun can work reliably up to MHz frequency. So microstrip line is used to make balun which is dispersive and it has no cut-off frequency. It has good isolation between adjacent traces. To realize the broadband applications, the amplifier design strategies transferred from the lumped circuit to distributed circuit (basically based on microstrip lines) [4]. In the MSL-based amplifiers, the parasitic issues, as well as the noise level of the specified circuits, are minimized by matching the load with input by using even and odd mode techniques, which in turn reduces the nonlinearity of the device [5–7]. In general, this is too achieved by canceling the mismatch errors at either even input or odd matching network. Balun is a three-port device to convert a single-ended signal into a double-ended signal and vice-versa [8–11]. There are many types of microstrip lines-based baluns that are categorized into Wilkinson power divider balun, Marchand balun and branch line configuration balun, and many more. In Marchand balun, a coupled line is applied to enhance the bandwidth, but it requires a high precision fabrication process [12–14]. To overcome this problem, Marchand balun with a vertically installed structure is proposed to achieve good isolation and port impedance matching [15]. Wilkinson power divider balun is designed to achieve the wideband and planar layout [16]. But, the above two methods produced DC short and degrade the bandwidth performance, i.e., the above two did not use for input impedance matching for the balanced port as well as isolation. Overall, the conventional microstrip line is used to overcome the DC short as well as the performance of the balun. The advantages of the microstrip line-based balun are not required any isolation between balance ports. Therefore, this work proposes a microstrip line-based balun using an even and odd mode-matching technique and achieves a wideband from 0.5 to 2.5 GHz. The work is organized as follows: Sect. 2 describes the structure of the circuit implementation of the proposed

balun. Section 3 presents the simulations and results of the stated matching circuit topology. Finally, the conclusion is followed in Sect. 4.

2 Balun Design and Consideration

Figure 1 shows the proposed balun matching circuit which is symmetrical. The current of the feed transmission line is minimized by using a symmetrical type of balun, which is why usually takes symmetrical type, i.e., widths and lengths of the upper half of the transmission line are the same as for the lower half. The only difference is the open circuit microstrip line that is denoted by MLOC38. The main function of this open stub is to provide a short circuit for the resistor. The several advantages of this balun matching circuit are: (1) Complex balanced load impedance can be easily transferred to standard 50 Ω unbalanced impedance (2) All ports are matched and more than 10 dB isolation between balanced ports (3) No short stub is required to avoid the DC power loss because both resistor and open stub act as short stub circuit which reduced the balun complexity proposed. This balun is treated as a four port network with an open fourth port at node e, while even and odd mode-matching techniques have been used to analyze the complete balun circuit. In order to find out the matching technique operation, the even mode forward transmission $T_{\text{even}} = 0$, odd mode input impedance $Z_{\text{in}} = 2Z_{\text{RF}}$, even mode impedance $Z_{\text{even}} = 0$ and even mode output return loss $S_{22\text{even}} = 0$. Table 1 shows dimension values of the upper half of the transmission lines.

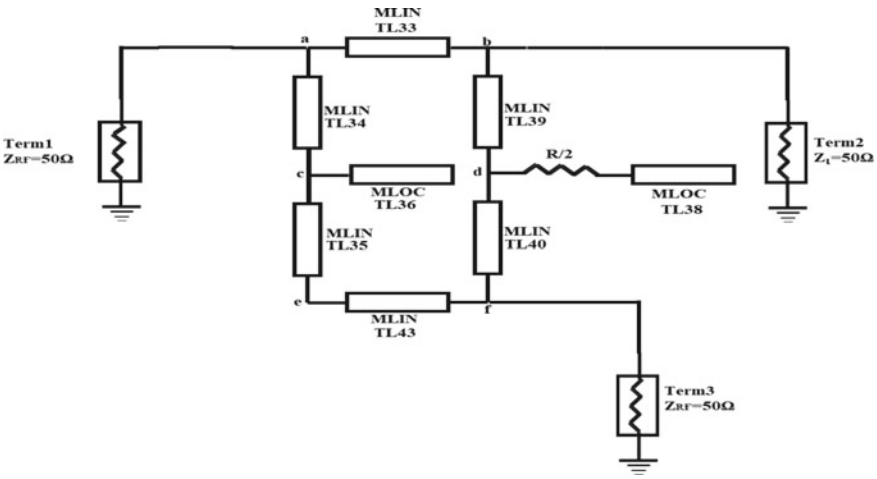


Fig. 1 Schematic of proposed balun matching circuit

Table 1 Components value of transmission line used in balun

Dimensions	W (mm)	L (mm)
TL33	1.1224	0.2
TL34	0.938	0.5
TL36	0.312	1.25
TL38	0.312	3.75
TL39	0.6255	5.66

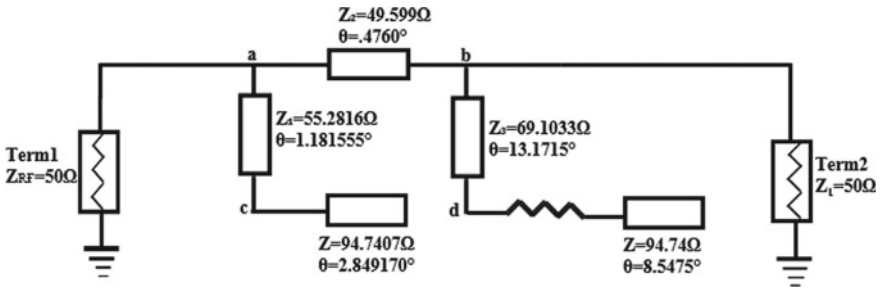


Fig. 2 Equivalent circuit of even mode analysis

2.1 Even Mode Circuit Analysis

In this subsection, the even mode circuit and analysis would be discussed. The even mode equivalent circuit of the proposed balun is depicted in Fig. 2, which is achieved by an open circuit of the nodes such as c and d at a symmetrical line. The open stub at node c combines with a series transmission line (TL34) to represent short at node a while open stub (TL38) provides a short circuit for the resistor R. To make the analysis simple, all the impedances are converted into the admittances. So, the admittance equation at node b can be written as,

$$Y_{\text{sb}} = -jY_2 \cot\theta_2 \tag{1}$$

$$Y_{\text{pb}} = Y_3 \frac{G + jY_3 \tan\theta_3}{Y_3 + jG \tan\theta_3} \tag{2}$$

where transconductance $G = 1/R$ and $Y_3 = 1/Z_3$

2.2 Odd Mode Circuit Analysis

The odd modes such as c and d matching network of the proposed balun are obtained by shorting the nodes at symmetrical line. The odd mode-matching circuit of the

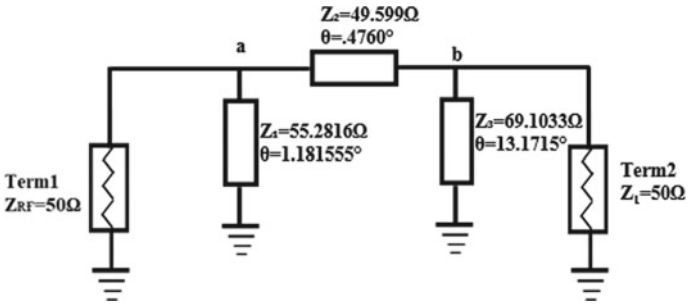


Fig. 3 Equivalent circuit of odd mode analysis

proposed balun is shown in Fig. 3. The admittance at node b for the odd mode are given by the following equations,

$$Y_b = Y_L - jY_3 \cot\theta_3 \tag{3}$$

$$Y_a = Y_2 \frac{Y_b + jY_2 \tan\theta_2}{Y_2 + jY_b \tan\theta_2} \tag{4}$$

To satisfy $Z_{\text{in}} = 2Z_{\text{RF}}$ with 50 Ω input port termination at node a, then

$$\text{Real}\,[Y_3] = 1/2Z_{\text{RF}} \,\text{and}\, \text{Imaginary}\,[Y_3] = -X$$

where $X = -Y_1 \cot\theta_1$.

3 Results and Discussion

In this section, the results and discussion about balun structure are described. The simulation results are carried to analyze the representation of balun as a broadband matching circuit by odd and even mode techniques. The results are obtained by using Keysight Technologies Advanced Design System (ADS) v.2019 tool. The simulated S-parameters of the proposed balun in even and odd modes are shown in Figs. 4, 5, and 6, respectively. Figure 4 shows the return loss and transmission coefficient plots of combined matching technique while Figs. 5 and 6 illustrates return loss and transmission coefficient plots for even and odd mode techniques, respectively. It is observed that input and output return loss and transmission coefficient values of the matching techniques are significantly depends on impedance, Z and electrical strength, θ of each transmission line. For various combinations of Z and θ, the matching technique provides more degree of freedom to design wide bandwidth, namely, $Z_2 = 49.599\ \Omega$, $\theta_2 = 0.4760$, $Z_3 = 69.1033\ \Omega$, $\theta_3 = 13.1715$, $Z_1 = 55.2816\ \Omega$, and $\theta_1 = 1.181555$, respectively. By increasing the value of θ_2 up to 1, we noticed a bandwidth

degradation which results a narrow band operation. Additionally, the forward and reverse transmission coefficients are also decaying. Hence, there is always a trade-off between return loss and transmission coefficients for specified values of impedance and electrical strength. Figure 7 depicts the layout of the proposed balun, where three ports have been taken, out of three; one port has been used for input while remaining two for the output ports. Rogers RT Duriod 5880 substrate with 0.51 mm thickness, permittivity of 3.68 and loss tangent of 0.0004 is considered for the proposed balun matching network design.

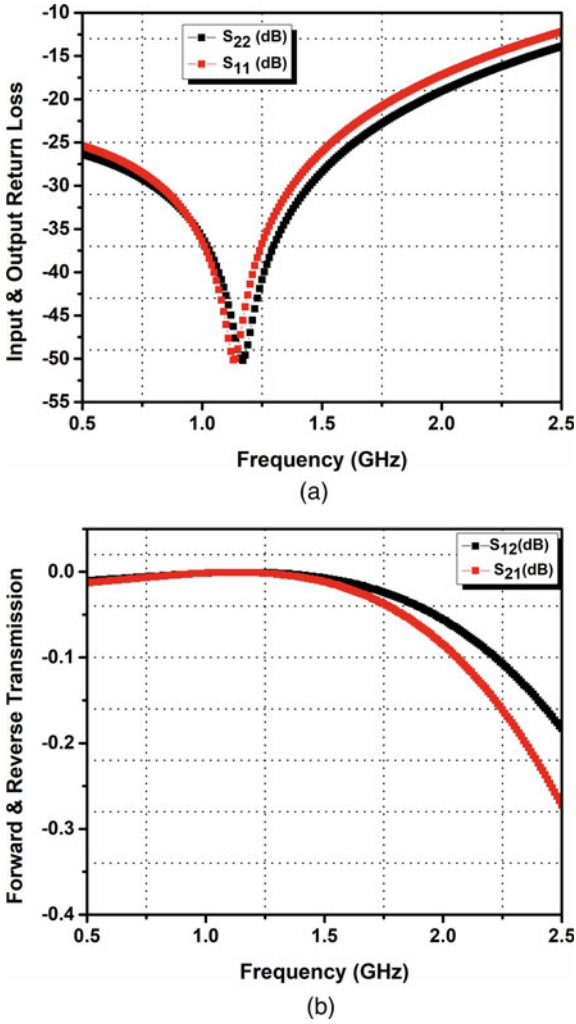


Fig. 4 **a** Input and output return loss versus frequency and **b** transmission coefficient versus frequency

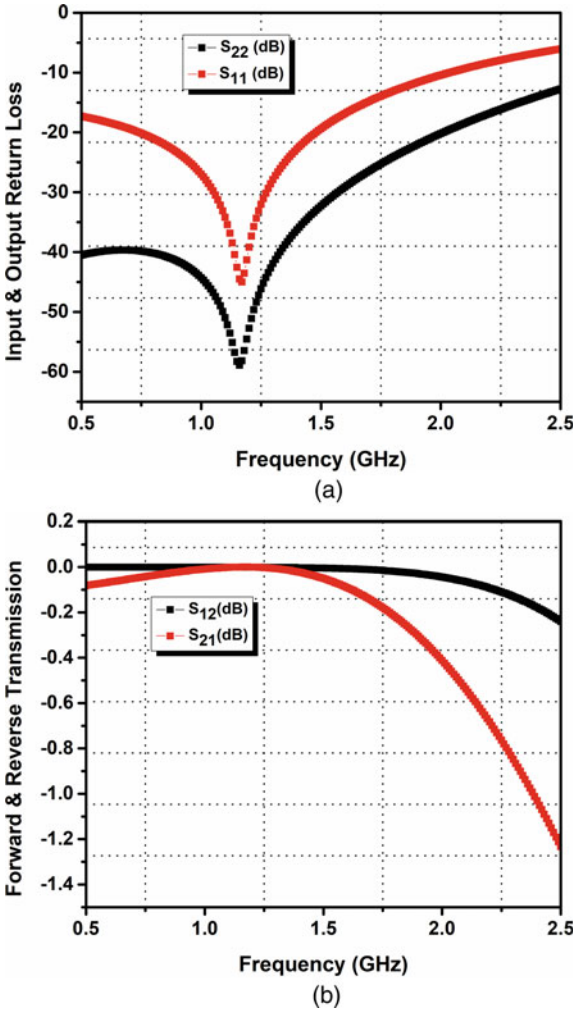


Fig. 5 Even mode with its **a** Input and output return loss and **b** transmission coefficient

4 Conclusion

This paper concludes a detailed analysis of the novel microstrip-based using open stubs and one lumped component which provides a short circuit. The complete design formula for this balun matching network is taken based on even and odd mode topology. A range of frequency bands (UHF-band, L-band, and S-band) can be realized by this balun matching network for the application of a low noise amplifier. It ensures that all properties related to balun are coming with fruitful results. So, this matching technique can be used in high-power amplifier applications with acceptable performance. Measurement results verify the design concept.

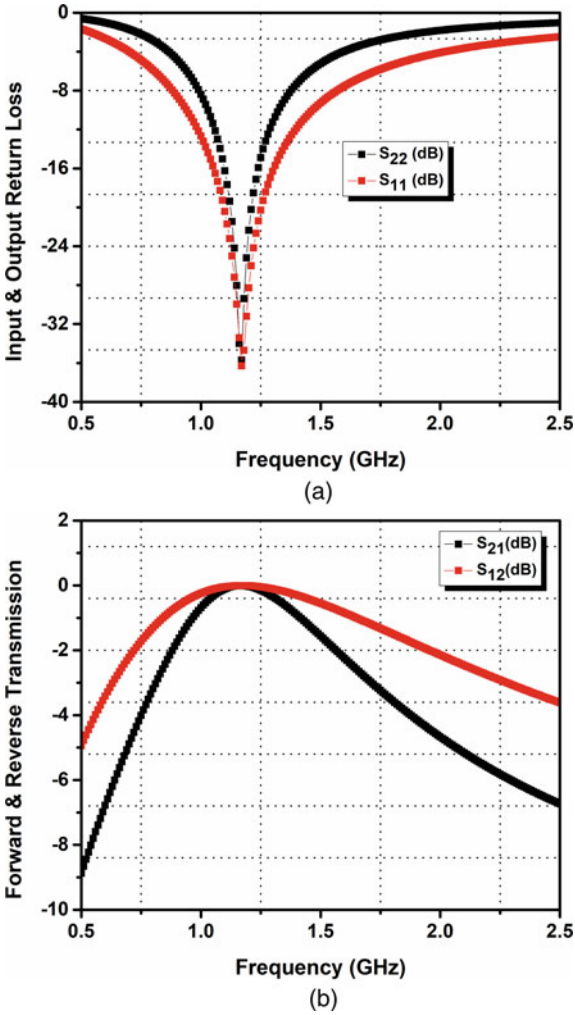


Fig. 6 Odd mode with its **a** Input and output return loss and **b** Transmission coefficient

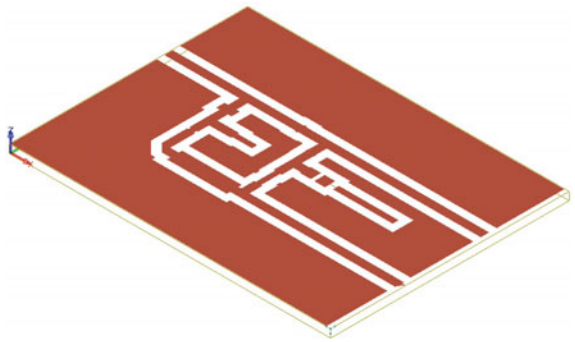


Fig. 7 Layout of proposed balun matching technique

Acknowledgements This work was supported by the SERB-DST Research Project, Government of India, Grant/Award Number: EEQ/2020/000144.

References

1. Photonics Dictionary. [Online]. Available: https://www.photonics.com
2. Baluns. What they do and how they do it (W7EL). Available: http://www.eznec.com
3. Zhang H, Peng Y, Xin H (2008) A tapped stepped-impedance balun with dual-band operations, vol 7, pp 119–122
4. Gupta R, Kairatova S, Hashmi M, Nauryzbayev G (2021) A dual band balun architecture with unequal port-terminations. EuMA, pp 848–851, 12–14 January (2021)
5. Zhou M, Shao J, Arigong B, Ren H, Ding J, Zhang H, Design of microwave baluns with flexibles structures, pp 1509–1531
6. Hedayatullah Maktoomi Md, Ren H, Klein V, Wilson R, Arigong B (2020) A wideband isolated real to complex impedance transforming uniplanar microstrip line balun for push-pull power amplifier, vol 68, No 11, 0018-9480
7. Wu T-Z. Transmission lines and components. Department of Electrical Engineering, National Taiwan University, Chap. 4
8. Smith RM, Lees J, Tasker PJ, Benedikt J, Cripps SC (2012) A novel formulation for high efficiency modes in push-pull power amplifiers using transmission line baluns. IEEE Microw Wireless Compon Lett 22(5):257–259
9. Saad P, Thorsell M, Andersson K, Fager C (2013) Investigations of push-pull microwave power amplifiers using an advanced measurement setup. IEEE Microw Wireless Compon Lett 23(4):220–222
10. Stameroff AN, Pham AV, Leoni RE (2010) High efficiency push-pull inverse Class F power amplifier using a balun and harmonic trap waveform shaping network. In: IEEE MTT-S International microwave symposium. Anaheim, CA, USA, pp 521–525
11. Stameroff AN, Ta HH, Pham AV, Leoni RE (2013) Wide-bandwidth power combining and inverse Class F GaN power amplifier at X-band. IEEE Trans Microw Theory Technol 61(3):1291–1300
12. Aliqab K, Hong J (2019) Fully embedded ultra-wideband multi-layer balun into organic packaging substrate. IET Microw Antennas Propag 13(3):322–325
13. Yang L, Zhu L, Choi WW, Tam KW, Zhang R, Wang Z (2018) Wideband balanced-to-unbalanced band pass filters synthetically designed with Chebyshev filtering response. IEEE Trans Microw Theory Technol 66(10):4528–4539
14. Sen Ang K, Robertson ID (2001) Analysis and design of impedance transforming planar Marchand Balun. IEEE Trans Microw Theory Technol 49(2):402–406
15. Ahn HR, Tentzeris MM (2017) Novel generic assymetric and symmetric equivqlent circuits of 90° coupled transmission line sections applicable to Marchand Balun. IEEE Trans Microw Theory Technol 65(3):746–760
16. Zhang ZY, Guo YX, Chuen Ong L, Chim MYW (2005) A new wide-band planar balun on a single-layer PCB. IEEE Microw Wireless Compon Lett 15(16):416–418

Principal Component Analysis of LISS—III Images Using QGIS

V. Vijayalakshmi, D. Mahesh Kumar, and S. C. Prasanna Kumar

Abstract Processing and Classification of Remote Sensing Data is essential for various purposes such as knowing the earth's surface and its terrain, natural resource management, natural disaster management, agriculture, soil fertility, reclamation of waste and degraded lands. Generating the Land Use Land Cover (LULC) map of the study area is the first and basic step to achieve most of the above objectives. LULC map is also useful to perform change detection studies, to know the terrain of the earth's surface, to monitor the growth of forests, to know the aftermath of a natural disaster such as forest fire, volcano, floods. Principal Component Analysis (PCA) takes a significant role in generating the LULC map as it gives the advantage of reducing the image dimension thus reducing system and time requirements for data analysis. In this article PCA has been performed using QGIS software which is an open source software. Along with data compression, dimensionality reduction and simplicity in data analysis, PCA also supports in representing the data over the feature vector, i.e., where there is maximum variation in data.

Keywords Remote sensing · Satellite imagery · Land Use Land Cover (LULC) · Geographic Information System (GIS) · Principal Component Analysis (PCA) · Quantum geographic information system (QGIS)

1 Introduction

Principal Component Analysis is a very significant phase in multivariate statistical analysis and the most widely used technique for data compression [1]. It is a non-parametric mathematical algorithm which involves reducing the dimensionality of images in large datasets. Multispectral and Hyperspectral images pose a complexity

V. Vijayalakshmi (✉) · D. Mahesh Kumar · S. C. Prasanna Kumar
Department of Electronics and Instrumentation, JSS Academy of Technical Education, Visvesvaraya Technological University, Belagavi, Karnataka, India
e-mail: v.vijayalakshmi80@gmail.com

S. C. Prasanna Kumar
e-mail: prasannakumar@rvce.edu.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_38

during data analysis due to the presence of a large number of bands which in turn contributes a large number of variables for analysis. This issue can be resolved by implementing PCA which scales down these large variables into much smaller ones and thus simplifies the process of image analysis. PCA generates a new set of variables which are independent of one another. A matrix can be generated in such a way that both magnitude and direction of variables can be represented. As a result of PCA, not only the memory requirements of the system are reduced but also it makes processing faster. Most of the times PCA ensures maximized accuracy which makes image classification optimal [2].

Standardization of variables is performed as an initial phase in PCA as it ensures equal contribution from all continuous initial variables and transformation of all variables onto the same scale. Mathematically the standardization procedure can be performed by subtracting the variable value with the mean and dividing by the standard deviation. The standardization procedure can also be treated as a mandatory prerequisite as it prevents the dominance of variables with large range over the variables with smaller ranges and thus yields unbiased results.

Another important phase involved in PCA is to establish the relationship between the mean of variables and to analyze the correlation that exists between them. This can be done by generating the covariance matrix in which the values of the principal diagonal elements will represent variances of each variable. A positive sign indicates that these variables have the same rate of change while the negative sign indicates inverse correlation.

The aim of PCA is to generate principal components which is a linear mixture of uncorrelated variables whose first component consists of all the information in a compressed form. The first component is the most important component with respect to constituting information and hence is the most significant when compared to the remaining components.

Variance indicates the distribution of variables in a data set and its relationship. The calculation of the mean of the dataset is the first step in calculating variance. Variance is computed using the equation below.

$$\mathrm{Var}(X) = \Sigma(X_i - X)2/N = \Sigma x_i 2/N \tag{1}$$

where N is the total number of scores, X is the mean of the N scores, X_i is the each individual value, x_i is the ith deviation score in the set of scores, $\mathrm{Var}(X)$ is the variance of all the scores in the set.

Covariance indicates the relationship between different bands in the dataset and their direction. It is used to find out whether the variables are in the same direction or not.

2 Literature Survey

With each Eigen vector paired with an Eigen value and their number representing dimensionality of data and variance in the principal components. Constructing a feature vector with the Eigen vectors with the most prominent columns, will thus represent the reduced dimensionality. The feature vector emphasizes the reorientation of the data along the principal components which can be achieved by generating the product of transposes of the original data set and the feature vector [2].

Pearson's based correlation study has been conducted to analyze the COVID-19 death rate in various affected countries. These countries were classified using PCA based on the number of patients and number of deaths. Thus using PCA the extent of spread in these affected countries was estimated [3].

ENVI Software has been used to generate the False Colored Composite (FCC) of the dataset. Spectral decorrelation stretching has been performed to remove redundancy. PCA based hybrid compression transformation is performed. The dataset consists of 27 images of variable spatial correlation for the state Tamil Nadu, India. The factors that are significantly considered are rate of distortion and the property of preserving and retaining information [4].

Preprocessing techniques such as Radiometric and Geometric processing operations were carried out before PCA. The study area selected was North Iran and IRS LISS 3 images have been taken up for analysis. The data of each band was represented as a matrix which contained the Digital numbers of the pixels. Nearest neighbor resampling technique was used to ensure that the pixel brightness value was retained. Maximum Likelihood Classification algorithm was used for classifying the study area into 5 classes. Using Kappa Coefficient, classification accuracy was estimated. The first four principal components from LISS 3 images were used for analysis. Through the confusion matrix, an accuracy of 99.37% was estimated and through the Kappa coefficient analysis an accuracy of 99.09% was estimated. As a result image enhancement and dimensionality reduction was performed in the most optimized way [5].

High Resolution satellite imagery was used to study the damages caused by Tornado Outbreak during 1999. PCA and NDVI operations were performed for the purpose of image enhancement and spectral signature enhancement techniques. NDVI change analysis was implemented to observe the study area before and after the disaster. Since PCA does not require prior knowledge of the study area, it is one of the most robust methods for land cover classification. The resultant principal components were orthogonal to one another with ordered variance properties [6].

Maximum Likelihood classification is the most efficient in classifying Land cover types since it enables the estimation of land cover types in the study area, from which the training pixels can be selected. These training pixels can be further used to compute the mean vector and covariance matrix of each land cover class. This algorithm ensures that every pixel to be classified in the land cover classes. The study area selected is Aurangabad, India. Analyzing the confusion matrix yielded a

maximum classification accuracy of 98.198% while the Kappa coefficient produced a maximum value of 0.96 [7].

Only the first three components were used which constituted for about 99.3% of the image information with respect to the first study area. The near infrared band exhibits high reflectance values in the vegetation zones. Thus it can be concluded that these components are capable of detecting vegetation. Urban areas were identified by the second component. Land surface analysis was effectively performed by implementing PCA [8].

A combination of PCA and Artificial Neural Networks were used to classify multi-temporal satellite imagery for the purpose of land cover classification. The number of input nodes was decreased to reduce the time taken for the training phase. The back propagation algorithm was implemented on the four bands of the first two components of PCA [9].

For various applications such as urban and regional planning, natural resource management, natural disaster management, it would be required to identify an optimal classification algorithm for LULC mapping. This could be achieved by accuracy assessment of every classification algorithm using Kappa Coefficient, Receiver Operational Curve (RoC), Index-Based Validation and Root Mean Square error (RMSE). In this article, six machine-learning algorithms, some of which are random forest (RF), support vector machine (SVM), artificial neural network (ANN), fuzzy adaptive resonance theory-supervised predictive mapping were used for implementation.

Fuzzy (ARTMAP), spectral angle mapper (SAM) and Mahalanobis distance (MD) were examined. Experimental analysis shows that more advanced techniques such as ANN, SVM, yield much higher accuracy than traditional classifiers. Along with the type of algorithm implemented, factors such as spatial and temporal resolution of image dataset, hardware setup and software chosen for processing also contribute to increase in accuracy of classification. The Landsat 8 Operational Land Imager (OLI) image dataset covers some parts of Sahibganj District of Jharkhand and Malda and the Murshidabad districts of West Bengal, India, were chosen as the study area. The LULC classes to be identified included agriculture land, built up area, sand bar, fallow land, vegetation and wetlands [10].

Once the dataset was preprocessed and underwent a rectification process, the images were then trained by taking samples from the 6 classes. All the mentioned classification algorithms were rigorously implemented to identify LULC in the area of study. Normalized Differential Vegetation Index(NDVI), Normalized Differential Water Index(NDWI) and Normalized Differential Built up Index(NDBI) were computed in order to compare the performance analysis of each of the 6 algorithms [10].

Appropriate Softwares such as TerrSet Geospatial monitoring and modeling system was used for ANN and Fuzzy ARTMAP, Environment for Visualizing Images (ENVI 5.3) was used for Spectral Angle Mapper, SVM and Mahalanobis Distance while R programming language was used for Random Forest (RF). It was found that RF was the algorithm that suited best for the selected study area [10].

With the help of Geographical Information System(GIS), remote sensing images of Uttara Kannada District were analyzed for LULC changes. Base maps were created using topographic maps and Google Earth which served as ancillary data. The objective was to classify the study area into the following classes: Evergreen forest, Moist deciduous forest, Dry deciduous forest, Scrub forest/Grassland, Forest plantations, Built up, Water, Crop land, Horticultural plantations and Open land [11].

Once the preprocessing operations such as atmospheric and geometric corrections were performed, the image dataset was georeferenced and the region of interest was extracted. The FCC was generated using bands 2, 3, 4 which represented Green, Red and Infrared, respectively. NDVI was computed using Red and NIR bands. Supervised Classification was performed to generate the spectral signatures for each individual class. Training phase was effectively done by making use of topo sheets and Google Earth images. Gaussian Maximum likelihood Classification rule was used to assign an unknown pixel to one of the LULC classes. Along with LULC change analysis, the Forest fragmentation process was also quantified and analyzed [11].

LULC analysis is very much needed to increase the land productivity by restoring degraded lands and conducting procedures and operations for reclamation of wastelands. Performing classification of the study area by both digital image processing method and Visual Interpretation techniques using GIS software was done in this article. Accuracy assessment for both the methods was conducted and thus performance analysis was made. The study area chosen was Chamarajanagar district, Karnataka. Initially at the Level-1 classification phase, the study area was classified into 5 main classes such as Agriculture land, Built up land, Forest, wastelands and water bodies. At Level-2 classification, the study area was further subdivided into detailed sub-classes of the main class. Image Characteristics of various land use/land cover categories of the study area as seen in FCC was tabulated. The area of each class was estimated and tabulated at both Level-1 and Level-2 classification. It was found that change in land use was mainly due to the hydrological factors. A detailed analysis of LULC was performed so that land terrain could be represented. Land use/land cover in the form of maps and its statistical data helps in spatial planning, management, utilization of land for agriculture, forestry, pasture, economic production, agricultural planning, settlement surveys, environmental studies and operational planning based on agro-climatic zones etc. [12]

PCA technique has been implemented using GIS for the study area Bangladesh for the purpose of soil and water conservation in the process of natural resource management. The main objective was to prioritize watershed by the application of PCA and also to conserve soil and water which are very important natural resources. The images were classified into four main classes viz. cultivated land, water body, sand bar and settlement. LULC maps were generated to identify the drainage network using ARCGIS software. PCA here has been used as a parameter reduction technique and to generate the significant Principal Components. The area, perimeter and basin length of individual watersheds were determined. The paper concludes by initiating remedial measures to conserve soil and water [13].

PCA is used to monitor change detection using LANDSAT-5 TM data for the study area chosen as Alaska. Wavelet Transformation has been implemented by fusing important parameters of PC1 after which HKFCM classification algorithm has been applied. This classifier is based on unsupervised Fuzzy concept which consists of input layer, output layer, activation function without hidden layer [14].

3 Study Area and Data

The area of study for this work is chosen as Mysore district of Karnataka which consists of 7 talukas viz. Mysore, Tirumakudalu Narasipura, Nanjangud, Heggadadevanakote, Hunsur, Piriyapatna and Krishnarajanagara and also has two important rivers flowing in it which are Cauvery and Kabini. The geographical co-ordinates are 11°30′ to 12°50′ North Latitude 75°45′ to 77°45′ East Longitude. It covers a geographical area of 676,382 ha and is considered as prominent IT hub of Karnataka state [15].

Linear Imaging Self Scanning (LISS III) images of Mysore district, Karnataka were downloaded from Bhuvan Portal. LISS III imagery dataset consists of a set of multiband or multispectral images. They have a resolution of 24 m and a swath of 140 km with a revisit period of 24 days. The size of the tile is 15′ and occupies a memory of around 2 MB. These images consist of four spectral bands which are Green, Red, Near Infrared and Short Wave Infrared [16].

4 Methodology

Principal Component Analysis (PCA) was performed on LISS III images of the study area using open source Quantum Geographic Information System (QGIS) software. Initially all the four bands of the image were loaded into the QGIS software for computation. After which the virtual layer raster and band set definition were performed using the Semi-Automatic Classification Plugin (SCP). Using the SCP, Variance and Covariance matrix values are generated for the layer stacked Virtual Bandset image. Then the Eigen values and Eigen vectors are computed automatically for further analysis. In this paper, three Principal Components (PC) viz PC1, PC2 and PC3 are generated for the image and analyzed.

PCA is a method of extracting the significant features from the entire band set definition and removing the features of lesser importance. The objective behind PCA is to understand the correlation between different features of the image dataset. Covariance matrix is also called as dispersion matrix.

PCA was performed to generate the first three Principal Components along with their Eigen values, Eigen vectors. The Correlation Matrix and Covariance Matrix was generated for all the four bands of the image dataset. Figure 1 represents the

generated virtual bandset for which all the four bands were used. Bandset feature of SCP was used to generate the virtual bandset and also the band sets were defined.

Figures 2 and 3 shows the first and second Principal Components respectively while Fig. 4 shows the matrices generated. Tables 1 and 2. shows the covariance and correlation matrix generated.

Using Symbology feature in QGIS, allows the user to perform land delineation and generate LULC classes. Three different classification schemes are available viz.

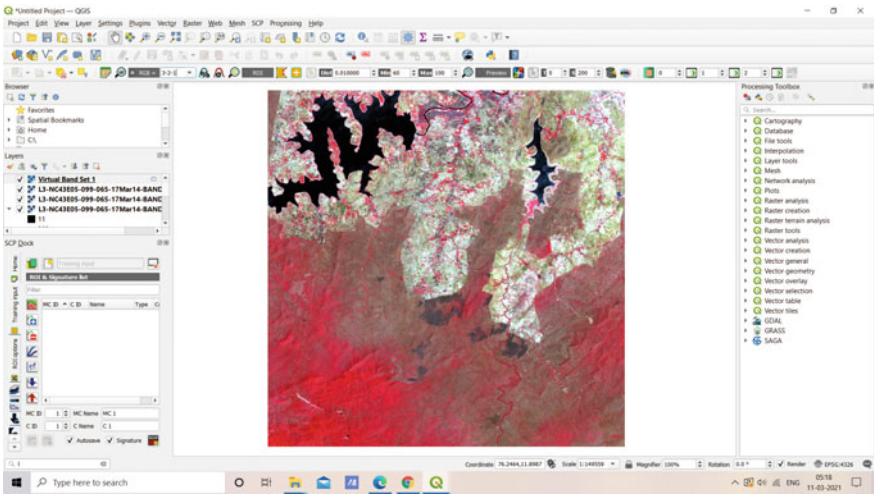


Fig. 1 Band set definition image in QGIS

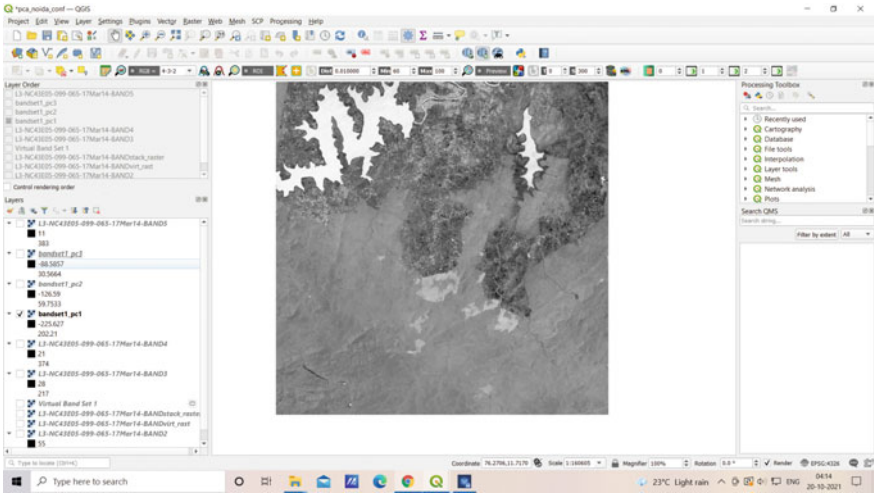


Fig. 2 First principal component

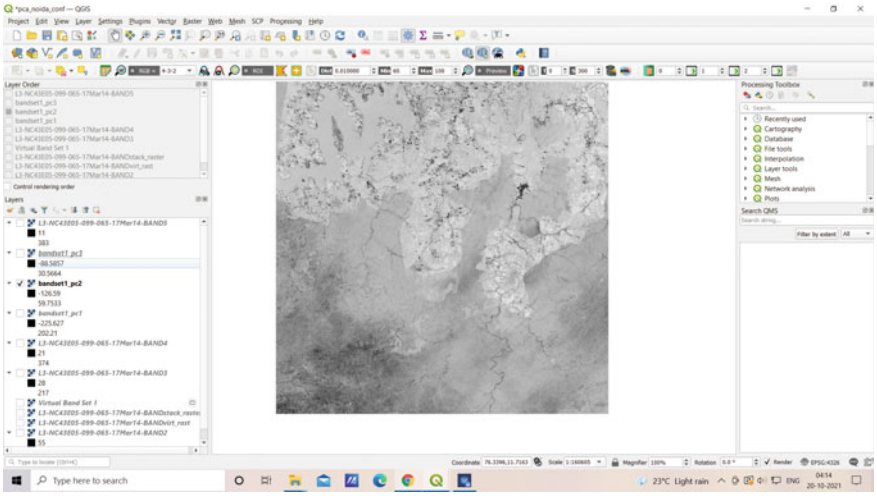


Fig. 3 Second principal component

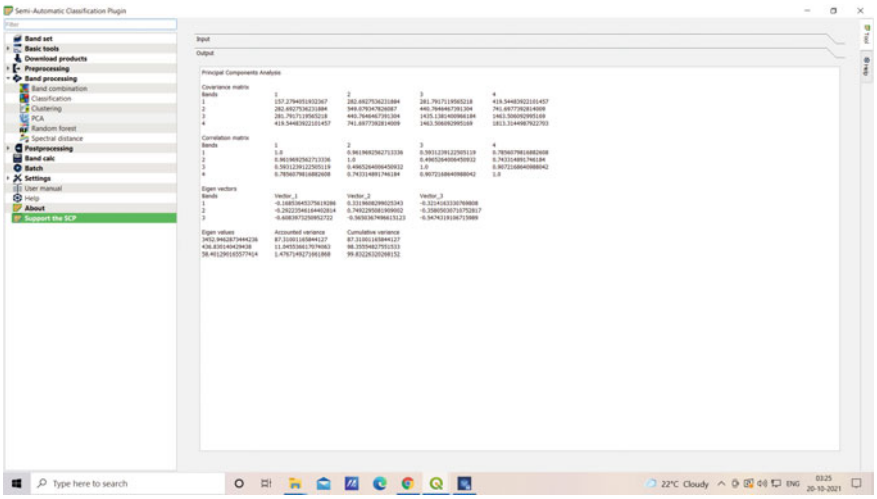


Fig. 4 Covariance and correlation matrices

Table 1 Covariance matrix

Bands	1	2	3	4
1	157.27940	282.69275	281.79171	419.54483
2	282.69275	549.07934	440.76464	741.69773
3	281.79171	440.76464	1435.1381	1463.5060
4	419.54483	741.69773	1463.5060	1813.3144

Table 2 Correlation matrix

Bands	1	2	3	4
1	1.0	0.96196	0.59312	0.78560
2	0.96196	1.0	0.496526	0.74331
3	0.59312	0.49652	1.0	0.90721
4	0.78560	0.74331	0.96721	1.0

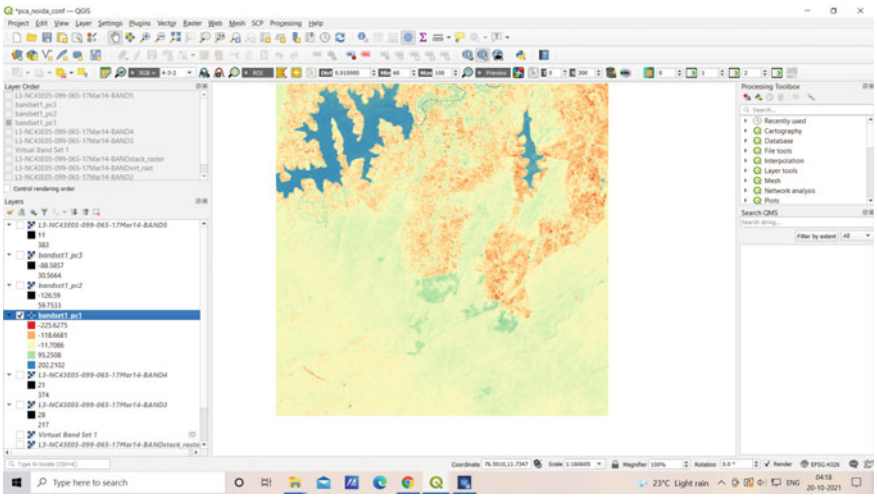


Fig. 5 Symbology of first principal component

Continuous, Equal Interval and Quantile. The number of classes can be chosen such that classification is optimal. Also a diverse range of colors are available on the color ramp for appropriate selection of colors for specific applications. Figures 5 and 6 represents the Symbology feature of SCP of the first and third Principal Components.

Using the PCA feature in SCP allows the user to select the bandset definition, and number of Principal components to be generated. SCP plugin is ideal for generating LULC maps along with performing change detection studies. Using SCP a wide range of datasets can be downloaded and loaded into the QGIS platform.

5 Results and Analysis

With the SCP Plugin, by making use of the PCA feature, the Variance–Covariance matrix can be automatically generated. For the dataset considered, Table 1 shows the above mentioned matrix. The principal diagonal elements represents the Variance values, for the 4 bands of the input dataset viz. Green, Red, Near Infrared and Short

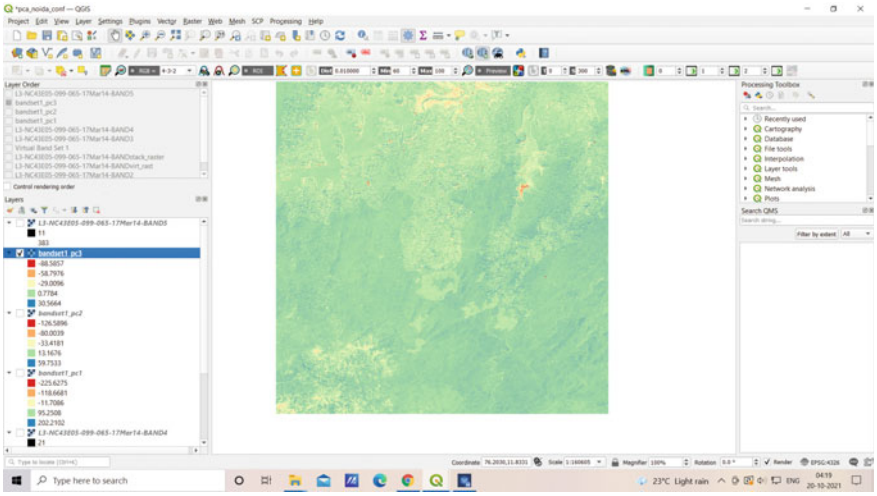


Fig. 6 Symbology of third principal component

Wave Infrared. The non-diagonal elements indicate the Covariance values for all the four bands of the considered dataset. All the Variance and Covariance values that were generated were found to be positive which means that they are correlated linearly. However, a negative value indicates an inverse non-linear relationship. In this process of PCA computation, the relationship with the associated bands are depicted. Table 2 shows the Correlation matrix whose principal diagonal elements are 1.0.

The first Principal Component (PC1) shown in Fig. 2 represents the combination of variables in a linear relationship with a maximum value of variance associated with it. The second Principal Component (PC2) shown in Fig. 3 also is generated with the same concept but with the condition that the correlation between PC1 and PC2 is 0. Thus the PC2 consists of all the remaining variance which was not a part of PC1. Table 2 represents the Correlation Matrix and its distribution associated with all the bands of the input dataset. It is found that the variables are highly correlated.

6 Conclusion

The diagonal elements in the covariance matrix represent the variance values between each bands. The off diagonal elements in the Covariance Matrix represent the Covariance values. If the Covariance values are positive then there exists a linear combination of values. If the values are negative then the values change in opposite directions. If the Covariance is 0 then there is no predictable relationship between the values. Variance and Covariance values represent the distribution of a set of points about their mean value. Variance is the measure of deviation from the mean value for points

in 1 dimension. Covariance is measured between 2 dimensions to see if there is a relationship between the 2 dimensions.

Thus this paper concludes that there exists a linear relationship between the bands and its variables of the input dataset. It is also to be noted that PC1 and PC2 are uncorrelated and thus the subsequent Principal Components to be generated would also not be correlated with one another.

PCA eliminates data redundancy and thus reduces memory requirement for the system. It simplifies the Image processing and analysis procedures by providing a smaller representation of the dataset. It reduces the noise in images and hence increases the accuracy of classification. Along with data compression, dimensionality reduction and simplicity in data analysis, PCA also supports in representing the data over the feature vector, i.e., where there is maximum variation in data.

References

1. Mishra S, Sarkar U, Taraphder S, Datta S (2017) Principal component analysis, Int J Livestock Res
2. https://builtin.com/data-science/step-step-explanation-principal-component-analysis
3. Mahmoudi MR, Heydari MH, Qasem SN, Mosavi A, Band SS. Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries, Alexandria Eng J
4. Thayammal S, Priyadarsini S, Selvathi D (2020) Edge preserved multispectral image compression using PCA and hybrid transform. Multimedia Tools Appl
5. Dolati MK, Bonyad AE (2016) Use of principal component analysis in accuracy of classification maps (case study: North of Iran). Res J Forest
6. Yuan M, Dickens M, Magsig MA (2015) Analysis of Tornado damage tracks from the 3 May Tornado outbreak using multispectral imagery. ResearchGate
7. Chatse M, Gore V, Kalyane R, Dr. Kale KV (2016) Performance evaluation of maximum-likelihood algorithm after applying PCA on LISS 3 image using python platform. Int J Sci Eng Res
8. Estornell J, Jesus M, Mart Gavila M, Teresa Sebastia A, Mengual J (2013) Principal component analysis applied to remote sensing. Model Sci Educ Learn
9. Chae HS, Kim SJ, Ryu JA (1997) A Classification of Multitemporal Landsat TM data using principal component analysis & artificial neural network
10. Talukdar S, Singha P, Mahato S, Pal S, Liou YA, Rahman A (2020) Land-use-land-cover classification by machine learning classifiers for satellite observations—A review. Remote Sens (2020)
11. Kuchay SA, Ramachandra TV (2016) Land use land cover change analysis of Uttara Kannada. Imperial J Interdiscip Res
12. Basavarajappa HT, Dinakar S, Manjunatha MC (2014) Analysis on land use/land cover classification around Mysuru and Chamarajanagara District, Karnataka, India, Using IRS-1D Pan+LISS-III satellite data. Int J Civil Eng Technol
13. Arefin R, Md. Mohir MI, Alam J (2020) Watershed prioritization for soil and water conservation aspect using GIS and remote sensing: PCA-based approach at northern elevated tract Bangladesh. Appl Water Sci (2020)
14. Ruhil N, Singh M, Mitra D, Singh A, Singh KK (2019) Detection of changes from satellite images using fused difference images and hybrid Kohonen fuzzy C-means sigma. Procedia Comput Sci

15. District Industrial Profile, Mysore
16. https://data.gov.in/keywords/resourcesat

A Review on Different Preprocessing and Feature Extraction Technique for SSVEP BCI Inference System

Mukesh Kumar Ojha and Dhiraj Gupta

Abstract The steady-state visual evoked potential (SSVEP) is a periodic signal contaminated with recorded electroencephalography (EEG) signal. Accurate detection of SSVEP signals from noise contaminated EEG signal is the key challenge to improve the performance of an SSVEP-based BCI system. Therefore, the use of a signal processing algorithm plays a significant role to detect the SSVEP signal with great accuracy. This paper describes the recent development in the use of various existing detection algorithms for the SSVEP BCI system. The signal processing technique related to preprocessing and feature extractions is discussed in this paper. This study report that technique that can be applied for non-stationary and nonlinear signals analysis are more employed as compared to traditional Fourier transform to improve the performance for SSVEP BCI system. Spatial filtering techniques are useful for channel selection and to eliminate the nuisance signal from multi-channel EEG signal.

Keywords Steady-state visual evoked potential (SSVEP) · Canonical correlation analysis (CCA) · Brain computer interface (BCI) · Electroencephalography (EEG) · Empirical mode decomposition (EMD)

1 Introduction

The brain-computer interface (BCI) system enables the subject to communicate with the computer directly through brain signals [1, 2]. The steady-state visual evoked potential (SSVEP) signal is an evoked signal contaminated with the recorded electroencephalography (EEG) signal used to develop the BCI system [3]. In recent times, the SSVEP-based BCI systems have become renowned among other BCI systems due to the relative high-information transmission rate (ITR), minimal training time, and high signal-to-noise ratio (SNR) associated with SSVEP [4].

M. K. Ojha (✉) · D. Gupta
Greater Noida Institute of Technology, Greater Noida 201310, India
e-mail: mukeshbit12@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_39

The SSVEP is a periodic wave that induces into a non-invasive EEG signal when the subject observes the visual-stimulus flicker at a particular frequency [5, 6]. The SSVEP signal appears in the recorded EEG signal when its frequency matches the stimulation frequency and harmonics with the spontaneous EEG signal and artifacts associated with the brain recording system. In recent years, many authors suggested that the presence of artifacts and other background EEG signals might deteriorate the activities and functions of the SSVEP BCI system [7, 8]. Therefore, the primary challenge is to detect and identify the SSVEP signal's presence from the recorded EEG signal for the development of high-performance SSVEP BCI systems. The solution of this question is quite straight forward. We need efficient signal processing algorithm for preprocessing and feature extraction of recorded EEG signal. This technique must be able to extract the SSVEP frequency component with great accuracy from recorded EEG Signal.

Consequently, many methods have been employed in the preprocessing part of the SSVEP-based BCI systems. Among them, the most common method is frequency filtering and more particularly, it is band-pass filtering [9]. Beside the frequency filtering approaches, the spatial filtering approaches [10, 11] have also been used at a larger extend. More specifically, the minimum energy filtering [7] methods have been employed to remove the components unrelated to SSVEP.

The central part of SSVEP BCI is the feature extraction by which SSVEP frequency component is extracted from the recorded EEG signal. In this direction, the spectral analysis is the most widely used approach to extract the frequency component of SSVEP signals [3, 4]. In addition to this, the canonical correlation analysis (CCA) is another efficient method to extract the frequency components of SSVEP signals [12]. In recent year, many researchers proposed several extension of CCA algorithm to optimize the reference signal in order to enhance the accuracy of SSVEP-based BCI system [13–16]. To this end, this present study explores the various existing preprocessing and feature extraction method and their limitation for SSVEP BCI system.

1.1 SSVEP BCI System

The block diagram of the SSVEP-based BCI system is shown in Fig. 1. It consists of multiple stages. They are signal acquisition module to acquire the EEG signal non-invasively, signal processing algorithm, and an output system to control the action of the device. The signal processing module consists of a preprocessing module, feature extraction module, and classification module. The aim of the signal preprocessing module is to eliminate the noise and artifacts from the acquired EEG signal. The feature extraction and classification module use the characteristics of the SSVEP signal to identify the subject's intent to control the device. Thus, the signal processing module provides the link between the subject and external device. The selection of

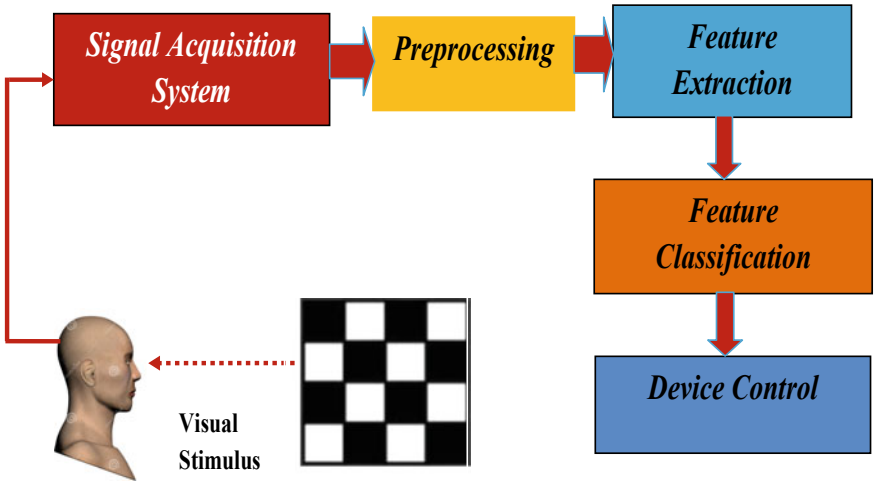


Fig. 1 SSVEP BCI inference system

appropriate signal processing modules plays a crucial role to enhance the performance of the SSVEP BCI system. Finally, the detected SSVEP frequency component is given as a command frequency to communicate and control the external device.

2 Signal Preprocessing

The aim of signal preprocessing stage is to minimize the nuisance signal as much as possible to get the clean EEG signal and to improve the recognition accuracy of the SSVEP signal contaminated with recorded EEG signal. In general, there are two filtering approaches are used for the SSVEP BCI system to minimize the nuisance and other artifacts. They are the frequency filtering approach and spatial filtering approach. The spatial filtering approach is also used for feature extraction.

2.1 Frequency Filtering Approach

Band-pass filter (BPF) and notched filter are two commonly used frequency filtering approaches to eliminate the nuisance signal in the SSVEP BCI system. BPF allows the signal to be processed within a certain frequency range corresponding to stimuli frequency and its harmonics [9]. The stimulus frequencies are classified as low frequency (2–12 Hz), medium frequency (12–30 Hz), and high frequency (above

30 Hz). Most SSVEP BCI preferred low frequency because it exhibits high amplitude SSVEP signal at low frequency [9]. However, the low stimulus frequency signal induces visual fatigue. In addition to this, the low and medium stimulus frequency signal also induces epileptic seizures [17]. For safety and comfort, the high frequency of the stimulus is preferred for SSVEP-based BCI system [18]. The function of the notched filter is to remove the power line interference which is 50 Hz in most countries [9]. Almost every reviewed paper uses both filters in the preprocessing stage to minimize the nuisance signal which is implemented either using hardware or software.

2.2 Spatial Filtering Approach

Spatial filtering pre-processes the signals recorded from multiple channels using the spatial domain. The main function of spatial filtering to enhance the signals related to SSVEP or to minimize the nuisance signal to improve the performance of the SSVEP BCI system. They use the linear combination of recorded EEG signals from different channels to perform the task. Minimum energy combination (MEC) [7] and maximum contrast combination (MCC) [18] are the two most frequent methods used for spatial filtering. The function of MEC is to eliminate the signal unrelated to SSVEP while the function of MCC is to improve the SNR. Other approaches such as common average references (CAR) [19], principle component analysis (PCA) [20], and analytical common spatial filtering process (ACSP) [21] have also been used in past as spatial filters to minimize the nuisance of signal and to enhance SNR. The advantage of spatial filtering is that it can be used for signal preprocessing as well as for feature extraction. In past, a different approach of spatial filtering method has been used as future extraction and the result shows that the MEC approach achieved better accuracy as compared to other methods.

3 Feature Extraction Method

As we know the signal processing unit of the SSVEP BCI system consists of signal preprocessing, feature extraction, and classifications subunit. Feature extraction is one of the key parts of the signal processing unit that play an important role to optimize the overall performance of the SSVEP BCI system. The various method used for feature extraction of SSVEP BCI systems is summarized into Table 1. This includes Fourier transform (FT), wavelet transform, Hilbert Hung transform (HHT) method, and canonical correlation transform.

Table 1 Different preprocessing technique for SSVEP BCI

Preprocessing technique	Description	References
Band-pass filter	BPF is designed to limit the signal into specific frequency range corresponding to stimulus frequency and its harmonics	[9, 17, 18]
Notched filter	Notched filter is designed to eliminate the noise due to power line interference	[9]
Spatial filtering	Spatial filter is mainly designed to combine the signals from different channel to improve the SSVEP response or to reduce the noise and artifacts	
Minimum energy combination	The aim of MEC is to minimize the noise as much as possible by combining the signals from different channel	[7]
Maximum contrast correlation	The objective is to improve the SNR. This is similar to MEC only difference is filter coefficients weights are calculated differently	[18]
Common average reference	CAR method subtracts the average value of all electrodes from the channel of interest to make EEG signal reference free	[19]
Principal component analysis	Principal component analysis decomposes the signal into SSVEP component and brain activity. Its main objective is to reduce the dimension	[20]
ACSP	A common spatial pattern method is designed to give both amplitude and phase information based on CSP	[11]

3.1 Fourier-Based Transform

Fourier transform (FFT) is one of the most widely used methods to detect the SSVEP frequency from recorded EEG signals. This approach is mainly used to find the power spectrum density (PSD) of recorded EEG signals to detect the SSVEP frequency [22, 23]. The main advantages of Fourier-based transform approach are its simplicity and less computational time. Table 2 indicates the list of reviewed paper that employed Fourier-based transform approach to detect the SSVEP frequency components. Most of the paper explores the Fourier transform method to compute the power of recorded EEG signal at stimulation frequency and its harmonics to detect the SSVEP frequency component [4]. In another study, the author computed the average power of recorded EEG signal centered around stimulus frequency, and the frequency with maximum power is considered as SSVEP frequency [24]. In multi-channel signal analysis, the selection of optimum electrodes from different subjects was also performed using the FFT method.

Table 2 Different feature extraction technique for SSVEP BCI

Feature extraction technique	Description	References
Fourier-based transform	An effective method to compute the power spectrum or to compute SSVEP phase	
Fast Fourier transform	FFT method has fast computational speed but require long segment length to compute the power spectrum. Suitable for stationary and linear signal	[5]
DFT	This method requires more computation time as compared to FFT	[29]
Wavelet transform	Wavelet transform has good time and frequency resolution due to adjustable window size. This method is suitable for non-stationary signal	
Filter bank decomposition	This approach decomposes the signal into several sub-bands. The sub-band corresponding to stimulus frequency is chosen for feature extraction	[23, 25, 26]
Wavelet packet decomposition	Decompose the signal into complete wavelet tree	[30]
Empirical mode decomposition + PSA	EMD technique is suitable for non-stationary and nonlinear signal analysis. It decompose the signal into several IMFs. FFT method computes the power spectrum of IMFs in which SSVEP component prominent	[4, 23]
EMD + RGZC	This approach finds the instantaneous frequency using refined generalize zero crossing method after decomposing the signal into several IMFs using EMD	[27]
CCA	An efficient technique to find the SSVEP component. This method finds the correlation multi-channel EEG signal and sine–cosine reference signal. The maximum correlation coefficient is considered as SSVEP frequency component	[12]
M-CCA	Uses optimal reference signal instead of sine–cosine to improve the performance of standard CCA	[13]
K-CCA	Kernel CCA approach project the dataset into high dimensional space to enhance the performance of traditional CCA of which data sets are linearly related	[14]
p-CCA	Phase information is used into reference signal	[15]

(continued)

Table 2 (continued)

Feature extraction technique	Description	References
MEMD + CCA	MEMD technique minimizes the mode-mixing problem of EMD by adding Gaussian noise with original signal	[16]

3.2 Wavelet Transforms (WT)

WT is another efficient method to compute the frequency component of the SSVEP signal. They have variable window sizes with a good time and frequency resolution [25]. Thus, they are suitable to process the non-stationary and nonlinear EEG signal. They use the filter bank approach to decompose the EEG signal into several sub-bands [23]. The sub-band which contains the stimulus frequency is chosen for further processing. The wavelet coefficients are chosen to find the power at relevant frequency.

The major problem associated with wavelet analysis is the selection of appropriate mother wavelet functions to improve the performance of the SSVEP BCI system [26]. The selection of appropriate wavelets to attain the good performance of SSVEP BCI is still worth of the research. In one study [26], the author experimented with various wavelet functions and reported that a complex Morlet wavelet was best. Although, wavelet analysis is more suitable for the analysis of non-stationary signals as compared to the fast Fourier transform (FFT) method.

Frequency resolution is the key issue with the FFT method. The frequency resolution is defined as the ratio of sampling frequency to the data length. The window length of the recorded EEG signal should be large to improve the frequency resolution when the sampling frequency is fixed. This may causes classification error during the change of stimulus.

FFT method is mainly designed for signals which are linear and stationary. But the recorded EEG signals are non-stationary and nonlinear. It is therefore presumed that each segment of the recorded EEG signal is linear and stationary. This presumption may affect the overall performance of the SSVEP BCI system.

3.3 Hilbert Hung Transform (HHT)

Hilbert Hung transform (HHT) is a self-data-driven technique suitable for nonlinear and non-stationary signal analysis. HHT technique decomposes the recorded EEG signal into several oscillating functions known as intrinsic mode function (IMFs) [4, 23, 27]. Each IMF represents the local characteristics recorded EEG signal. In a recent study, several authors have used the EMD technique to decompose the original-SSVEP signal into many IMFs. SSVEP related IMFs are selected by computing

the power spectrum density of each IMF and then IMF whose frequency component closest to stimulation frequency is identified as target frequency [4]. In another study [28], the IMFs which have maximum correlation coefficients corresponding to stimulus frequency are used to identify the target frequency. In recent years, multivariate empirical mode decomposition (MEMD) [28] is developed to overcome the mode-mixing problem of EMD that occurs due to signal intermittences. MEMD technique added the white noise with the original signal and then decomposes the resultant signal into several IMFs. HHT is more stable as compared to FFT, which means recognition accuracy does not change significantly with a change in the data length. In addition, HHT does not require a pre-defined basis function like wavelet to decompose the signal into several oscillating functions.

3.4 Canonical Correlation Analysis (CCA)

CCA is another efficient method to detect the SSVEP components obtained from multi-channel EEG signals. CCA is a multivariate statistical analysis that computes the correlation between two sets of data [12]. This approach linearly transforms the observed dataset into a new dataset and then finds the maximum correlation coefficients between transformed two data sets [12–14].

In the case of detection of SSVEP components, the CCA method finds the correlation between the SSVEP signal and the reference signals which includes sine and cosine signals of the same frequency as the stimulus frequency. The maximum correlation coefficient is considered as SSVEP frequency components. The author [22] compares the CCA method with PSDA and reported that CCA performs superior as compared to PSDA for multi-channel SSVEP signal analysis. In recent development, different CCA approach has been used to enhance the performance of SSVEP BCI system. A multiset CCA approach has been proposed by the author [13] to find the optimal reference signal rather than sine–cosine as a reference signal to improve the detection accuracy of the SSVEP BCI system. Another paper [14] proposed the Kernel CCA approach to achieve better detection accuracy as compared to the standard CCA approach. In addition, a phase constrained CCA (p-CCA) was proposed by the author [15] to improve the accuracy using the phase of SSVEP signal as a reference signal.

The study of several papers using the CCA approach reported that the first maximum correlation coefficient is only utilized to find the detection accuracy of the SSVEP BCI system. The use of other coefficients is still worth researching.

4 Conclusion

This study mainly outlines the several preprocessing and feature extraction method used for SSVEP BCI system. Band-pass filter and notched filters are mainly used to

eliminate the noise. Spatial filtering method is mainly used for multi-channel signal processing to improve the SNR and to minimize the effect on inter-subject variation.

FFT and CCA are two most popular methods employed for feature extraction. FFT approach finds the power spectrum density at stimulus frequency and its harmonics. The advantage of FFT approach is its simple approach and require less computation time. However, FFT approach is mainly applicable for linear and stationary signal. In recent development, wavelet transform and Hilbert Hung transform are employed for nonlinear and non-stationary signal.

Another approach CCA computes the correlation between multi-channel SSVEP signal and reference signal as sine–cosine. The maximum correlation between SSVEP signal and reference signal at stimulus frequency is considered as SSVEP frequency. In recent development, several authors optimize the reference signal to detect SSVEP frequency component with great accuracy.

References

1. Walpow JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clin Neurophys 113:767–791
2. Chaudhary U, Birbaumer N, Ramos-Murguialday A (2016) Brain-computer interfaces for communication and rehabilitation. Nat Rev Neurol 12:513–525
3. Xiaogang C, Yijun W, Shangen Z, Shangkai G, Yong H, Xiaorong G (2017) A novel stimulation method for multi-class SSVEP-BCI using inter modulation frequencies. J Neural Eng 14
4. Ojha MK, Mukul MK (2021) A new novel approach based on EMD to improve the performance of SSVEP based BCI system. Wirel Pers Commun
5. Hwang H-J, Lim J-H, Jung Y-J, Choi H, Lee SW, Im C-H (2012) Development of an SSVEP-based BCI spelling system adopting a QWERTY-style LED keyboard, J Neurosci Method. 208:59–65
6. Gao S, Wang Y, Gao X, Hong B (2014) Visual and auditory brain computer interfaces. IEEE Trans Biomedical Eng 61:1436–1447
7. Friman O, Volosyak I, Gräser A (2007) Gräser A Multiple channel detection of steady-state visual evoked potentials for brain–computer interfaces. IEEE Trans Biomed Engg. 54:742–750
8. Jiang X, Bian GB, Tian Z (2019) Removal of artifacts from EEG signals: a review. Sensors 19:97
9. Müller SM, Diez PF, Bastos-Filho TF, Sarcinelli-Filho M, Laciar VME (2011) SSVEP-BCI implementation for 37–40 Hz frequency range. In: Proceedings of annual international conference of the IEEE engineering in medicine and biological society, pp 6352–6355
10. Garcia-Molina G, Zhu DH (2011) Optimal spatial filtering for the steady state visual evoked potential: BCI application. In: Proceedings of 5th international IEEE/EMBS conference on neural engineering, pp 156–160
11. Falzon O, Camilleri K, Muscat J (2012) Complex-valued spatial filters for SSVEP-based BCIs with phase coding, IEEE Trans Biomed Eng 59:2486–2495
12. Lin Z, Zhang C, Wu W, Gao X (2007) Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Trans Biomed Eng 54:1172–1176
13. Yu Z, Guoxu Z, Jing J, Xingyu W, Andrzej C (2014) Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis. Int J Neural Syst 24
14. Zhang ZM, Deng ZD (2012) A kernel canonical correlation analysis based idle-state detection method for SSVEP-based brain-computer interfaces. In: Proceeding of 2nd international conference on material and manufacturing technology, pp 341–342, 634–640

15. Pan J, Gao X, Duan F, Yan Z, Gao S (2011) Enhancing the classification accuracy of steady-state visual evoked potential-based brain-computer interfaces using phase constrained canonical correlation analysis. J Neural Eng 8:036027
16. Rai A, Sharma D, Rai S, Singh A, Singh KK, IoT-aided robotics development and applications with AI. Emergence of cyber physical system and IoT in smart automation and robotics. Adv Sci Technol Innov
17. Fisher RS, Harding G, Erba G, Barkley GL, Wilkins A (2005) Photic-and pattern-induced seizures: a review for the epilepsy foundation of America working group. Epilepsia 46:1426–1441
18. Bir P, Karatangi SV, Rai A (2020) Design and implementation of an elastic processor with hyperthreading technology and virtualization for elastic server models. J Supercomput 76:7394–7415
19. Muller SMT, Bastos TF, Sarcinelli Filho M (2013) Proposal of a SSVEP-BCI to command a robotic wheelchair. J Control Autom Electr Syst24:97–105
20. Pouryazdian S, Erfanian A (2009) Detection of steady-state visual evoked potentials for brain-computer interfaces using PCA and high-order statistics. In: Proceedings of world congress on medical physics and biomedical engineering, vol 25, pp 480-483
21. Picton T (1990) Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine. J Clin Neurophysiol 7:450–451
22. Hakvoort G, Reuderink B, Obbink M (2011) Comparison of PSDA and CCA detection methods in a SSVEP-based BCI system
23. Ojha MK, Mukul MK (2021) A novel approach based on EMD to improve the performance of SSVEP based BCI system. Wirel Pers Commun. 2455–2467. https://doi.org/10.1007/s11277-021-08135-6
24. Muller-Putz GR, Pfurtscheller G (2008) Control of an electrical prosthesis with an SSVEP-based BCI. IEEE Trans Biomed Eng 55:361–364
25. Samar VJ, Bopardikar A, Rao R, Swartz K (1999) Wavelet analysis of neuro-electric waveforms: a conceptual tutorial. Brain Lang 66:7–60
26. Zhang Z, Li X, Deng Z (2010) A CWT-based SSVEP classification method for brain-computer interface system. In: Proceedings of international conference on intelligent control and information processing, pp 43–48
27. Wu CH, Chang HC, Lee PL, Li KS, Sie JJ, Sun CW, Yang CY, Li PH, Deng HT, Shyu KK (2011) Frequency recognition in an SSVEP-based brain computer interface using empirical mode decomposition and refined generalized zero-crossing. J Neurosci Methods 196(2011):170–181
28. Chen Y-F, Atal K, Xie SQ, Liu Q. A new multivariate empirical mode decomposition method for improving the performance of SSVEP-based brain computer interface, J Neural Eng https://doi.org/10.1088/1741-2552/aa6a23,2017
29. Ortner R, Allison BZ, Korisek G, Gaggl H, Pfurtscheller G (2011) An SSVEP BCI to control a hand orthosis for persons with tetraplegia. IEEE Trans Neural Syst Rehabil Eng 19:1–5
30. Bian Y, Li HW, Zhao L, Yang GH, Geng LQ (2011) Research on steady state visual evoked potentials based on wavelet packet technology for brain-computer interface. Proc Eng 15:2629–2633

Digital IC Tester Using Advanced Microcontroller Board Interfaced with GUI MATLAB

Shylaja V. Karatangi, Amrita Rai, Puneet Kumar Mishra, Krishanu Kundu, and Reshu Agarwal

Abstract In today's day, all devices to be compact and easy access to verify the functions. So with the use of software and hardware IC tester will be made easy to use, and testing of all TTL 74xx series. The digital integrated circuit tester described in this paper uses an advanced microprocessor to test 14-pin digital integrated circuits (ICs) from the 74xx series. First, the system greets the user and prompts them to input their IC number. It receives input from the IC's input pins and tests it using a GUI on a PC. It then verifies the IC (in the 14-pin socket) according to the IC number and sends the output to the PC, displaying the result on the GUI display as "GOOD" or "BAD" depending on the gate number. If the IC number is given as input incorrectly or more than two digits entered, an error signal "ERROR" is displayed.

Keywords Arduino UNO board · Digital IC · Tester · GUI · IC under test

1 Introduction

Engineers test ICs on two levels to see if they are in working order: wafer level testing and packaging level testing. Shorting conditions, leakage conditions, supply

S. V. Karatangi (✉) · A. Rai · P. K. Mishra · K. Kundu
Department of Electronics and Communication Engineering, G. L Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: shilajachalageri@gmail.com

A. Rai
e-mail: amritaskrai@gmail.com

P. K. Mishra
e-mail: puneetmishra1988@gmail.com

K. Kundu
e-mail: krishanukundu08@gmail.com

R. Agarwal
AIIT, Amity University, Noida, India
e-mail: ragarwal1@amity.edu

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_40

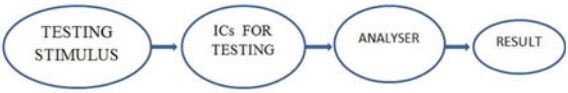


Fig. 1 Working diagram of system

conditions, and input and output conditions are the four general tests used to determine the IC's status. The acronym IC stands for "integrated circuit" and refers to any semiconductor-based chip that contains an integrated set of digital circuitry. A conventional logic IC is a single, tiny integrated package that includes all necessary components and functions for a logic circuit.

Applying a set of test stimuli to the Arduino UNO board's input pins and analyzing the system's reaction with an analyzer is expected to test various integrated circuits [1–3]. The IC under test is fault-free if it delivers proper output responses for all input stimuli [3, 4]. Those IC that fail to reach the desired reaction during testing is considered faulty or defective [3, 5]. This project adopts a user-friendly technique to test 74xx series digital ICs using a MATLAB graphical user interface and an Arduino UNO board [3]. Figure 1 shows the block diagram for sampling digital ICs.

MATLAB serves as the test stimulus generator for the IC in the testing system [3, 6]. The GUI connects to the Arduino UNO board and provides a user-friendly and interactive environment for the entire testing process. The response analyzer is a MATLAB source program that displays test results on the GUI's screen [3, 7, 8].

As previously stated, MATLAB is used to apply a stimulus to the IC under test and to collect the IC's reaction to the stimulus fast. It then compares the IC's answer to the desired/resultant response to determine whether the device is defective [10, 11].

The truth table shows the correct output for a digital 74xx series IC. The MATLAB source code, which acts as a response analyzer, analyzes every conceivable production to the truth table of a specific IC [6, 9].

2 Working of the Proposed System

2.1 Auto-Search Method

The number of pins on the IC to be verified is entered initially in this process. The device then begins manifesting all potential input signals to the IC and returning its answer for each one. If a response matches the output of an IC in its database, it considers that IC to be excellent.

2.2 Manual-Checking Method

The IC number is entered first in this approach (e.g., 7432). The essential details of that IC are specific away in the following section. The user has known the option of using a truth table at the start of the verification procedure. This alternative must be chosen to display the truth tables. Following that, the MCU starts the signal-processing duty.

The MCU supplies 5 V to pin 14 and 0 V to pin 7 of the IC 74xx series. Because the IC has many gates, each one is verified individually. The MCU feeds each gate the correct combination of inputs based on the truth table and receives outputs from the IC as input.

The MCU then draws its conclusion on that gate by comparing observed results to expected results based on IC specifications. Finally, the number of good and bad gates and the overall condition of the IC as indicated. All testing of logic gate family IC is based on truth table shown in Table 1. Each IC series have four such gates inside the integrated chips, proposed tester check each gate and display the actual condition of each gate as well as whole IC.

The suggested system's flow chart depicts in Fig. 2 explain the process of the GUI display screen initializing and showing the welcome screen after receiving supply power. After that, when the IC is connected to the Arduino via socket, the controller

Table 1 Truth table of all digital logic gates for IC tester

TRUTH TABLE FOR 74xx SERIES ICs

TRUTH TABLE OF AND GATE

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

TRUTH TABLE OF OR GATE

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

TRUTH TABLE OF NOT GATE

A	Y
0	1
1	0

TRUTH TABLE OF XOR GATE

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

TRUTH TABLE OF NAND GATE

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

TRUTH TABLE OF NOR GATE

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

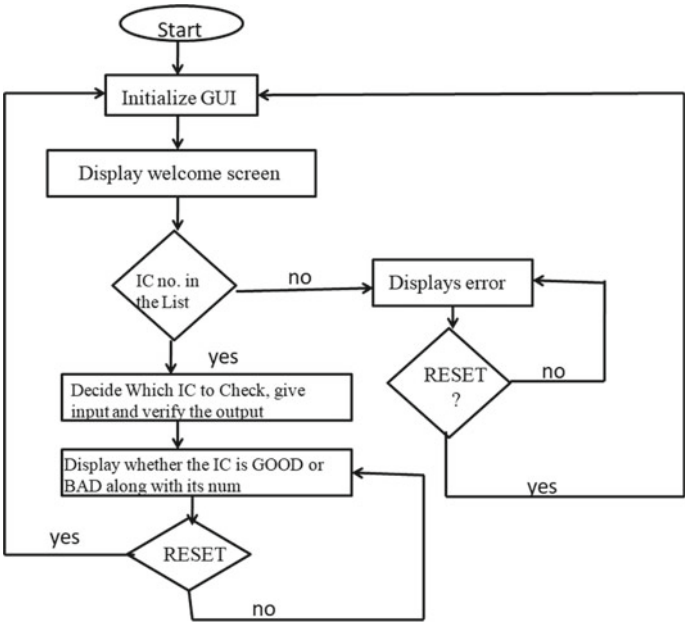


Fig.2 Flow chart of proposed system and their operations

looks up the IC number in the list and provides the appropriate input to the IC, compares and verifies the result, and ultimately displays the effect on the GUI to indicate whether the condition is good or bad.

3 Working with MATLAB and Arduino Board

The MATLAB R2014b version was used to create the GUI application program. Open the project's source code files after appropriately setting up the package path. Both source code files should be kept in the same folder. Replace the COM port with the port number on your PC where the Arduino Uno board is installed (in the line $a =$ arduino ('COM4')). Run the file and click the "Connect" button to connect MATLAB and the Arduino Uno board. Select the appropriate IC from the drop-down menu selections after successful contact has been established.

Two inputs are taken in the truth table discussed before, and the desired output is received after the procedure. When the result matches the production of a specific 74xx IC, the ICs can be categorized based on their behavior. If none of the outputs match the input, the GUI displays BAD on the screen.

The source code program is based on the switch case principle, which means that when a specific IC number matches a case in the source code, the related condition

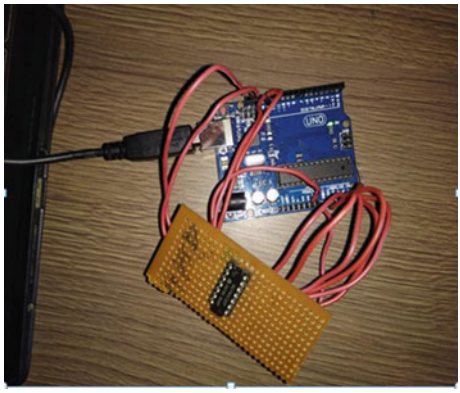


Fig. 3 Hardware connection of IC

is executed. The input for digital ICs will be binary 1 and 0 and will verify the truth table for each IC.

3.1 Hardware of the Proposed System

The various components of the IC tester are given below:

3.1.1 Arduino Uno Board

The Arduino Uno is an ATmega328-based microcontroller board. Fourteen digital input/output pins (six of which can be used as PWM outputs), six analog inputs, a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button on this board. As shown in Fig. 3 used for testing digital IC with the help of Arduino Uno board and microcontroller. It comes with everything you will need to get started with the microcontroller; plug it into a computer with a USB cable or power it with an AC-to-DC adapter or battery (Table 2).

3.2 Software of the Proposed System

For high-productivity research, development, and analysis in the industry, MATLAB is the tool of choice. Toolboxes are a type of application-specific solution available in MATLAB. Toolboxes are essential for most MATLAB users since they allow you to understand and use specific technologies. It will enable you to answer many numerical problems in a fraction of the time that writing a program in a lower-level language would take. MATLAB aids in understanding and implementing topics in various fields, including engineering and mathematics. To make robots or human-like

Table 2 Summary of arduino uno board

Summary	
Microcontroller	ATmega328
Operating voltage	5 V
Input voltage (recommended)	7–12 V
Input voltage (limits)	6–20 V
Digital I/O pins	14
SRAM	2 KB (ATmega328)
EEPROM	1 KB (ATmega328)
Clock speed	16 MHz analog input pins
DC current per I/O pin	40 mA
DC current for 3.3 V pin	50 mA
Flash memory	32 KB

machines, it takes a combination of electrical engineering, electronics engineering, and computer science, to mention a few. Researchers and engineers in robotics use MATLAB to create and modify algorithms, simulate real-world systems, and write code automatically—all from a single software environment.

MATLAB is a high-performance programming language for technical computing. It is a package that combines computers, visualization, and a programming environment. MATLAB is also used in this study to create a user-friendly GUI and an efficient application for various IC testing.

Microchip's ATmega328 is an 8-bit, 28-pin AVR microcontroller that uses RISC architecture and includes a 32 KB flash-type program memory. The Atmega328 microcontroller found in the Arduino UNO, Arduino Pro Mini, and Arduino Nano boards. It has a 1 KB EEPROM memory and a 2 KB SRAM memory. It features eight pins for ADC operations, which come together to form Port A (PA0–PA7). Cost-effectiveness, low power consumption, programming lock for security, and valid timer counter with independent oscillator are only a few of its outstanding qualities.

The ATmega328 microcontroller executes the program in the flash memory. The tester operation is implemented using a MATLAB application. To check the condition of the IC in the given list, the program switch case concept is employed. When the selected IC number matches the program's supplied cases, the system operates and displays the IC's status and gate state. Figure 4 shows the hardware and software interface of proposed IC tester.

4 Results of Proposed IC Tester

After the hardware interface and simulation of in built GUI in MATLAB, the different digital IC were tested and results are shown in figure below. The GUI firstly initialized

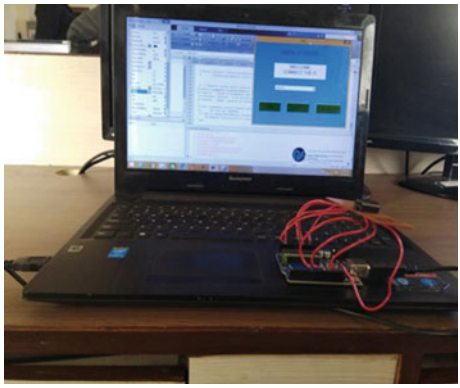


Fig. 4 Software stimulation of proposed IC tester using MATLAB

Fig. 5 Initialization of GUI with interfacing of testing board

with the massage display WELCOME text on the GUI screen when the supply resumed and connected board is activated, as shown in Fig. 5.

Following the placement of the IC, the processor generates the appropriate input to the IC, displays the IC's status, and displays the gates involved in the IC. When the IC is connected to the Arduino board, for example, the GUI screen shows that the IC 7432 is in GOOD condition and the state of the gate, as shown in Fig. 6.

In case, if the IC is not responding to the result as per the source code appropriately, it will display the BAD condition and the status of the gate as shown in Fig. 7.

5 Conclusion

Analog circuits work with continuously changing signals, referred to as analog signals. In contrast, digital circuits work with discretely variable signals, referred to as digital signals, in which the signal only has two levels: 0 and 1 (binary digital

Fig. 6 Displaying GOOD condition of IC 7432

Fig. 7 Displaying BAD condition of IC7432

signaling). It is critical to conduct quality control testing on several samples of integrated circuits to verify that they are serviceable. The GOOD and BAD results for integrated circuits tests, as well as their source of supply, can be maintained in a database. Without the use of an external keypad, this integrated circuit tester designed to test digital ICs. However, it can readily expand this device to accommodate many ICs as long as the memory allows. There is no need for rewiring; instead, you will need to make the necessary changes to the source code. If you need to test more pins, you can utilize two microcontrollers in a master slave configuration. One was used for testing, and the other was used for driving.

References

1. Tarkunde MA, Prof. Shinde AA (2012) IC Tester using 89S52 microcontroller. Int J Comput Eng Res (ijceronline.com) 2:24–27
2. Darji C (2021) Design and Implementation of Microcontroller Based Digital Logic Gate IC Tester. https://doi.org/10.13140/RG.2.2.36066.25288/1

3. Pabiania MD, Peralta JT, De Luna LAT (2014) Microcontroller based design of digital IC tester with multi-testing and loop testing functions. Int J Artific Intell Mechatronics 2:183–186
4. Electronics for you(EFY) (2017) Engineering Projects for You
5. Houcque D (2005) Book "Introduction to MATLAB for Engineering Students". Northwestern University, version 1.2
6. Kumar S (2016) Digital IC tester (for Industries and Engineering Colleges). Panjab University, IEEE Maker Project, UIET
7. Rai A et al (2021) Real time vehicle recognition system for secure environment. J Phys: Conf Ser 2007, 012049:1–7
8. Taha Y, Elamin Y (2015) Microcontroller based integrated circuit tester. Int J Eng Res Appl, 5(2):118–122, ISSN: 2248–9622
9. Shylaja C et al (2021) Modelling and simulation of 16-bit vedic multiplication using FPGA. J Phys: Conf Ser 2007, 012049:1–7
10. Devika S et al (2016) IC tester using Pic microcontroller. Int J of Res Eng Technol 05:442–445
11. Kanade DG et al (2019) Digital IC tester using arduino. Int J Trend Res Develop 6:102–104

A Simulation-Based Model for Detecting Fall of Elders Using Proteus

R. Harsha, Swikriti Gupta, and K. N. Veena

Abstract The People's body undergoes a variety of physical changes as they become older. We know that majority of fall occur at home. The elderly, disabled, as well as ordinary people will fall to the ground due to poor balance, fainting, and other causes. In the proposed system, piezoelectric sensors are integrated under the tiles of the bathroom floor. The values of the piezoelectric sensor output voltage and the pre-set threshold voltage are compared by the Arduino microcontroller. A recue signal will be sent to the guardian. In case of emergency, the guardian can activate the solenoid lock system, which is already built into the doors, to unlock them. A simulation model is being formulated here to check how the hardware system might work.

Keywords Piezoelectric sensors · Arduino o microcontroller · GSM module

1 Introduction

Per year, more than 1.6 million U.S. adults are treated in emergency rooms for fall-related injuries, according to a senior health report about falls. Fractures, loss of freedom, and even death are the most common outcomes of these injuries. Ageing and obstacles in the home setting are two of the most common causes of falls.

People's bodies undergo a variety of physical changes as they age, making them more vulnerable and susceptible to falls. For example, vision deteriorates over time, making it difficult to explore one's surroundings and spot tripping hazards. Medications are prescribed to reduce the symptoms of ageing as the human physicality deteriorates. This, ironically, raises a person's risk of falling because certain medications reduce mental function.

R. Harsha (✉) · S. Gupta
Department of Electronics and Communication Engineering, Dr. Ambedkar Institute of Technology, Banglore, India
e-mail: harsh4no1@dr-ait.org

K. N. Veena
School of Electronics and Communication Engineering, Reva University, Banglore, India

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_41

The average person's living area is full of possible fall hazards, which is why the majority of falls occur at home. Causes and risk factors are a senior health article. Slippery floors, clutter, dim lighting, shaky furniture, obstructed pathways, and pets are all common hazards in the house.

The injured person will be saved and revived in this paper using an IoT-based device. Elderly and disabled people, as well as regular people, will fall to the ground as a result of falling, fainting, or other causes. As a result of the fall, the individual may sustain physical injury. If the injury is severe, such as blood loss, the injured person's life must be saved immediately.

As a result, our devices can easily detect a person's fall inside the bathroom. The proposed system incorporates piezoelectric sensors beneath the bathroom floor tiles. When a person walks, his or her feet inevitably press against two or three tiles, generating low voltage in the piezoelectric sensors. If a person falls to the ground, which means they land on tiles, their body can exert pressure on more than four tiles simultaneously. As a result, the piezoelectric sensors are subjected to increased pressure and generate increased voltage. The voltage levels have been validated, and the microcontroller has established the voltage threshold value. The Arduino microcontroller compares the output voltage of the piezoelectric sensor to a pre-set threshold voltage. If the performance of the piezoelectric sensors meets the threshold, the controller sends a signal to the GSM module, which sends a message or calls a registered number of someone who can assist the injured person, or a community alarm is sounded. The rescuer's number and the rescue message should have already been programmed into the microcontroller and GSM module via AT commands. When the ring sound is heard, the SIM card has been successfully registered in the network via the GSM module. As a result, the rescuer will receive a message or phone call indicating that a parent or friend has been injured, allowing him to immediately go to the house and save them. If the individual enters bathrooms, he or she may respond to the message or call in the same manner to activate the solenoid lock system already installed in the doors. This paper sought to not only detect physical collapse in people, but also to bring technology closer to them in order to save their lives.

2 Background

Resuscitating injured people requires careful accident management and patient safety. Falls and fainting in toilets and stairwells account for 32% of home deaths, according to the World Health Organization. These forms of household injuries raise the risk of death for injured individuals, and they must be handled to prevent deaths. Accidental nervous episodes can cause people to become paralysed, lose their memories, or fall into a coma.

As a result of technology improvements, certain health issues have grown more prominent in society. In terms of health and emergency injury management, the elderly and disabled must be cared for at home. It needs to be tracked in real time to

ensure quick response and recovery, as well as a quicker response to save the injured individual. Even if the accident is mild, if there is blood loss or major injuries, it will result in death.

2.1 Rationality Behind Choosing the Project

The World Health Organization predicts that 28–35% of persons aged 65 and up die each year, with 32–42% of those aged 70 and up dying. As people age and become more fragile, they are more likely to fall. Indeed, falls rise exponentially as a result of age-related biological changes, resulting in a high rate of falls and fall-related accidents in ageing populations. The incidence of falls-related accidents is anticipated to climb by 100% by 2030 unless prompt preventative measures are taken. Assistive technologies that can help minimise this major health issue are a societal imperative in this situation. The use of fall detectors is currently being researched.

The amount of time that older people spend on the floor after falling is a crucial element that fall detectors can assist to lessen (long lie). When determining the magnitude of a fall, this is a vital component to consider. Many elderly people are unable to stand without assistance, and lying down for lengthy periods of time can cause hypothermia, dehydration, bronchopneumonia, and pressure sores. This is especially important if the person lives alone or has lost consciousness as a result of a fall. Falls in seniors was written by SR Lord, C Sherrington, and Hb Menz. Several studies on the long lie have been undertaken in order to assess the risk, causes, and prevention techniques. Long lies are linked to higher mortality rates in the elderly and are suggestive of fragility, disease, and social isolation, according to the researchers. Despite the fact that there was no immediate harm from the fall, over 20% of patients admitted to the hospital after a fall spent an hour or more on the ground, and their 6 month morbidity rates were extraordinarily high.

3 Literature Survey

Taramasco et al. [1] conducted research on a sensor-based monitoring system for detecting falls in the elderly. Using various types of sensors and algorithms, infrared sensors were utilised to warn elderly persons about probable falls. Thermal sensors with a poor resolution for categorising falls and alerting caregivers. This innovation was built on a thermal sensor array for senior citizens who live alone. Three recurrent neural network algorithms were implemented: Bi-LSTM, LSTM, and GRU, all of which performed well when detecting falls, with the model Bi-LSTM standing out. He planned to test comparable heat sensors in additional locations in future to validate the device in less controlled clinical settings.

Lingmei Ren [2] et al. to combat fall difficulties have conducted research on detecting and preventing falls utilising sensing techniques and classifier models. New

fusion-based methods based on a combination of several sensors or classifier models were employed to improve the performance of fall-related systems. The systems used a variety of sensors to collect meaningful signals for further processing and analysis, as well as a variety of analysis algorithms to handle the data. Fall detection systems detect body impact and sound an alarm, whereas fall prevention systems extract gait data to provide an early fall alert.

Minvielle et al. [3] have conducted research into fall detection utilising smart floor sensors and a monitoring system based on sensitive floor sensors. A piezoelectric polymer sensor was used. The sensor is a thin film that is directly applied to the concrete and then coated with a flooring solution. As a result, when a person falls, the floor sensors detect it and send a warning signal. Fall detection was monitored because monitoring floor vibrations is less intrusive than other passive detection approaches.

Santiago et al. [4] have conducted research on a fall detection system for the elderly in order to generate a rescue signal in order to save the person. A wearable device and a cell phone were the two main components of the system. The wearable was capable of communicating with a cell phone that was within a 100-ft radius. When the wearable device detects a fall, it sends an alert to the cell phone, which subsequently sends an alert to the user's emergency contacts. In addition, this device included a panic button that could be used to notify emergency contacts if the user believed a fall was imminent.

Jiangpeng Dai et al. [5] a pervasive fall detection system using mobile phones, a prototype system called per fall detector is constructed on the Android G1 phone as a platform for pervasive fall detection system, and trials are undertaken.

Tang1 et al. [6] fall detection sensor system for the elderly fall detection sensor devices are worn by the elderly on their waists or wrists. The sensor device has an algorithm built in that detects and measures the users' body location. The system will sound an alarm if the sensor devices identify the body position in falling mode.

Jahnvi Gupta et al. [7] posture detection and underlying sensor technology falling detection devices like sensors will be worn by elderly people on their wrist or waist. The embedded algorithm will run by a sensor device, and it will detect the geographic location and body measures of the person. If body position detected by sensor device is in falling mode the system embedded alarm will be triggered. There are a variety of sensors that can be used to collect data about human posture. Using machine learning techniques, this data is then utilised to classify a posture as healthy or unhealthy. After the model has been trained, data will be categorised in real time after being sent to it through Wi-Fi. An overview of how IoT networks can be used to recognise human posture is offered. Fall detection is aided by the underlying sensor technology utilised in IoT networks engaged in human posture.

Hoa Nguyen et al. [8] fall detection using smartphones to enhance safety and security of older adults at home, to detect falls, motion sensors record accelerations that are compared to pre-set criteria. Data on acceleration is gathered from 15 participants. In three axes, the change in the root-sum-of-squares of the total acceleration is examined. A fall event is detected when the amplitude moves from the lower to the higher threshold. By taking location into account, the algorithm has been enhanced

to avoid false positives then a fall is detected, this programme gives a button on the user interface for users to acknowledge whether it is a real fall or not.

Neha Mangala et al. [9] fall detection and gesture alert system making progress on a gesture recognition algorithm, to establish signals for fall detection using sensors using an array of capacitive proximity sensors and pressure sensors, the gesture recognition and fall detection detector produces a map of the positions. Accelerometer-equipped gyroscopes sensors are incorporated into a single board in the form of an MPU 6050 to receive an alert on the built web application. The movements are subsequently processed using the gesture recognition algorithm. The data acquired by the sensors is transferred to a database, where it is synthesised and presented in the form of graphs to the user or a helper. The technology assists in providing immediate support and necessary care to the user, who may take this portable device to any area with ease.

Authors in [10–13] proposed that how by using image processing techniques to capture photos of people and subsequently detect visual falls. Pervasive detection, affordability, and acceptance are all issues with these systems were discussed in [14–16]. Within the monitoring environment, which is expensive to set up, the detection area is limited. Also jeopardised is people's privacy were elaborated in [17–20].

4 Objectives

1. To design a simulation model to detect a person's fall inside the bathroom.
2. To develop a model to identify the fall of a person without any near human body contact.
3. To cut down on the amount of time elderly people spend lying on the floor after falling.
4. To send out a rescue signal via GSM communication in order to save the individual.

5 Methodology

5.1 Proteus Design Suit

As shown in Fig. 1, the Proteus design set is a proprietary software programme for electrical design automation. Electronic design professionals and technicians use the programme to develop schematics and electronic prints for printed circuit board manufacture. Labcenter Electronics Ltd created it in Yorkshire, England, and it is available in English, French, Spanish, and Chinese.

PCB, in 1988, the company's chairman, John Jameson, produced the first version of what is now the Proteus design suite for DOS. In 1990, schematic capture support was added, followed by a transition to the Windows environment. In 1996,

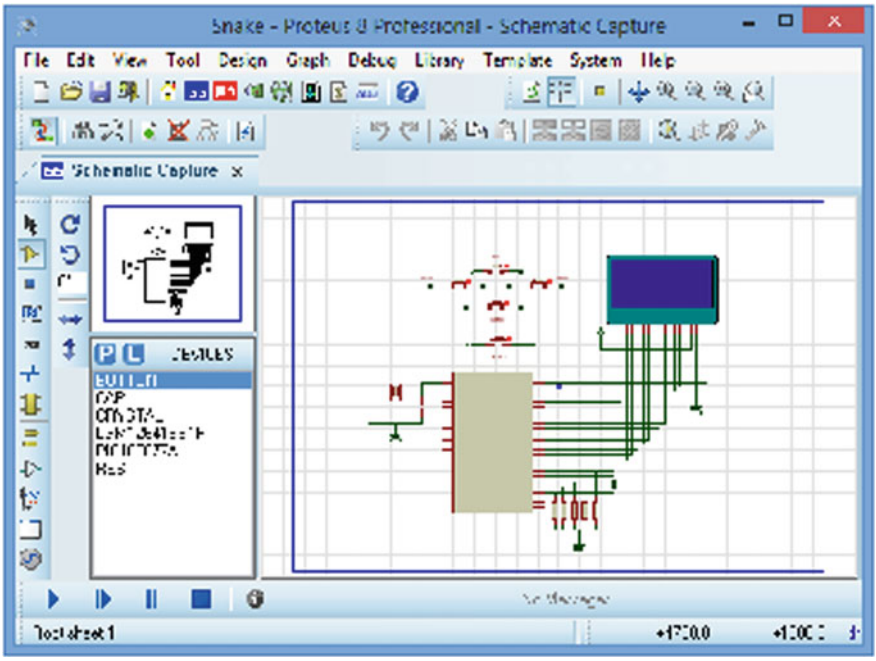


Fig. 1 Proteus design suite

Proteus added support for mixed-modal mode SPICE simulation, and in 1998, it added support for microcontroller simulation. In 2002, shape-based auto routing was enabled, and in 2006, 3D board Visualisation was included as part of a major product upgrade. MCAD import/export was implemented in 2015, and a separate IDE for simulation was created in 2011. In 2017, support for high-speed design was added. Feature-driven product releases happen every two years, whereas maintenance-based service packs happen whenever they are needed.

The Proteus build suite is a Windows programme for recording, simulating, and designing PCB layouts. It comes in a number of forms, depending on the size of the designs being created and the microcontroller simulation requirements. Auto routing and rudimentary mixed mode SPICE simulation capabilities are included in all PCB design solutions. Both design simulation and the design phase of a PCB layout project employ the Proteus design suite's schematic capture functionality. As a result, it is a crucial feature that all product configurations incorporate.

The PCB layout module gets a net list of connection information from the schematic capture module. It makes use of this information, as well as user-defined design requirements and several design automation tools, to help create error-free circuit boards. Although the size of the design is limited by the product configuration, PCBs with up to 16 copper layers can be created.

The 3D viewer module gives you a three-dimensional perspective of the board you are working on, as well as a semi-transparent height plane that mimics the enclosure.

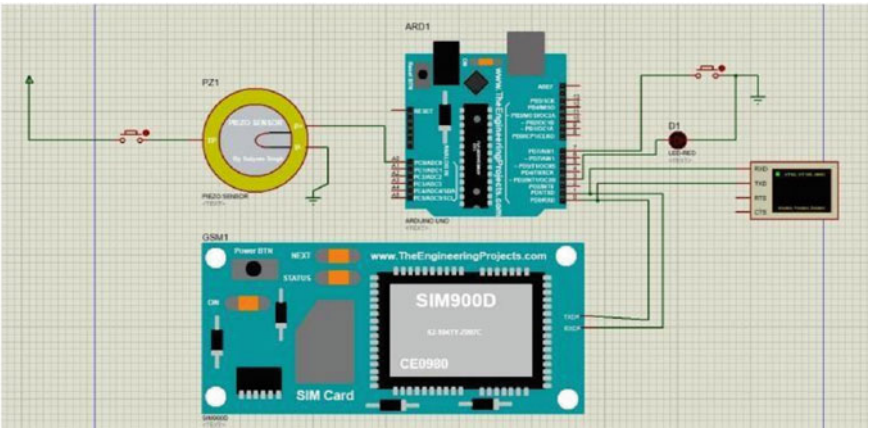


Fig. 2 Circuit diagram

For precise board positioning and installation, the STEP file may be imported into mechanical CAD software like AutoCAD or Autodesk.

5.2 Circuit Diagram

See Fig. 2.

5.3 Components

5.3.1 UNO Arduino

The most advanced version of the Arduino Proteus Library, including a total of six Arduino board:

- Arduino Uno
- Arduino Pro Mini
- Arduino Mega 2560
- Arduino Mega 1280
- Arduino Nano
- Arduino Mini.

In total, seven Arduino Proteus Libraries was created. First, separate Proteus Libraries for each of these six boards was created, and then integrated them all in the seventh library.

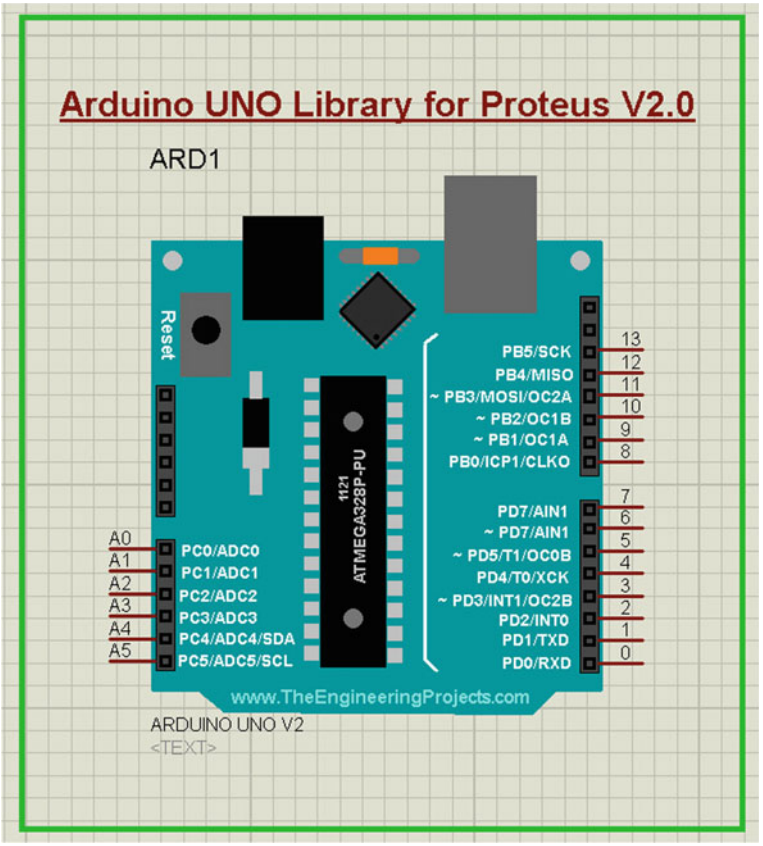


Fig. 3 Arduino UNO

So, if only Arduino UNO is needed, download the Arduino UNO Library, but if want to work with many boards, download the combined version (7th).

As shown in Fig. 3, the Arduino UNO is the most popular board from the business. The ATmega328P microcontroller is used in the Arduino UNO. Digital and analogue input/output (I/O) pins, shields, and extra circuitry make up the board.

5.3.2 Piezoelectric Sensor

Sensors are electronic devices that detect or sense specific physical quantities in the environment. All possible inputs include light, heat, motion, moisture, pressure, and vibrations. The output is usually an electrical signal that matches the applied input. The output signal is transferred through a network to be processed further, and this output is used to connect the input [14].

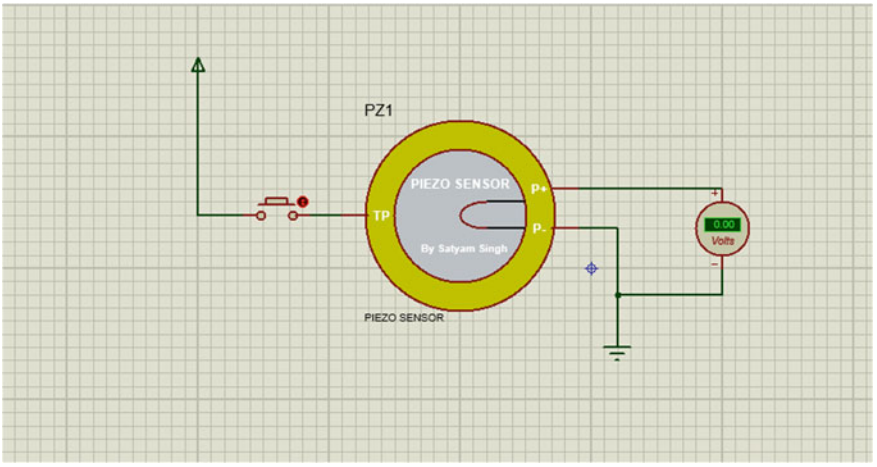


Fig. 4 Piezoelectric sensor

It is frequently necessary to replicate the circuit's ancestor before creating it. The emulation of the circuit attitude when it is executed in real life is known as simulation. When simulating any circuit, it is critical to remember that the simulation copy of all the devices used in your circuit must be available in the software's library. In a nutshell, it is for simulating the management of devices using a simulation copy [15].

As shown in Fig. 4, a piezoelectric sensor is one that responds to piezoelectricity's fundamental. In addition, piezoelectricity is a state in which an element generates electricity when mechanical stress is applied to it. Not all elements, however, exhibit piezoelectric properties. Natural single-crystal elements have a higher solidity than synthetic single-crystal elements, but they are slightly more sensitive. A quartz piezoelectric sensor is typically employed in the project [16].

5.3.3 GSM Module

A GSM modem, also known as a GSM module, is a physical device that connects to a distant network using GSM mobile phone technology. They are essentially equivalent to a standard mobile phone in the eyes of the mobile phone network, including the necessity for a SIM card to identify them to the network. TTL-level serial interfaces are usually provided by GSM modems to their hosts [17].

- First and foremost, get the GSM library for Proteus.
- Following the download, you will receive three files:
- GSM Library TEP.IDX
- GSMLibraryTEP.LIB
- GSM Library TEP.HEX
- Place all these files in the Libraries folder of your Proteus software.

- Now, as shown in Fig. 5 that we have three GSM modules in our Proteus software.
- These three GSM modules are identical in terms of functioning, as you can see. They all have two pins, TX and RX, and they only differ in colour, but they all work on the same serial port.
- The first is light blue in hue, which is also our theme colour, the second is green, and the third is red.
- There are a few basic commands that this version of the GSM module now supports.
- It will not be able to send or receive SMS at this time because these features have not yet been introduced.

5.3.4 Virtual Terminal

As shown in Fig. 6, virtual terminal is an important part of Proteus. When working with serial modules like GSM, GPS, or X-Bee, it is quite beneficial. It reads data from the serial port (DB9) and sends data to the serial port (DB9).

Virtual terminal is a Proteus application that can be used to examine and transmit data to and from the serial port (DB9). In Windows XP, there is a built-in programme called hyper terminal that can be used for the same purpose, but this virtual terminal is extremely handy for Windows 7 users. The Arduino UNO only has one serial port on pins 0 and 1, whereas the Arduino Mega 2560 has four. Serial port is supported by PIC microcontrollers, as well as 8051 microcontrollers [18].

To send and receive data through a serial port, Proteus uses a virtual terminal. A serial port is a 9-pin connection that is often seen on computers and is utilised in embedded system projects for data exchange. In student projects, data is often transferred from hardware to computer via serial port, and the user then constructs an application on their computer to view the data in some representable manner.

Now, there are some testing processes in projects that, if utilised correctly, can be quite beneficial, and these testing steps demand the use of specific tools to test the process. Consider the case where a student created the hardware to send data to the computer and the application to receive it, but when he tested it, he received no data. The student is uncomfortable at this point since they do not know where the mistake is, thus it is possible that there is a hardware or software error [19].

To be assured, he must now test both of them separately, which is where the virtual terminal comes into play. Connect your equipment to the computer first, then operate it to see if the virtual terminal displays data. This terminal can be used in a variety of ways. If you wish to make a circuit in Proteus that contains a serial port, for example, you can include this terminal and test it before moving on to the hardware [20].

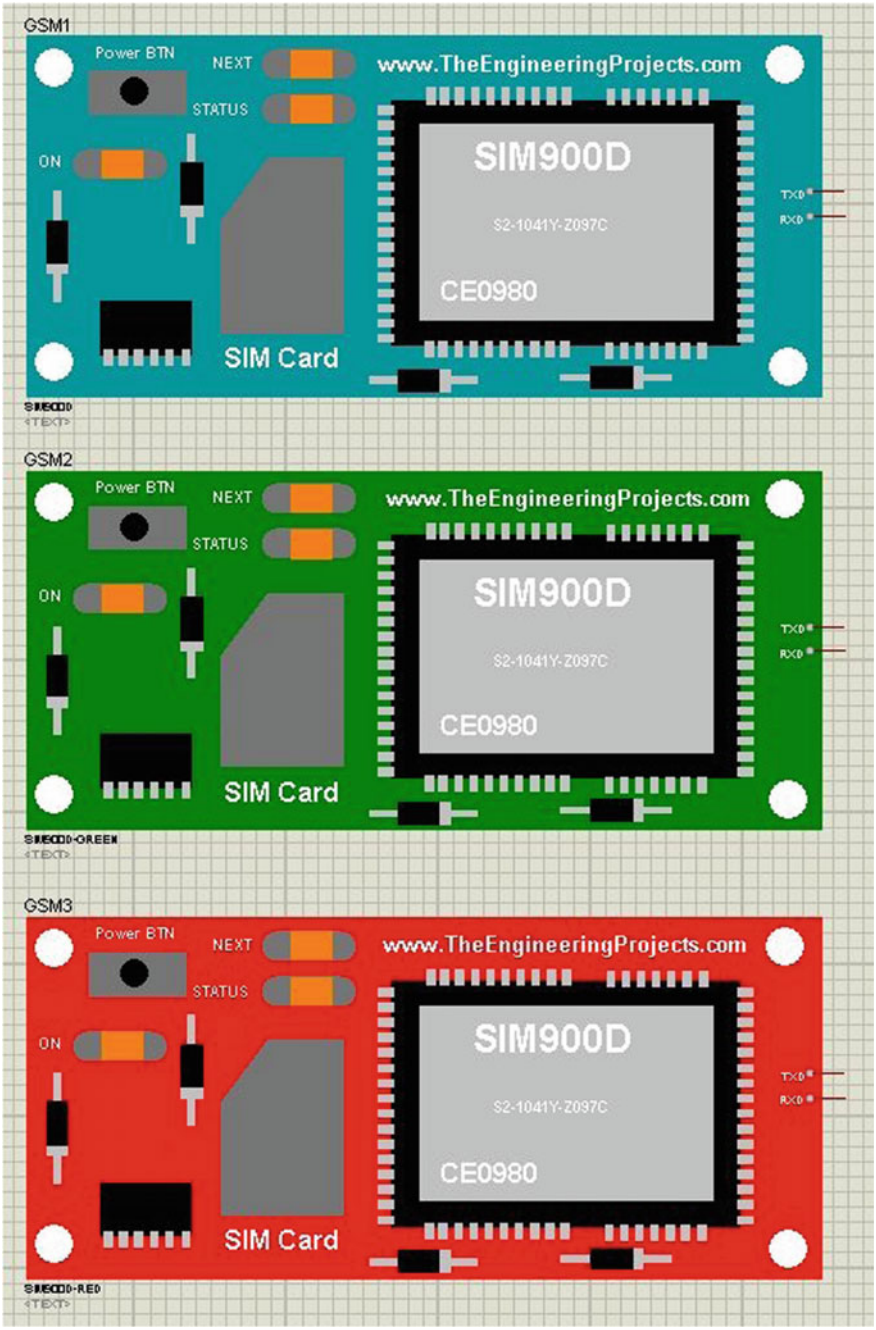


Fig. 5 GSM module

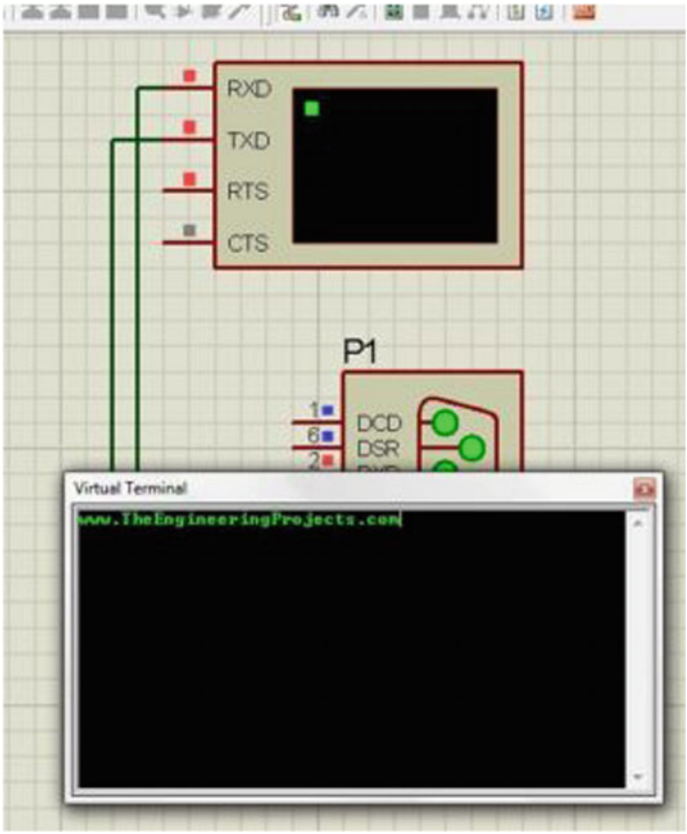


Fig. 6 Virtual terminal

6 Advantages

- The designed systems play a major role in monitoring frail elderly people at homes.
- Signals through GSM communication helps in quick rescue of the person.
- The system is cost-effective, affordable, and implementation is not complicated.
- Fast reliable method for sending signals for help.
- Fall detection systems will provide people with an extra measure of security.

7 Novel Approach

The technique was designed with older people who live alone in their homes in mind. We describe a novel method for detecting falls in the home environment in this

research. We discovered that our study findings indicate 95% accuracy for non-fall activities and 96.67% accuracy for fall activities using the proposed new approach. The proposed simulation model attained an average accuracy of 98.3% with a sensitivity of 100% and a specificity of 97.7%. There are just a few commercial fall detection products on the market. The recommended solution is cost-effective due to the use of low-cost piezoelectric sensors. Sensors are placed beneath the tiles, which can be done during construction if needed, making installation simple. If not, two to four tiles can be easily removed for a low cost and afterwards used to place sensors. Because we rely entirely on technology here, the safety lines are integrated into a system that does not rely on an individual's ability to request assistance. This method does not require any external equipment or belts, and it will not interfere with people's daily lives.

8 Conclusion

As the number of senior people who live alone and are at risk of falling increases, fall detection and prevention become more important in their everyday life. It is vital to be able to identify and report a fall in real time. We may deduce from the previous explanation that we can accurately detect a human fall by passively monitoring the floor vibration patterns and measuring the change in the threshold voltage. Because the sensors can simply be placed in the area to be monitored (e.g., in the bathroom), this health monitoring device with piezometric sensors has the advantage of being contactless and non-invasive, with no need for people to wear or carry any devices or change their normal habits and behaviour. This system's main purpose is to detect when someone has fallen and activate the alarm system by calling the person in touch.

References

1. Chen J, Yang S Application of proteus software in MCU teaching
2. Rahman MM, Jannat NE, Islam MO, Salakin MS Arduino and GSM based smart energy meter for advanced metering and billing system
3. Mutinda M, Kamweru P Arduino uno, ultrasonic sensor HC-SR04 motion detector with display of distance in the LCD
4. Development of Home Intelligent Fall Detection IoT System based on Feedback Optical Flow Convolutional Neural Network, Special Section on Intelligent Systems for the Internet of Things
5. TheFootstep Power Generation using Piezo- Electric Transducer. Int J Eng Innovative Technol (IJEIT) 3(10), ISSN: 2277–3754
6. Intuitive IoT-based H2U healthcare system for elderly people. In: 2016 IEEE 13th inter-national conference on networking, sensing, and control (ICNSC), Mexico City, 2016
7. Harrou F, Zerrouki N, Sun Y, Houacine A (2017) Vision-based fall detection system for improving safety of elderly people. IEEE Instrum Meas Mag 20(6):49–55

8. Yu X (2008) Approaches and principles of fall detection for elderly and patient. In: Proceedings on IEEE 10th international conference on e-health networking, applications and services (HealthCom), 42–47
9. Wang C et al (2016) Low-power fall detector using triaxial accelerometry and barometric pressure sensing. IEEE Trans Ind Informat 12(6):2302–2311
10. Sixsmith A, Johnson N (2004) A smart sensor to detect the falls of the elderly. IEEE Pervasive Comput 3(2):42–47
11. Fan X, Zhang H, Leung C, Shen Z (2017) Robust unobtrusive fall detection using infrared array sensors. In: Proceedings on IEEE international conference on multisensor fusion and integration for intelligent systems (MFI), Daegu, South Korea, pp 194–199
12. Karim F, Majumdar S, Darabi H, Chen S (2018) LSTM fully convolutional networks for time series classification. IEEE Access 6:1662–1669
13. Graves A, Fernández S, Schmidhuber J (2005) Bidirectional LSTM networks for improved phoneme classification and recognition. In: Proceedings international conference on artificial neural networks. Berlin, Germany, Springer, pp 799–804
14. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation
15. Mashiyama S, Hong J, Ohtsuki T (2014) A fall detection system using low resolution infrared array sensor. In: Proceedings on IEEE 25th annual international symposium on personal, indoor, and mobile radio communication (PIMRC), pp 2109–2113
16. Taniguchi Y, Nakajima H, Tsuchiya N, Tanaka J, Aita F, Hata YA falling detection system with plural thermal array sensors. In: Proceedings on international conference on soft computing and intelligent systems (SCIS) and 15th international symposium on advanced intelligent systems (ISIS).
17. UValpoLabitec. Repository Github. Accessed 14 Jul 2018. [Online]. Available: https://github.com/UValpoLabitec/FallDetectionModels
18. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
19. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings on International conference on machine learning 2015, pp 448–456
20. Bravo D, Hughes E (2017) Las personas mayores que viven solas en Chile. Centro UC Encuestas y Estudios Longitudinales, Santiago, Chile, Tech. Rep

Simulation of a Microgrid with OpenDSS an Open-Source Software Package

Anjali Jain, Ashish Mani, and Anwar S. Siddiqui

Abstract The power system network is a very complex and exorbitant entity. Moreover, the condition becomes more complex once a distributed generation, renewable energy sources, energy storage devices are added to the grid/microgrid. For adding any source in the grid, proper analysis for load flow, short circuit studies, transient studies, etc., is required to understand the performance of different components. It is dangerous to directly load the network to check the performance of the system, hence, an alternative mechanism is adopted wherein the real-time scenario is modeled virtually using some simulation tools, and thereafter, different scenarios can be analyzed for different cases. Simulation using simulation tools is a well-known technique to assess the performance of the system in a virtual environment. Simulation tools artificially create models, and it helps in analyzing the performance of the system over multiple scenarios. Hence, an effort has been made in this paper in compiling the non-exhaustive list of simulation software package to tackle microgrid capabilities, wherein microgrid is comprised of distributed generation and renewable energy sources. Also, a detailed review has been done to discuss the features and shortcomings of different simulation software packages. After a detailed review of simulation software packages, it has been found that OpenDSS works well for distribution systems or microgrids and works efficiently not only for balanced systems but also for unbalanced systems. Therefore, a tutorial has been presented to model microgrids with the help of OpenDSS. Apart from this, an example using IEEE 13 node feeder is discussed with distributed generation and renewable energy sources to showcase the performance of OpenDSS.

A. Jain (✉) · A. Mani
Department of Electrical and Electronics Engineering, Amity University, Uttar Pradesh, Noida, India
e-mail: anjalijain.121@gmail.com

A. Mani
e-mail: amani@amity.edu

A. S. Siddiqui
Department of Electrical and Electronics Engineering, Jamia Milia Islamia, New Delhi, India
e-mail: assiddiqui@jmi.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_42

Keywords Open-source simulation software · OpenDSS · Renewable energy sources · Distributed generation · Distribution system · Microgrid

1 Introduction

Simulation software packages that have been developed by different researchers have their features as well shortcomings. No tool can be perfect for all types of scenarios simultaneously, and hence, the user needs to identify which tool will be best suited for its requirements. There is a huge list of tools that are available in the market. The three main sub-categories discussed in [1] are transient state simulation tools, real-time software application, and grid analysis software.

Transient state simulation tools are time-based simulation tools that have no limitation to the duration for running the simulation that is why they are also being called off-line simulations. Few of such simulation tools are EMTDC, Simpowersytem, EMTP, etc. The main objective for such tools is to evaluate the impact of the things which cause a transient state in a system like a lightning stroke, any short circuit contingency, turning off or turning on of circuit breakers, etc., in the power system network. Next, we have real-time software applications which are time-based synchronized in real-time with external inputs/outputs such as RTDS, Opal-RT, and Hypersim. Third, certain grid analysis software applications in the frequency-based state allow for load flow calculations, such as PSAF, ETAP, or PSS/E.

Nonetheless, all the tools which are available in the market are not accessible to users for free. Some tools available for simulation are available in the market as either proprietary or open-source. Proprietary tools come with a license for which the user has to pay for the users such as Power System Simulator (PSS), PowerWorld2, and DigSilent PowerFactory3. Such tools are provided with the detailed lab manual along with not only support and maintenance from the service provider but also with timely updates for the package. Proprietary tools are well-validated before providing to the users. The main issue with the use of such tools is the heavy cost associated with them which is difficult at times for individuals.

To overcome the mentioned issue, users are shifting to the use of free open-source tools available in the market. The discussion in this study is focused on understanding the features of open-source tools and identifying the best suitable tool specifically for modeling microgrid or distribution systems. The main objective of this paper is to evaluate different available open-source software tools for modeling a microgrid (MG) with renewable energy sources (RESs) and electric vehicles (EVs). Although there is a big list of available software, in this paper, the popular software used for modeling the distribution system or microgrid are discussed.

Section 2 briefly introduces open-source software, and then, an effort has been made to discuss different available open-source simulation tools which are available for the distribution system and microgrids. After analyzing different software tools,

modeling using OpenDSS is discussed in Section 3. Section 4 utilizes the implementation of OpenDSS to model an IEEE 13 node network with distributed generation and renewable energy sources. Section 5 is for results and discussion. The paper is concluded in Sect. 5.

2 Open-Source Software

Open-source software is freely available and can be easily downloaded to do simulation studies. The benefits associated with open-source software are that the students who are unable to buy the license to carry out their simulation work for proprietary tools can take help from open-source tools. As already discussed, there is a huge ocean of simulation software packages but no software package stands perfect for all types of studies. Some software packages work well for load studies, others may work well for transient studies. Some may be more legitimate for transmission system, whereas other performs well for a distribution system.

The problem under study is the modeling of microgrids or distribution systems, hence a detailed study has been made on a few open-source software packages depicting their feature and shortcomings when used in the analysis of microgrid or distribution systems. Following are the detailed features and shortcomings of few software packages. The open-source simulation software which is discussed here are:

- InterPSS
- MATPOWER
- psst
- APREM
- Anylogic
- Repast
- RAPSim
- Power World Simulator
- GridLAB-D
- OPenDSS

The detailed discussion about the above-mentioned tools are discussed in the next section.

2.1 Internet-Based Power System Simulator (InterPSS)

InterPSS is an Internet-based open-source tool that has been developed by a group of researchers from China, the United States, and Canada. The main motive behind the development of this software simulation package for designing, analyzing, and simulating real-world engineering problems with the support of Internet technology.

InterPSS edition is broadly classified into two editions: One is called desktop edition (DE) and the other is web edition (WE). DE has a graphic editor which can manage various projects at the same time, i.e., we may run load flow and transient stability simulation at the same time and hence saves time. The result obtained by the software can be easily annotated on the single line diagram which makes it easy to understand. The web edition of InterPSS is in Chinese. With a web browser, we can solve load flow, short circuit study, etc., and straightaway upload the data file to the site. Once the data is online, the different calculations for load flow or short circuits can be performed online from the server and the results can be fetched at the users' screen [2].

Features of InterPSS

- As the name suggests, the Internet is utilized as a virtual platform for this software package.
- It can be integrated with the components which are developed using other platforms.
- The programming language used in InterPSS is Java.
- The other features supported by this software package are forecasting, CIM6 modeling, and cascaded failure analysis.
- Different formats can be generated easily as per the requirement.

Shortcomings of InterPSS

- The language used for InterPSS is Java, wherein earlier version of java was outperformed by Fortran, C, or C++ .

2.2 Matpower

MATPOWER is an open-source simulation software package that is based on MATLAB. It can perform a high-level set of power flow, optimal power flow (OPF). This software package is released in the year 1997. Its easy integration with MATLAB makes it popular in the industry as well as in academia. There are readily available scripts for load flow analysis, continuous power flow, optimal power flow, unit commitment, etc. [3].

Features of MATPOWER

- MATPOWER is used for the OPF formulation
- It is compatible with Octave.
- Codes are simple to understand and easy to modify.
- It is utilizing a high-level programming language which makes it suitable for lengthy numerical computation for power system simulations.
- It has easy-to-use functions for different matrices such as admittance matrix, and susceptance matrix.

Shortcoming for MATPOWER

- Although MATPOWER is free and open-source software, it requires a licensed MATLAB for that wherein MATLAB is paid software.

2.3 Psst

Psst is abbreviated for power system simulation toolbox. It is also an open-source toolbox specially designed for simulating and analyzing power system networks [4].

Features of psst

- It is freely available and can be modified easily.
- It can also be utilized for large scale models in power system.
- It has wide application for optimal power flow, security constrained economic dispatch, etc.
- It uses Python as its programming language and has fast prototyping and development
- The programs can be easily customized and hence expanded according to requirement.

2.4 Pandapower

Pandapower is an open-source simulation software package that is used for static as well as quasi-static balanced power system networks [5].

Features of pandapower

- The solver used in pandapower is Newton Raphson.
- The capability of pandapower includes modeling of constant current loads, grids with multiple reference nodes, and a connectivity check.
- Different application of pandapower includes optimal power flow (OPF), state estimation, short circuit calculations, etc.
- The language used is Python, and hence, it is easy to use and can be easily extended with other libraries.
- Python libraries are used to handle input and output parameters.
- Pandapower has application in automated time-series calculation, state estimation, short circuit calculation, and also graph analysis using NetworkX library.

2.5 Anylogic

Anylogic is also an open-source simulation package that has wide application not only in power systems but also in other problem areas which are based on discrete event simulation, agent-based simulation, and system dynamics simulation [6].

Features of Anylogic

- This model can model and simulate the network at different levels. Hence, network can be studied from the lowest level to the highest level of analysis.
- It is being used for various applications such as power flow analysis, economic dynamics of generation and consumption of power, and performance analysis of different entities,
- The interface is easy to use.
- It has a user-friendly graphical user interface to set up a model
- Java is used to extend the default models easily.

Shortcomings of Anylogic

- It does not allow WCF communication
- It requires the study of lengthy documentation which makes it difficult to use at times.

2.6 Repast

Recursive Porous Agent Simulation Toolkit (Repast) was developed in 2008 by Macal, C. M., & North, M. J. to develop models which are flexible and interacting [6].

Features of Repast

- It is compatible with Linux, Apple Mac, and MS Windows.
- It has wide application not only in the field of power systems but also in supply chains, social science, infrastructure, etc.
- Agents utilized to design a system make the model more autonomous and intelligent.
- The result can be visualized in 2D or 3D as per requirement.
- It can be interfaced with MATLAB with the help of a software package called the JMatlink package

Shortcoming of Repast

- Repast is a flexible tool but the increased flexibility also leads to more complex interaction. Hence, it becomes difficult for non-programmers to use this software simulation package.
- Also, to work on this software, good knowledge of the JAVA language is required.

2.7 RAPSim

RAPSim is an open-source simulation package especially used in microgrids with renewable energy sources. Hence, it is also called SGS (It has very little application as compared to other software packages such as OpenDSS and GridLab [6, 7].

Features of RAPSim

- The programming language used in JAVA.
- It has a friendly GUI
- It is easily adaptable by users
- Time simulation thread allows up to one-minute resolution.
- Results obtained after running the software are in CSV formats.

Shortcoming of RAPSim

- Battery models are not present in RAPSim.

2.8 Aprem

APREM is abbreviated for Analyse Paramétrique des Réseaux Électriques [parametric analysis of power grids] with MATLAB. It functions with MATLAB [1].

Features of APREM [1]

- The method used to solve electrical circuits is an increased node matrix.
- Users can easily introduce changes in loads, generators, and different other components of the grid to do simulation studies.
- It is easy to learn.
- In APREM, a first grid is defined and then any number of components can be easily associated with grid elements.
- It is widely used in an academic institution because of its easy use.

Shortcomings of APREM [1]

- APREM offers fewer functions than other software packages such as OpenDSS and GridLab-D.
- Grid automation is not possible with APREM.

2.9 PowerWorld Simulator

PowerWorld simulator is an open-source software package created by T. Overbye to teach power system students [8].

Features of PowerWorld Simulator [8]

- It has an intuitive GUI; hence, it provides a user-friendly interface.
- A large number of elements is shown on the ribbon interface which improves the visibility of each element and hence becomes easy to interact with.
- It has a model explorer to help manage different case information and displays at a glance.
- It supports different models like generators, wind generators, load models, etc.

Shortcomings of PowerWorld Simulator [8]

- PowerWorld simulator works for only balanced power flow.
- It is designed for transmission system simulations

2.10 GridLAB-D

GridLab-D is an open-source software package to model power flow especially for residential load modeling and the retail market [7, 8].

Features of GridLab-D [8]

- It is specifically used for residential modeling of components such as dishwashers and refrigerators.
- It is used for load flow, time-series calculation, volt/VAR management, and power loss calculation.
- The pre-processing and post-processing of the data can be done with CSV files.
- It is compatible with MATLAB, MySQL
- It can be used both for an unbalanced and balanced system.
- One of the special features of GridLab-D is the climate module which helps in the integration of renewable energy sources precisely.

Shortcomings of GridLab-D [8]

- It can work for simple radial networks. It cannot handle complex networks.
- Its interface is not user-friendly.
- It does not have a built-in conversion tool to export or import data from other platforms.
- There is not enough documentation available in the literature to learn this software.

2.11 OpenDSS

OpenDSS is an open-source simulation tool that is specifically designed for the distribution system [2–4]. The salient feature of OpenDSS is the interconnection of distributed generation to the utility grid [8, 9].

Features of OpenDSS [8]

- It is used for modeling of components such as capacitor, line, reactor, and transformer, generator, PV system, storage, and load, capacitor bank control, storage controller, and DG dispatch controller.
- The pre-processing and post-processing of the data can be done with CSV files.
- It is compatible with MATLAB, Excel, VBA, etc.
- It can be used both for an unbalanced and balanced system.
- In addition to power flow analysis, it is capable of performing harmonic analysis, short circuit analysis, and transient stability analysis.

Shortcoming of OpenDSS [8]

- It does not support online editing, zooming, or dragging, and dropping shapes.

3 Modeling of the Network Using OpenDSS

OpenDSS has been picked in this study because of its said features in the above discussion and its ability to work in an unbalanced system along with renewable energy sources, demand response, and electric vehicles. Moreover, OpenDSS can be easily integrated with MATLAB for optimizing objective functions easily [8, 9].

The structure of OpenDSS has been shown in Fig. 1, wherein the main simulation engine is getting inputs from user-written dynamic link libraries (DLL). The simulation engine interacts with communication interfaces and text scripts. Output scripts can be saved in CSV files which are further used to pre-process or post-process the data.

The detailed component blocks for the open DSS main simulation engine have been shown in Fig. 2. The main simulation engine is employed to model different power delivery elements such as line, transformer, capacitor, and reactor. It is also having power conversion elements for load, generator, V source, I source, and storage

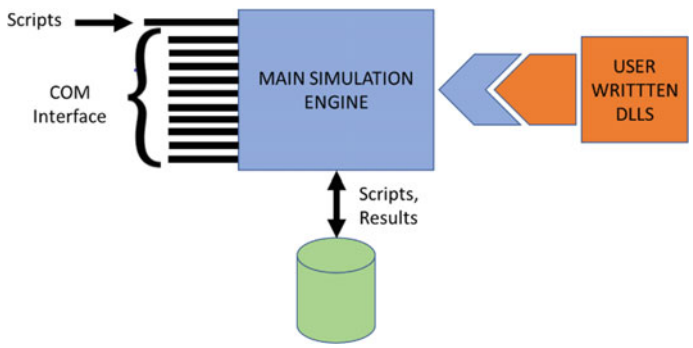


Fig. 1 Structure of OpenDSS [8]

devices. The control in the modeling may be regulator control, capacitor control, relay control, reactor, and fuse control. There also exist meters in OpenDSS engine such as energy meter, voltmeter, ammeter, and sensors. OpenDSS is also used to model general parameters such as line codes, line geometry, wire data, load shapes, and TCC curves. The new circuit is created having different components, and the solution is found based on a constant voltage, constant admittance, or constant current mode. The communication interface is also available in the simulation engine to interact with third-party applications such as Excel, VBA, and MATLAB.

As discussed in the previous section OpenDSS is a good option to be utilized for modeling of the distribution system and microgrid and hence, step-by-step procedure has been showcased for modeling of different components of a microgrid.

Step 1: The first step is to create a circuit and name the circuit.

The command used to create a new circuit with the name "test" is

$$\text{new circuit.test} \tag{1}$$

Step 2: After creating the circuit, start adding the loadshape of different loads.

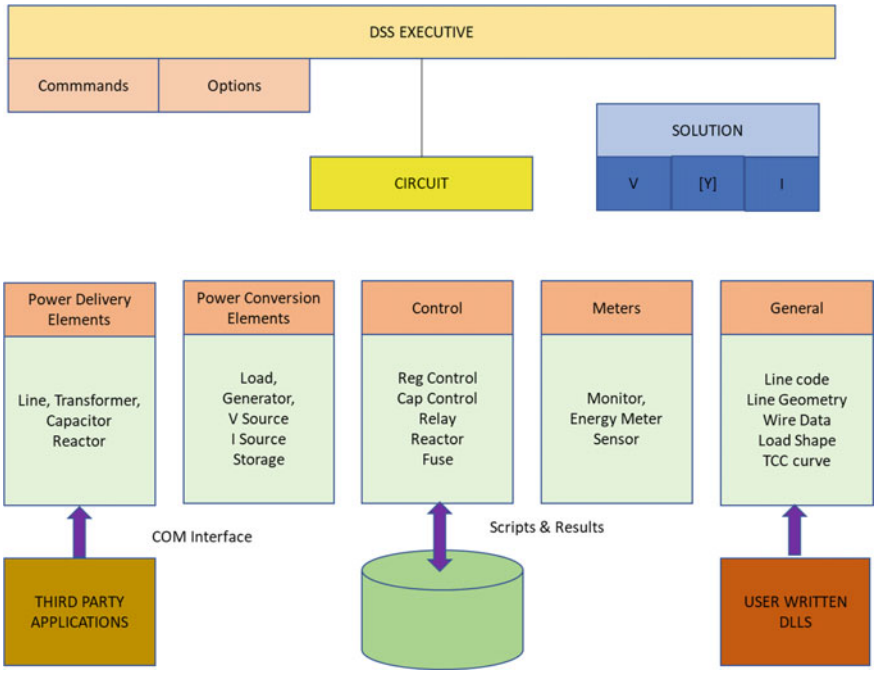


Fig. 2 Detailed simulation engine of OpenDSS [8]

These data can be uploaded using a CSV file. CSV file needs to be stored in the same folder where your program for OpenDSS is being kept. The data for load shapes are loaded in the object called "loadshape," e.g., in this case "loaddemand." The CSV file from where the information is fetched for load shape here is named as "demand." In the given command, number of intervals is 8760 for 1 year.

$$\text{new loadshape.loaddemand Npts} = 8760 \text{ Csvfile} = \text{demand.csv hour} = 1 \tag{2}$$

Step 3: Add the load shape for different generating units such as wind and solar.

Here, the loadshape name given for wind and solar are "wind" and "solar" and the CSV files used to load data for 8760 h is "datawind" and "datasolar." CSV files as under for a year data where the name of loadshape given is "wind" and CSV file used for feeding data for 8760 interval points with the duration of 1 h.

$$\text{new loadshape.wind Npts} = 8760 \text{ Csvfile} = \text{datawind.csv hour} = 1 \tag{3}$$

$$\text{new loadshape.solar Npts} = 8760 \text{ Csvfile} = \text{datasolar.csv hour} = 1 \tag{4}$$

Step 4: After that, step-by-step addition of grid elements is done.

For adding grid elements, we need to first name them.

(1) *Addition of source*

The source element is connected between a bus and ground via "vsource" object. The name defined here for adding a source is "source." Then, the bus number needs to be defined that is at which bus generator is connected, after that the phases are to be defined that is at which phase generator is connected, then the nominal KV (which is considered 14.43 kV for 25 kV of line voltage) is to be defined and its p.u. value

$$\text{new vsource.source Bus1} = \text{sourcebus Phases} = 1 \text{ basekv} = 14.43\text{pu} = 1 \tag{5}$$

(2) *Addition of line*

The line element is created via "Line" object. Here, the line is named as "Line1." After that position of line, i.e., between which buses the said line is connected. Here, line is connected between bus1 (named sourcebus) and bus (named B1) after that phases are defined and then the user needs to enter the value of R1 and X1 electrical parameters:

$$\text{new Line.Line1 Bus1} = \text{sourcebus Bus2} = \text{B1 Phases} = 1 \\ \text{R1} = 0.29\,\text{X1} = 0.9875.\text{new Load} \tag{6}$$

(3) *Addition of load*

The load element is created via "Load" object. The name used here is "load1," after that bus location at which that load is connected between that bus and ground, here it is "B1" then phase at which this load is connected, here it is phase 1, then nominal voltage, nominal capacity in kv and kw, power factor and then load type, also the time of simulation needs to enter which is yearly in this case considered for 8760 intervals points of one-hour duration each for loadshape which is "loadshape1" in our discussion.

$$\text{new Load.load1 Bus1} = \text{B1 Phases} = 1\ \text{Kv} = 14.43\ \text{Kw} = 10,000\ \text{Pf} \\ = 1\ \text{Model} = 6\ \text{Yearly} = \text{newsystem} \tag{7}$$

(4) *Addition of wind generator*

The generator object is created Vis "Generator" object. The name given for generator here is G1 which is connected to bus B1, phases are 1, nominal kv and kw are defined, after that power factor, model and then yearly load shape of "wind" is taken as defined earlier.

$$\text{new Generator.G1 Bus1} = \text{B1 Phases} = 1\ \text{Kv} = 14.43\ \text{Kw} \\ = 10,000\ \text{Pf} = 1\ \text{Model} = 1\ \text{Yearly} = \text{wind} \tag{8}$$

Step 4: Add energy meter to record values

Add energy meter where the parameter needs to be recorded. This is achieved by adding an object "energymeter" to the element where the parameters need to be recorded. Here, the name given to energy meter is M1, which is connected to Line L1 and the local data is being recorded only.

$$\text{new energymeter.M1 Line.L1 LocalOnly} = \text{yes} \tag{9}$$

Step 5: Specify simulation parameters

After the addition of energy meters, the simulation parameters need to be specified. In this example, the mode considered is yearly, it may be hourly, monthly. Then, test case name is specified where the result will be stored in the form of csv file in this case it is "Anjali," set the duration for running the simulation, i.e., 1, it means that simulation will run for 1 year of time, also define demand interval and overload report which will highlight if there is any overloading in any element.

$$\left.\begin{array}{c}\text{set mode} = \text{yearly}\\ \text{set casename} = \text{param}\\ \text{set year} = 1\\ \text{set demandinterval} = \text{true}\\ \text{setoverloadreport} = \text{true}\end{array}\right\} \tag{10}$$

Step 6: Launch network

After inputting all the required details, launch the circuit and save the results

$$\text{Solve} \tag{11}$$

$$\text{edit energymeter.M1 action} = \text{save} \tag{12}$$

Step 7: The simulation results are recorded in an csv file

As specified in the simulation parameters which is "param" as specified.

4 Implementation of OpenDSS for Simulating a Microgrid

After discussing the steps to model different elements of the power system using OpenDSS, let us start simulating a microgrid for the following test case.

4.1 IEEE 13 node system (without any distributed generator)
4.2 IEEE 13 node system with conventional generators
4.3 IEEE 13 node system with conventional generator and solar and wind power generator

The test case considered here is IEEE 13 node circuit which is an unbalanced network. The data set available for this test system can be obtained from [10]. Here, in this case, no distributed generation is considered. There are 13 nodes as defined in Fig. 1. The two capacitors are installed on nodes number 673 and 611 (Fig. 3).

As shown in Fig. 4, one distributed generator is added to node 680. The details for DG can be found from [10] and are shown in Fig. 5. The wind and solar generator daily outputs would follow the multiplication of rated kW values and duty schedules generated from historical data files. Wind output is essentially a stochastic disturbance and solar power contributes only during the daytime [10]. The pictorial representation of the modified feeder is shown in Ref. [10].

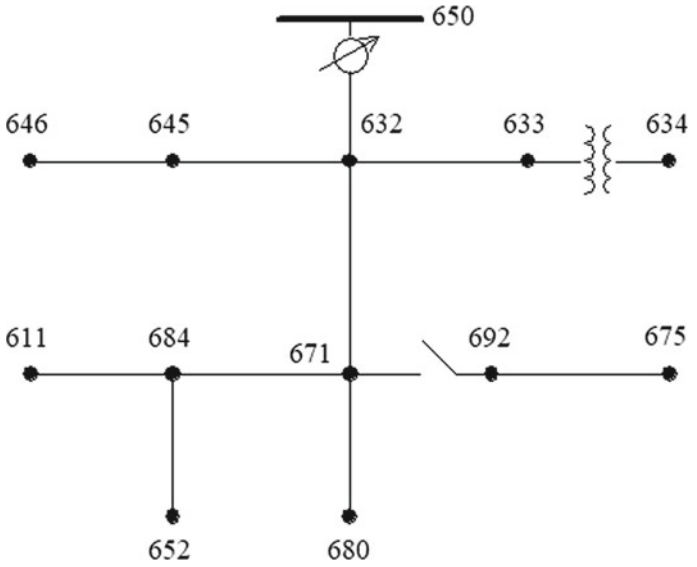


Fig. 3 IEEE 13 node Feeder [10]

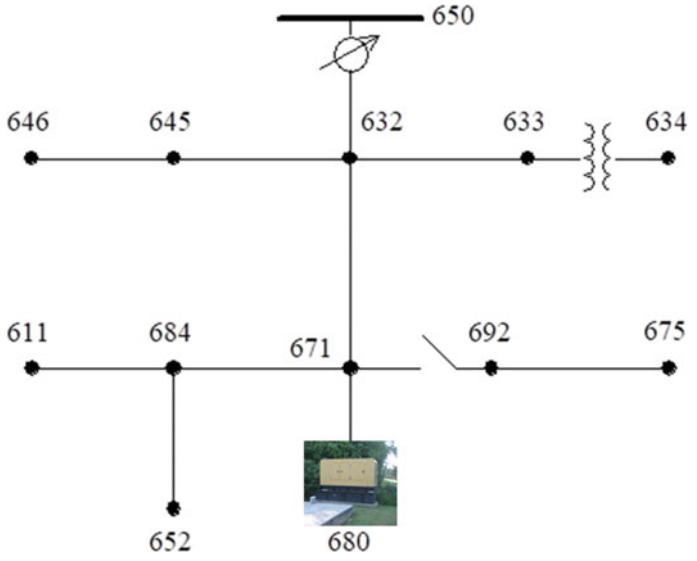


Fig. 4 IEEE 13 node Feeder with the distributed generator at node 680 [10]

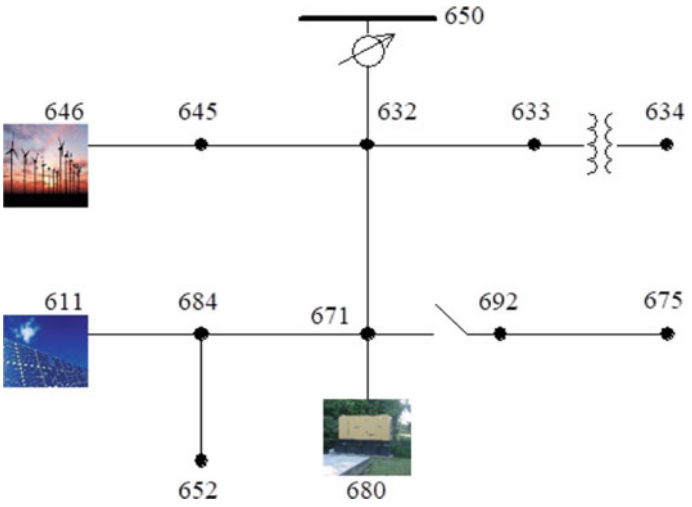


Fig. 5 IEEE 13 node Feeder with RES and distributed generator [10]

5 Result and Discussion

5.1 IEEE 13 Node System

In IEEE 13 node feeder, the circuit has 38 devices, the total number of buses considered in OpenDSS is 16, and the control iteration took to solve power flow is 3. The voltage lies between the following range of $1.056 \leq V_{p.u.} \leq 0.96083$.

The active power and reactive power of the network which is calculated are 3.56721 MW and 1.73659, respectively. The active and reactive losses encountered in the system are 0.112409 MW, (3.151%) and 0.327912 MVAR (18%), respectively.

5.2 IEEE 13 Node System with Conventional Generators

When one conventional generator is added to IEEE 13 node feeder at the bus no. 680, the total devices become 39, and control iteration is still 3. The voltage lies between the following range of $1.0499 \leq V_{p.u.} \leq 0.96708$. It has been found that minimum and maximum points of $V_{p.u.}$ is approaching toward unity.

Moreover, the active power and reactive power of the network which is calculated are 2.5311 MW and 1.60015 MVAR, respectively. This shows that the reactive power available in the network has been increased. Also, the active and reactive power losses obtained with the help of OpenDSS are 0.717954 MW (2.837%) and 0.196112 MVAR (12%), respectively. This shows that by the addition of a conventional generator the active and reactive power losses have been reduced.

5.3 IEEE 13 Node System with Conventional Generator and Solar and Wind Power Generator

When solar power and wind power generator are added to IEEE 13 node feeder along with a conventional dispatchable generator, the total number of reading devices by OpenDSS is 41. The control iteration required by the software package is 4.

The voltage lies between the following range of $1.0499 \leq V_{p.u.} \leq 0.96708$. It has been found that minimum and maximum points of $V_{p.u.}$ is approaching towards unity.

Moreover, the active power and reactive power of the network which is calculated are 1.91394 MW and 1.5496 MVAR, respectively. Also, the active and reactive power losses obtained with the help of OpenDSS are 0.0530248 MW (2.77%) and 0.141045 MVAR (9%), respectively. This shows that by the addition of a conventional generator the active and reactive power losses have been reduced.

6 Conclusion

This paper has been written to understand the need for open-source software packages and to discuss different packages available for researchers to carry out their work. After a thorough analysis of different available software packages, it was found that OpenDSS works well for an unbalanced distribution system. Hence, OpenDSS modeling steps are discussed in detail and thereafter, a small unbalanced benchmark system of IEEE 13 node feeder is discussed. The OpenDSS is utilized to model an IEEE 13 node feeder, and then, load flow is run to find the losses in the system. After that, a conventional dispatchable generator is added to the test system and it has been found that the addition of a conventional generator reduces the active and reactive power losses. The losses are further reduced by adding solar and power generators to the system. The detailed result has been showcased in Table 1 as well.

Table 1 Load flow results obtained by OpenDSS for different test cases for static power Flow

Parameter	IEEE 13 node circuit	IEEE 13 node circuit with conventional generator	IEEE 13 node circuit with conventional generator and solar and wind generator
Max p.u. voltage	1.056	1.0499	1.0499
Min p.u. Voltage	0.96083	0.96708	0.97582
Total active power	3.56721 MW	2.5311 MW	1.91394 MW
Total reactive power	1.73659 MVAR	1.60015 MVAR	1.54496 MVAR
Total active losses	0.112409 MW	0.0717954 MW	0.0530248 MW
Total reactive losses	0.327912 MVAR	0.196112 MVAR	0.141045 MVAR

Future work includes establishing the MATLAB interface with an optimization algorithm with the test case and obtain the optimized cost of generation.

References

1. Moffet MA, Sirois F, Beauvais D (2011) Review of open-source code power grid simulation tools for long-term parametric simulation. Canmet Energy. Tech Rep 137:1–41
2. Zhou M, Zhou S (2007) Internet, open-source and power system simulation. In: 2007 IEEE power engineering society general meeting, IEEE, pp 1–5
3. Zimmerman RD, Murillo-Sánchez CE, Thomas RJ (2010) MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans Power Syst 26(1):12–19
4. Krishnamurthy D (2016) 2016psst: an open-source power system simulation toolbox in python. In: 2016 North American power symposium (NAPS) IEEE, pp 1–6
5. Thurner L, Scheidler A, Schäfer F, Menke J-H, Dollichon J, Meier F, Meinecke S, Braun M (2018) pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. IEEE Trans Power Syst 33(6):6510–6521
6. Kondoro A, Dhaou IB, Rwegasira D, Kelati A, Shililiandumi N, Mvungi N, Tenhunen H (2017) Simulation tools for a smart micro-grid: comparison and outlook. In: 21st conference of FRUCT association, pp 1–7
7. Jdeed M (2016) Comparison of the smart grid simulation tools RAPSim and GridLAB-D, pp 1–53
8. Gao DW, Muljadi E, Tian T, Miller M (2017) Software comparison for renewable energy deployment in a distribution network. No. NREL/TP-5D00–64228. National Renewable Energy Lab. (NREL), Golden, CO (United States), pp 1–17
9. https://smartgrid.epri.com/SimulationTool.aspx: Accessed on 15 Sept 2021
10. Ma J, Yang F, Li Z, Qin SJ (2012) A renewable energy integration application in a microgrid based on model predictive control. In: 2012 IEEE power and energy society general meeting, pp 1–6

Heart Condition Forecasting Using High-Accuracy ANN Algorithms

Pranshi Gupta, Piyush Yadav, Rajeev Agrawal, Varun Sharma, Ritik Dixit, Shubhashish Varshney, and Komal Kashish

Abstract Increasing life expectancy has many challenges like day-to-day illness, chronic diseases, allergies, and cardiac arrest. In reference to technical science, there are many wearable devices to measure basic health parameters but their accuracy and precision are also a challenge. The heart plays a pivotal role in the human body so measuring its heartbeat accuracy and precision is the highest priority because little mistakes and manipulation can cause severe illness, fatigue, and death of a person. In this paper, our major aim is to improvise the heart disease prediction by using the best possible algorithms of high accuracy. To predict cardiac arrest there are different technologies used such as artificial intelligence (AI) and machine learning (ML), and to improve accuracy and precision we took the combination of four most efficient algorithms, i.e., K-nearest neighbors, light gradient boosting machine, support vector machine, and decision tree classifier. This research paper throws light on the improvision done with different algorithms for predicting the probability of heart diseases, and after processing our dataset, we came to know that SVM and KNN are the two best of all algorithms that can ever predict the result with the highest accuracy.

Keywords Artificial intelligence · Machine learning · Technology · Support vector machine

1 Introduction

In today's era life expectancy is decreasing due to an increase in the high rate of chronic diseases. As life is busy in this internet world, people forget and do not have time to examine their health at a regular time interval, and regular usage of

P. Gupta (✉) · P. Yadav · R. Agrawal · V. Sharma · R. Dixit · S. Varshney
Department of Electronics and Communication Engineering, G L Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: pranshigupta46@gmail.com

K. Kashish
Hamburg University of Technology, Hamburg, Germany

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_43

tobacco, alcohol, and overstress with less or no physical activity causes the high rate of chronic diseases [1]. For elders, it is tough to regularly visit hospitals for a regular health examination. This imposed a major challenge to both the medical as well as engineering sectors. There are many reasons for the shortening of life like gene disorder, allergy, accident, etc. Numbers of people in this world are suffering negatively from heart diseases (HD) [2]. Shortness of breath and body weakness are some of the major symptoms of HD [3]. As the developing countries are not advanced in terms of modern technologies, the early diagnosis and treatment of HD are getting difficult [4]. The lives of many can be saved if the disease gets detected earlier. It is noted that the amount of money spent on medical compensation due to cardiac health issues is very high [5]. The accuracy and computation of the HD diagnosis method are not reliable [6]. The machine learning techniques give effective results to control the HD death ratio and also nowadays development of smart recorders provides a chance to record the information at various time intervals [7]. A report from a World Health Organization (WHO) indicates millions of people die due to overweight increased cholesterol levels, high blood pressure, lungs cancer, severe heart stroke, brain hemorrhage, kidney failure, and 17.9 million people die each year from cardiovascular disease which is 31% of all deaths in the world [8]. To overcome and help the people to move in their day-to-day work environment is the modern technologies, and ICT can be an effective solution to overcome this problem [9, 10]. The major drawback of bio-mediation is that the process is complicated and time-consuming [23]. In this work identification of HD is easy by machine learning (ML). ML models include K-nearest neighbor (KNN), support vector machine (SVM), linear regression (LR), decision tree (DT), naïve Bias (NB), and artificial neural network (ANN) for identification of HD. ANN is the core of AI and acts as a powerful tool because of its brain-like function where the neuron is the fundamental unit of a network that gives output according to the number of inputs that is fed on the neuron. This type of problem can be solved by using ML. In ML several predictive and classification algorithms such as RM, DT, LR, SVM, and ANN, the occurrence of cardiac arrest can be predicted by certain algorithms as mentioned. In this research paper, the proposed solution is to find out the best algorithm that can be used for the prediction of heart diseases in such a critical situation of COVID-19 as the curve is increasing rapidly [25]. Various parameters that are used include blood pressure, chest pain, heart rate, blood sugar level, etc. Using different algorithms on this dataset and then comparing their improvised accuracies through our different approaches, it is verified that SVM and KNN are the two best possible algorithms for the prediction of heart diseases. This paper gives us brief details about the process of heart diseases prediction along with the accuracies of different algorithms. Section 2 investigates the related work. Section 3 is the methodology and analysis consisting of the outcome of this research work. Section 4 presents the results and discussions. Finally, Sect. 5 gives the conclusion of the paper.

2 Related Work

In the twenty-first century IoT-based system and machine learning play a key role in the medical field and appliances. To develop IoT-based medical appliances many researchers are working on it. Some research-related work is mentioned below. The authors in [11] proposed a method for fetal health monitoring and identified that accuracy up to 99.9% can be achieved by using the ANN-based method efficiently. In [12] authors suggested a more accurate solution in doctor diagnosis by using several machine learning models rather than having a single model. In [13] authors used multilayer perception in this proposed system that was used because of its efficiency and accuracy. In [14] authors implemented smart devices, wireless sensors, and remote monitoring systems that can identify the real-time status of a heart patient and can be used in the future as wireless technology enabled smart home health monitoring systems. The work-study in [15] gives an idea of how machine learning techniques are used to process the raw data and give a new and insightful solution toward heart disease. The research work indicates that the error analysis can help to select the appropriate model for their respective data [24]. The authors in [16] prove that the machine learning algorithm performed on the dataset can be very effective in predicting the occurrence of cardiac arrest in patients. The work in [17] shows how machine learning shows effective results in making decisions and predictions for treating heart diseases such as random forest (RF), decision tree (DT), naïve Bayes (NB) are used to classify various heart disease attributes in predicting heart diseases. In [18] authors implemented and designed a reasonable, effective, and portable health monitoring system that is based on the concept of the wireless health monitoring system.

3 Methodology and Analysis

3.1 Data Sources

In this proposed software system, the dataset used to test as well as train the model was taken from an online community website of AI and Data Science named as Kaggle. It is ancillary to Google which gives us a practical dataset. The dataset consists of almost 105+ data inputs based on real-life readings. We had used 75% trained data and 25% test data for the prediction of heart disease in our model.

3.2 Architecture of the Model

This machine learning model basically completed its work in four major stages as shown in Fig. 1. Firstly, we collected the data from Google-based Kaggle data

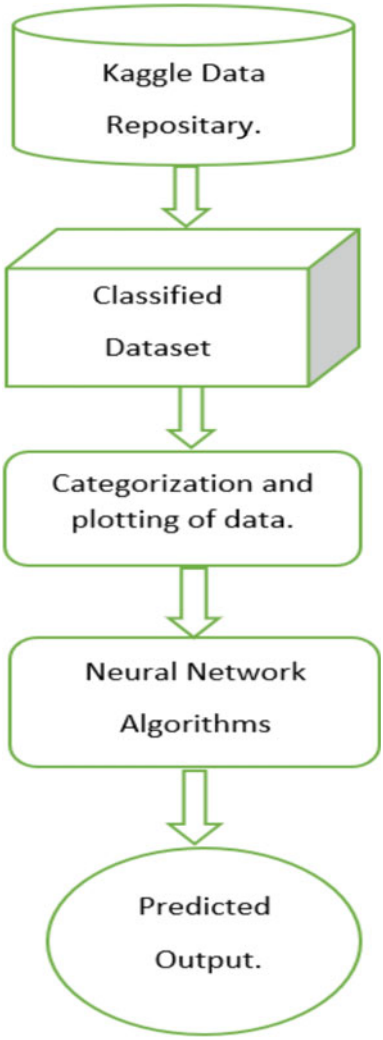


Fig. 1 Structure for representing the model

repository for training the model. Then, we have to classify this data using the decision tree. After classification, our next process is the categorization and plotting of data with the help of KNN. Now classification and categorization are followed by two other algorithms LGBM and support vector machine for enhancing the efficiency of the model. Finally, with all the above processes, the model predicts the output.

3.3 Algorithms Explanation

This section presents a detailed study of algorithms used in this software model. We had basically used four different and most efficient algorithms for the prediction of output namely:

i. Decision tree
ii. K-nearest neighbors
iii. Light gradient boosting machine
iv. Support vector machine.

Decision Tree Classifier: Decision tree is a type of supervised learning used for regression and classification and can work on categorical data or criteria depending on which it gives output in the form of Yes or No, 1 or 0, and True or False. Starting from the top it works down and down classifying the data and resulting in leaves. The test is performed on the features/attributes. This algorithm generates a model or tree-structured classifier based on these attributes. After that, we gave an unknown/unlabeled input to this classifier, and these models allot or give a class to our input based on its past dataset.

The accuracy of the decision tree is very high compared to other classification algorithms, and the main reason behind this is its tree-like structure which consists of the root node, leaf node, and interior node, and hence decisions made can be easily explained with this model. Now the main point here to note is that how can we decide the most efficient attribute to classify the data or the attribute with the most informative value. The answer for that can be explained using the term entropy. Entropy refers to the impurity or any randomness in the information being processed. It is calculated in terms of entropy as given in [13], for which Eq. (1) calculates the entropy of each sample specifically in terms of probability of occurring and not occurring, and the entropy of the whole sample space or dataset is calculated in Eq. (2), and then we used the difference of entropy of sample space and entropy of each sample for finding the information gain in terms of probability as represented in Eq. (3) Entropy of Sample Space

$$- p(\text{yes}) \log_2 p(\text{yes}) - p(\text{no}) \log_2 p(\text{no}) \tag{1}$$

If number of yes = number of no, i.e., $p(s) = 0.5$ Entropy$(s) = 1$
If it contains all yes or all no, i.e., $p(s) = 1$ or 0 Entropy$(s) = 0$
Entropy of Sample

$$\sum_{i=1}^{c} - p(\text{yes}) \log_2 p(\text{yes}) \tag{2}$$

$$\text{Information Gain} = \text{Entropy of Sample Space} - \text{Entropy of a Sample} \tag{3}$$

K-nearest Tree: KNN is the other supervised learning algorithm used to train this model in order to classify the data. It is a non-parametric and lazy learnings classification method with low efficiency as compared to other algorithms. KNN works on 'Feature Similarity', assuming that a similar type of data exists in close proximity or nearest neighborhood. Hence to find the location of the new point on the graph, the similarity between the existing points is compared based on the distance we use like mathematical logic of calculation of Euclidean, Manhattan, or hamming distance. Considering the case of regression, here it is decided by the mean average value for the new instance. The most common method for calculation is the Euclidean distance which is used to calculate the distance between two points in Euclidean space calculated in Eq. 4.

$$\text{Euclidean distance} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \tag{4}$$

Where (p_i, q_i) are the data points.
Now, the first step is to select the value of K, by similar featuring and then using the nearest neighborhood method which had the minimum Euclidean distance select the value of k. And then assigned a class to a new test instance.

Light Gradient Boosting Machine: Light GBM is a framework model and open-source library used for increasing the accuracy of the model. This model is based on the decision tree concept and helps to reduce memory consumption. Basically, it works on two principles that are gradient-based one-side sampling and exclusive feature bundling. The major advantage of using LGBM is it can easily handle a bigger amount of data and is preferred not to be used for small datasets as it can lead to overfitting. It provides 100+ parameters. GBM model can also be used as the final model for the predictions in the case of regression as per the requirements. It supports regression, binary classification, multi-class classification, cross-entropy, and LambdaRank [22].

Support Vector Machine: Our last and most used and flexible supervised learning algorithm is the support vector machine algorithm. SVM is the most used and efficient algorithm for classification due to its ability to handle highly multi-dimensional and categorical data easily [22]. In SVM, firstly we allocate each data item as a point in n-dimensional space to classify the data points into the classes. Support vectors are basically the coordinates of specific observation. SVM works better in the clear margin of separation and high-dimensional spaces. Classifiers of SVM use a subset of training data points resulting from the reduction of memory usage [13]. SVM is basically used in image processing, face detection, text categorization, image classification. So here we used two algorithms to train our model and the other two for better accuracy and efficiency. We had used overall 75% trained data and 25% test data in this dataset.

4 Results and Discussions

The sole motive of this work is to predict whether the person had the heart disease or not and the work is implemented of i5 Processor, 6 GB RAM, 1 TB external storage, and all libraries and programming were done in Python using Anaconda ML environment on Jupyter Notebook. As we noted the dataset is divided into tests, and we have applied two classifiers for the prediction of heart disease by classifying

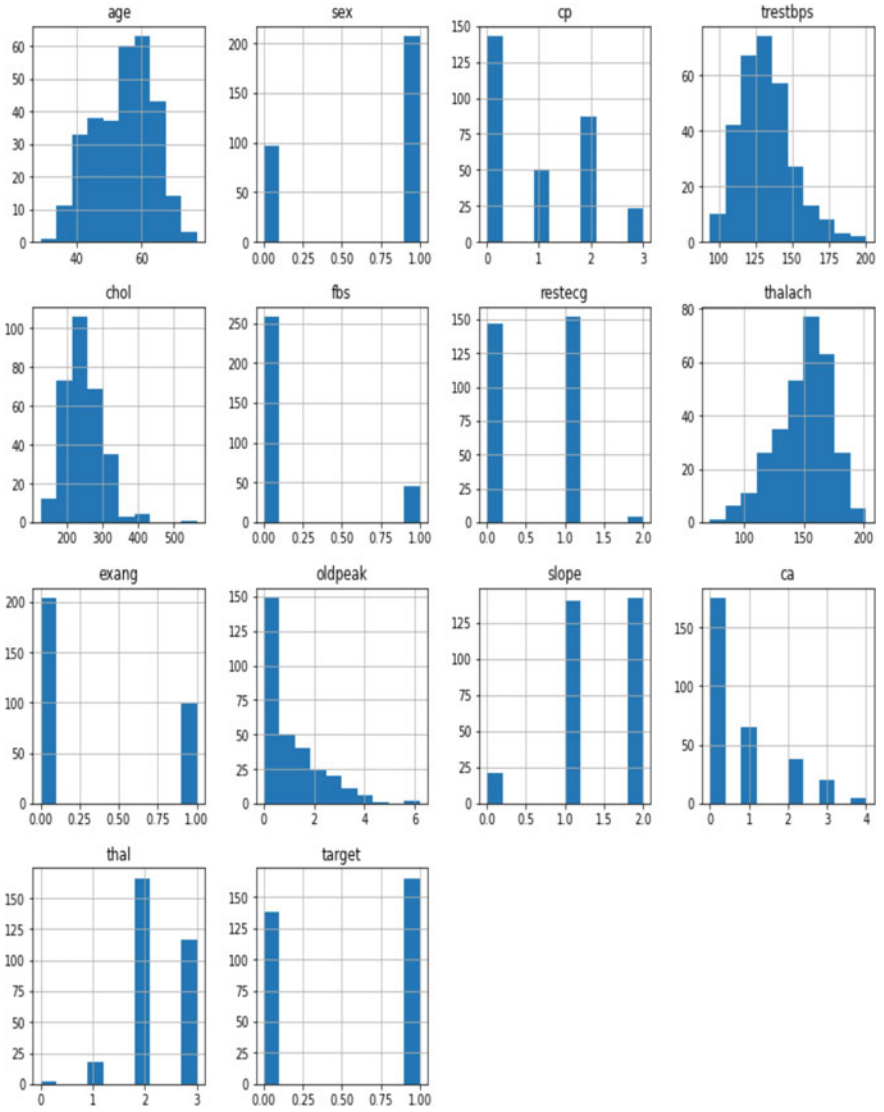


Fig. 2 Histogram of each data column

the different attributes and using two more algorithms for the higher rate of accuracy in our prediction. The two algorithms used for the classifier are decision tree and K-nearest neighbor. Fig. 2 and Table 1 give a depiction of the dataset used in this model having 14 attributes. The attributes in Table 2 are the basic body measures that vary rapidly and critically when there is any deflection in the functioning of the heart. Based on these body attributes or symptoms, we could predict the condition of the heart. A brief comparison of two characteristics of a normal human heart and

Table 1 Attributes with the description w.r.t Fig. 2 [17]

Characteristics	Illustration
Age	Individual's age in years
Sex	Gender status (1 = male, 0 = female)
Cp	Chest pain category
Trestbps	blood pressure of the person in resting condition (in mm hg)
Chol	Serum cholesterol (in mg/dl)
Fbs	Blood sugar level on empty stomach and gt; 120mg/dl (true:1, false:0)
Restecg	Electrocardiograph while resting condition
Thalach	Maximum value of heart rate recorded
Exang	Triggered angina by physical work (1 = yes, 0 = no)
Old peak	ST depression induced by physical work w.r.t to resting
Slope	The slope of the maximum exercise ST segment
Ca	Count of major vessels(0–3)colored with flourosopy
Thal	3 =stable, 6 = final defect , 7 = reversible defect
Target	0 or 1

Table 2 Accuracy collection of different combination model of algorithms of heart diseases prediction

Authors	Combinations of algorithms	Accuracy (%)
Shah et al. [17]	Naïve Bayes	88.15
	KNN	90.7
	Decision tree	80.2
	Random forest	86.84
Otoom et al. [19]	Naïve Bayes	84.5
	SVM	85.1
	Functional tree	84.5
Vemandasamy et al. [20]	Naïve Bayes	86.4
Chaurasia et al. [21]	J48	84.35
	Naïve Bayes	82.31
	Bagging	86.5
Proposed model	Decision tree	86.5
	KNN	92.6
	LGBM	81.5
	SVM	89.5

a heart suffering from any kind of defect in its operability is shown in Fig. 3, where the maximum heart rate of a heart having diseases is very high as compared to the maximum heart rate of a regular heart.

Figure 4 shows the prediction depicting as blues dots refer to no disease and red dot refers to having a disease. We had worked on improvising the accuracy of all four algorithms placing SVM and KNN at the highest place. The accuracy comparison of the combinations of the model used in this paper with the other recent research work is shown in Table 2. As summarized in Table 2, Author Shah et al. [17] in his paper focused on the accuracy of KNN making it highest among all his four algorithms used having an accuracy of 90.7%. When we see further, AF Otoom et al. [19] in 2015 took three algorithms which give results with similar accuracy making the average accuracy of the whole model 85%. Author Vemandasmy et al. [20] in 2015 worked on only one algorithm for prediction resulting in an accuracy of 86.4%, and author Chaurasia et al. [21] in his proposed model compared and improvised the accuracy of J48, naive Bayes, and bagging algorithms with the previous literature models. After that, the results from our proposed model are presented with the accuracy comparisons for all four algorithms having two out of four algorithms with higher accuracy. In our proposed model, we took four algorithms, giving SVM the highest

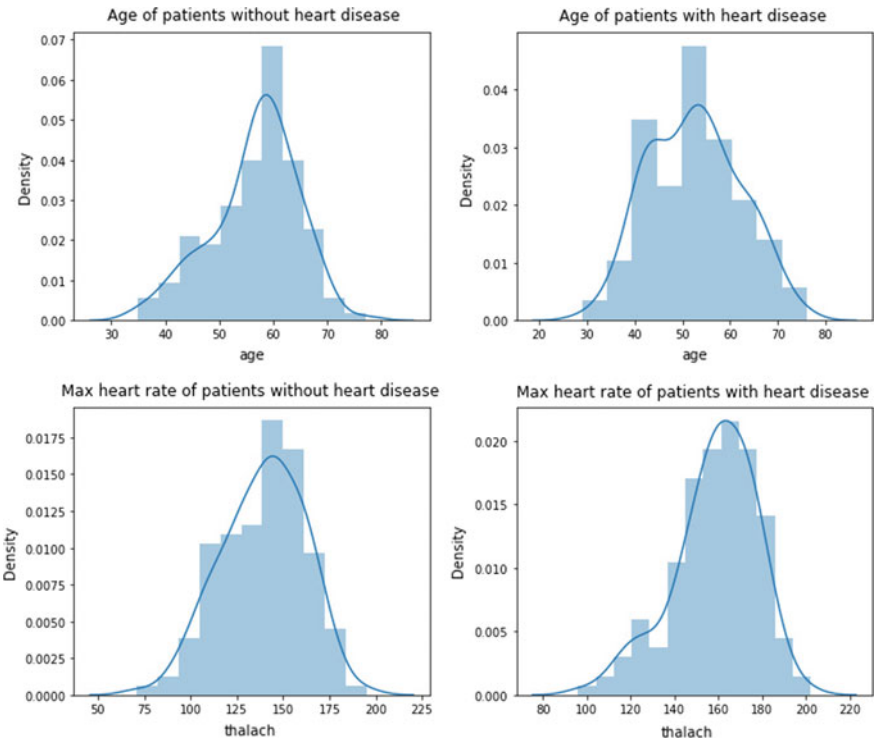


Fig. 3 Implementation of the algorithms

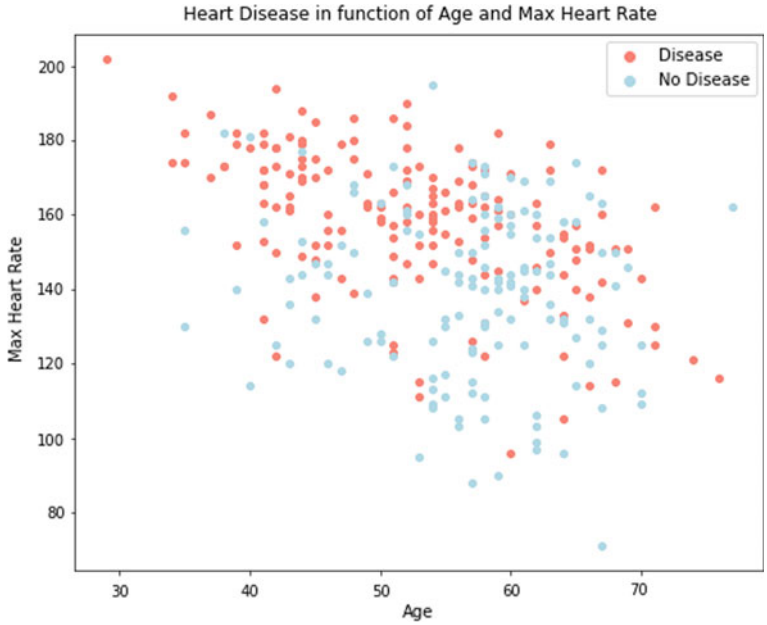


Fig. 4 Heart disease prediction b/w age and heart rate

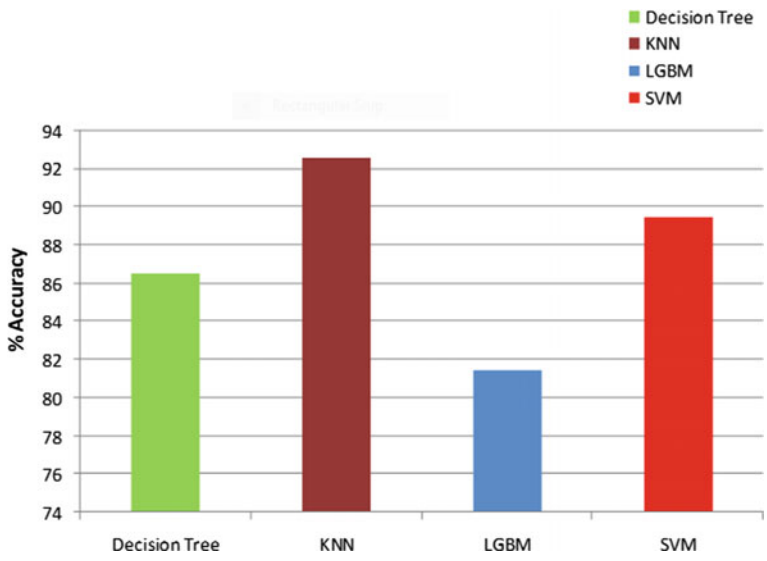


Fig. 5 Comparison of all four algorithms by applying the proposed model

accuracy of 89.5% followed by KNN with 92.6% in comparison to the other published work as shown in Table 2.

Figure 5 presents the accuracy of all the four algorithms considered in our proposed work. The decision tree gives a result with an accuracy of 86.5% having a tree-like structure for the data classification and followed by the KNN giving result with a 92.6% of accuracy rate, and the LGBM classifier gives an accuracy of 81.5%. Now, when we applied the SVM classifier and KNN differently to the datasets it gives us prediction but not as sure as when we applied both the SVM and KNN on the same datasets. And to add some more positivity, we use the other two on similar datasets. In a nutshell, when we applied all four algorithms, i.e., decision tree, KNN, LGBM classifier, and SVM collectively on the same datasets, these combinations of algorithms give us the best accuracy to predict heart diseases, with SVM giving the result with highest among all.

5 Conclusion

Various heart diseases or cardiovascular disease that disturb the stable functionality of the heart are the primary cause of death globally over the past few years. It is the need of time to take some responsible measures associated with the risk factor involved in heart diseases and to get a reliable approach for the early detection or prediction of heart condition so one can control or manage the condition before any casualty or harm. This research work summarized the process of data mining and machine learning.

References

1. Yeri V, Shubhangi DC (2020) IOT Based Real Time Health Monitoring System. Int Conf Res Comput Intell Commun Netw 980–984 (2020). https://doi.org/10.1109/ICIRCA48905.2020.9183194.
2. Bui AL et al (2011) Epidemiology and risk profile of heart failure. Nat Rev Cardiol 8:30–41. https://doi.org/10.1038/nrcardio.2010.165
3. Durairaj M, Ramasamy N (2016) A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate. Int J Control Theory Appl 9:255–260
4. Ghwanmeh S et al (2013) Innovative artificial neural networks-based decision support system for heart diseases diagnosis. J Intell Syst 5:176–183. https://doi.org/10.4236/JILSA.2013.53019
5. Al-Shayea QK (2011) Artificial neural networks in medical diagnosis. Int J Comput Appl 8:150–154
6. Gudadhe M et al (2010) Decision support system for heart disease based on support vector machine and artificial neural networks. IEEE Int Conf Comput Commun 741–745. https://doi.org/10.1109/ICCCT.2010.5640377.
7. Anooj PK (2012) Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules. J King Saud Univ Comput Inf Sci 24:27–40. https://doi.org/10.1016/j.jksuci.2011.09.002

8. Cardiovascular diseases (CVDs).: World Health Organization, Sep 2018. [Online]. Available: https://www.who.int/cardiovasculardiseases/en/
9. Yadav P, Agrawal R, Kashish K (2018) Protocol performance investigation using ad hoc WLAN for healthcare application. Pertanika J Sci Technol 26:1333–1354
10. Yadav P, Agrawal R, Kashish K (2018) Heterogeneous network access for seamless data transmission in remote healthcare. Int J Grid Distrib Comput 11:69–86
11. Mazumdar S, Choudhary R (2017) Innovative method for fetal health monitoring based on artificial neural network using cardiotocography measurements. IEEE Int Conf Res Comput Intell Commun Netw 265–68. https://doi.org/10.1109/ICRCICN.2017.8234518
12. Atallah R, Mousa AA (2019) Heart disease detection using machine learning majority voting ensemble method. Int Conf New Trends Comput Sci. https://doi.org/10.1109/ICTCS.2019.8923053
13. Gavhane A, Pandya I, Kokkula G, Devadkar K (2018) Prediction of heart disease using machine learning. Int Conf Electron Commun Aerosp Technol 1275–1278. https://doi.org/10.1109/ICECA.2018.8474922
14. Majumder JA, Saadany M (2019) A real-time cardiac monitoring using a multisensory smart IoT system. Comput Soft Appli Conf 281–287
15. Mohan SK, Thirumalai C, Srivastva G (2019) Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 7:81542–81554. https://doi.org/10.1109/ACCESS.2019.293707
16. Chauhan U, Chauhan V, Tiwary S, Kumar V (2019) Cardiac arrest prediction using machine learning algorithms. Int Conf Intell Comput Instrum Control Technol 886–890. https://doi.org/10.1109/ICICICT46008.2019.8993296
17. Shah D, Patel S, Bharti SK (2020) Heart disease prediction using machine learning techniques. SN Comput Sci 1:345–350
18. Peterek T, Gajdos P, Dohnalek P, Krohova J (2014) Human fetus health classification on cardiotocographic data using random forests. Intell Data Anal 2:189–195
19. Otoom AF, Abdallah EE, Kilani Y, Kefaye A, Ashour M (2015) Efective diagnosis and monitoring of heart disease. Int J Softw Eng Appl 9:143–56
20. Vembandasamy K, Sasipriya R, Deepa E (2015) Heart diseases detection using Naive Bayes algorithm. Int J Innov Sci Eng Technol 2:441–4
21. Chaurasia V, Pal S (2014) Data mining approach to detect heart diseases. Int J Adv Comput Sci Inf Technol 2:56–66. https://doi.org/10.1109/JEC-ECC.2012.6186978
22. Dwivedi AK (2018) Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Comput Appl 29:685–693
23. Kitjanukit S (2018) Attitude toward bioremediation–related technology and relation with company social responsibility. Evergr Jt J Nov Carbon Resour Sci Green Asia Strateg 6:86–92
24. Md Matiar Rahman A, Pal K, Uddin K, Saha TB (2018) Statistical analysis of optimized isotherm model for maxsorb III/ ethanol and silica gel/water pairs. Evergr Jt J Nov Carbon Resour Sci Green Asia Strateg 5:1–12
25. Bhatnagar P, Karuna S, Rajan S (2020) Predictive models and analysis of peak and flatten curve values of CoVID 19 case in India. Evergr Jt J Nov Carbon Resour Sci Green Asia Strateg 7:458–467

Internet of Thing (IoT): Routing Protocol Classification and Possible Attack on IOT Environment

Bhupendra Patel, Khushi Patel, and Anand Patel

Abstract We are living in contemporary era and use different technologies which facilitate humans to make their life comfortable and easier. Moreover, Internet of Things (IoT) helps us to integrate computing devices, different entities, animals or persons which will have the capability to transfer data over a network. There are different routing protocols, helpful to share the information from devices to devices, but as a security prospective there are many cyber-attacks as lots of vulnerable computing devices are getting linked with Internet every day. This paper gives brief of how different information exchange protocols works and how various attacks effect to Internet of Things (IoT).

Keywords Routing protocols · RPL · Security attacks · Internet of Things (IoT) · Network stack · DoS · DDoS

1 Introduction

Every day, technologies are lips and bounce, and it changes the human life significantly. For instant, as people were not able to communicate with other people within a minute in past, but nowadays, the help of technologies they can communicate

B. Patel
U & PU Patel Department of Computer Engineering, Chandubhai S. Patel Institute of Technology (CSPIT), Faculty of Technology and Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, Gujarat 388421, India
e-mail: bhupendrapatel.ce@charusat.ac.in

K. Patel (✉)
Department of Computer Engineering, Devang Patel Institute of Advance Technology and Research (DEPSTAR), Faculty of Technology and Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, Gujarat 388421, India
e-mail: khushipatel.ce@charusat.ac.in

A. Patel
Department of Information Technology, Faculty of Technology, Dharmsihn Desai University (DDU), Nadiad, Gujarat, India
e-mail: anand.patel73@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*,
Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_44

around the globe in short of time using mobile phone, email and different social media platform. Due to that our whole life is changed and now we are addicted to the technologies. Furthermore, the evolution of Internet of Things (IoT) made human life not only easier but also comfortable, and it is more widespread in this technology era for machine-to-machine communication. The wireless network, Bluetooth and RFID are the mediator to share our data on the Internet.

The term proposed by Kevin Ashton "Internet of Things" for the first times in the year of 1998. According to him "The Internet if Things has the would be changed the world, just as the Internet did. May be even more so" [1]. In 2001, the MIT Auto-ID center presented their idea on the same. Today, the IoT is being used in human life for daily activities at home as well as offices. For example, wearable devices are used for taking care of fitness of humans by monitoring heart beats, outdoor walking, etc. Moreover, smart health care is one of the areas which is using various IoT devices to offer services like telesurgery, diagnosis, etc., from remote places through IoT devices. As a result, healthcare services are available anywhere–anytime [2].

In IoT, there are number of wireless devices are connected with each other remotely and share their information over the Internet in form of IoT. To transfer data from source to destination secularly, we should have to use suitable routing protocol, which is helping us to send data secularly without any manipulations. Basically, there are three types of routing protocols used in IoT for exchanging the information: proactive, reactive and hybrid [3].

As a security point of view, many threats are making the service unavailable and share or manipulate the information while data transfer from source to destination. The main problem of is DOS and DDOS attack in IoT, also their cyber attackers share the malware over the network as well as some are connecting the malicious device in existing network and try to perform false operations [4].

The next section covers the introduction of IoT as well as its architecture. In next sections, different routing protocols use in IoT for transferring the data. Section 3 show the different kind of attacks possible in IoT. In last, Sect. 4 describes the significant effects of attacks on IoT network.

2 Introduction and Architecture of IoT

The Internet of Things (IoT) can be considered as physically connected devices (AC, lights, fans, etc.) or machines through the communication technology which can interact with one another and share the data and perform action as they received [1]. The IoT means, connected different things or objects which dealing with each other to provide smart services without any involvement of human. The Internet is establishing connection between computer nodes, while the IoT expands the connection between physical devices or objects as mentioned in former sentence. In all physical devices, there is installed/embedded the electronics circuits which will help to act as a particular node connected with Internet—IoT, also they can send and receive

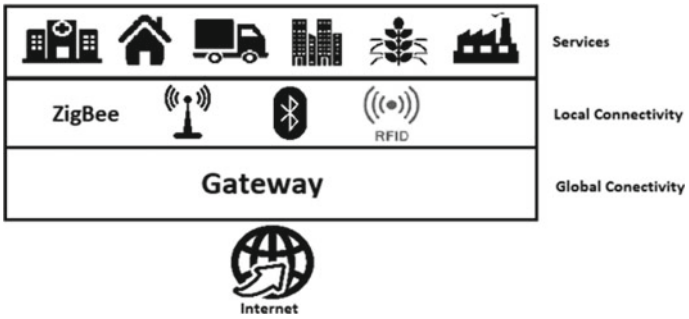


Fig. 1 Layered architecture of IoT [1]

the information and perform the particular task. Moreover, the layered model of IoT depicts how various devices connected with each other for communication.

Figure 1 depicts the various technologies which can be used at various layers to connect the devices with each other to setup IoT network. First layer is service, next is local connectivity layer, and last is global connectivity layer using gateway. For the global connection, we have an important thing that is Internet. The various web-enabled devices such as processors, sensors and communication hardware collect data from the environment and send it to cloud or edge connected devices for analysis or analyze locally. Majorly, in IoT, sensors and actuators are required to build IoT environment.

There are different types of sensors are available but a sensor is sensing the values in digital form and send or process the information in particular manner. Unlike sensors, an actuator convers electrical input into physical action. The example of different types of actuators is an electric motor, a hydraulic system or a pneumatic system.

Basic working process of IoT is as shown in Fig. 2. The various entities linked with other entities interact with others to offer the various kind of services [5]. Using sensors we can identify the physical objects. Sensors collect data in terms of different parameters such as detect smoke, temperature, light, humidity, pressure, etc., based on that particular sensor being used for sensing the data. The data sensed by the sensors will be communicated to the cloud or edge devices for performing various operations to identify the properties of data sensed.

Figure 3 compares the basic three layered network stacks of IoT and the five layered network stacks proposed by The International Telecommunication Union

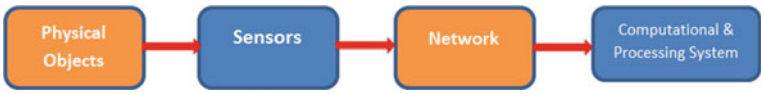


Fig. 2 Basic working process of IoT system [1]

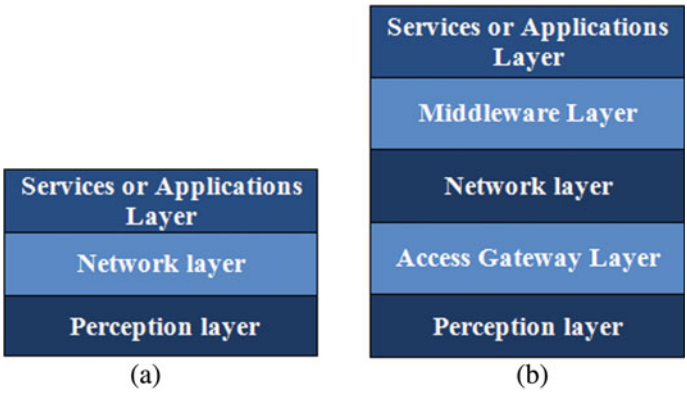


Fig. 3 IoT network stack. **a** Three layered network stack. **b** Five layered network stack [6]

(ITU) mentions [6, 7]. The layers of network stack proposed by Jia et al. and Domingo are 1. perception layer, 2. network layer and 3. service layer [7].

Now let's start with perception layer, it is collects information/data from the physical devices or sensors. In three-layer network stack, the role of network layer is to offer the bridge among perception layer and service/application layer. The network layer offers secure communication among perception layer to application layer while in the five-layer network stack, there is access gateway layer, which offers secure communication among perception layer and network layer. Moreover, in five-layer network stack, the role of middleware layer is to offer interface between networking devices and applications. The application layer offers services through which one can perform various operations on the information received from upper layers.

3 Routing Protocols Classification

In IoT, the main thing is how to transfer data from source to destination. In this section, brief about the different routing protocols which are repeatedly used in IoT applications for routing. The routing protocols, which are handle the transfer the data from source to destination and vice versa. Majorly four routing protocols are used most in IoT: RPL, CTP, LOADng and CORPL.

3.1 Routing Protocol for Low-Power and Lossy Networks (RPL)

RPL routing protocol frequently uses in IoT environment. It is a distance-vector routing protocol that is intended to work on upper of several link layer as well as

IEEE 802.15.4 PHY and MAC layers [8]. These MAC layer can remain embarrassed, hypothetically lossy, but not confined, less power wireless or power line communication (PLC) technologies. RPL main objectives are collection-based networks, where hosts or node frequently send gained measurements to a specific point. The main feature of RPL is offer a particular routing resolution for low power as well as lossy networks. This protocol was established to be highly adaptive to network environments and to provide substitute routes, when default paths or routes are unreachable. RPL is also provides a mechanism to distribute information over the dynamically designed network topology [8].

3.2 Collection Tree Protocol (CTP)

CTP is also known as distance-vector routing protocol that was established for a wireless sensor networks for routing. It works as an ancestor to RPL, and it was considered the de-facto routing standard for TinyOS. Also, it established like a tree-based topology or network with the source at the sink of particular network, CTP uses beacon signal mechanism to broadcast routing control messages in entire network. Furthermore, CTP depends on specific link-layer technical mechanism for establish of topology, CTP was famous for efficient energy consumption as well as high packet reception ratio (PRR) mechanism [8].

3.3 Lightweight On-Demand Ad Hoc Distance-Vector Routing Protocol-Next Generation (LOADng)

JOSÉ V. V. and his team defined that LOADng is simplifies the ad hoc on-demand distance vector (AODV) routing protocol [9] to handle the requests of LLNs. It has responsive feature and make a path or route among the nodes only, when data packets want to transferred. As a consequence, a particular source (SRC) node wants to share or send a data packet needs to commence a route discovery process and build a path till the destination (DEST) node [10]. The important thing for route discovery are the route request (RREQ), route reply (RREP) message and last is routing set, which can hold some route entries [11]. Whereas control messages are exchanged between the nodes to determine and build paths for transmit data packets, the formed routes information stored in the routing set and also manage control message coding as well as offer information to the forwarding of data packets.

3.4 CORPL

CORPL means cognitive radio environments. It will preserve the directed acyclic graph constructed approach of RPL and introduced novel modifications in cognitive radio environments in same time [12]. CORPL uses are adaptable progressing methodology that encompasses of two important steps, the first one is selection of a forwarder set for example, every node in the network selects manifold next hop nodes, and a coordination method to confirm that only the top receiver of each packet forwards it (exclusive forwarder selection). In CORPL, every node retains a forwarder set such that the forwarding node (next hop) is speculatively designated. The DAG creation process in CORPL strictly follows a related technique as process of RPL. After discovering process an unoccupied channel, the gateway node transfers a destination information object (DIO) message. The forwarder set is built in such a way that the forwarding nodes are within the transmission range of each node. Furthermore, throughout the DIO transmission, every node also reports some surplus information using the option field of the DIO message. Every node updates the next hop information via DIO message transmission. Depends on that the neighborhood information, each node vigorously prioritizes its neighbors in order to make the forwarder list [12].

4 Possible Attacks and Its Effects on IoT Network

Basically, there are different types of attack possible on different layer of IoT. On perception layer, there are many possible attacks such as jamming, tampering, exhaustion and collision [13]. These attacks have huge impact on the privacy and security of the user data as lots of data are stored on cloud platform [14]. Here, some of the possible attacks on IoT network are as follow [15–20].

4.1 Denial-Of-Service (DOS)

The denial-of-service (DOS) attack main aim is point out the particular host or network and try to effect the connection of network, generally it strive or makes service inaccessible to its anticipated users. DoS attack is sending high amount of information or data to the targeted user or network with high amount of traffic, or sending it information to stop the services [15]. It is one of the most common cyber-attack in security of network. Victims of DoS attack are generally the web servers of renowned organizations such as bank, stock market and media companies.

4.2 Distributed-Denial-Of-Service (DDOS)

The distributed-denial-of-service (DDOS) attack generally made at network layer and application layer. In particularly, DDOS attack there are many victims in the network. Then, attacker ready to avoid traffic of work-flow in network. The foremost goal of DDOS attack is try to crash the network or infrastructure and interrupt data overflow. DDOS attack majorly happened when device as well as system both is connected. Commonly, DDoS attack accomplishes its goals by preventing the normal work-flow from access essential destination.

4.3 Spoofed, Alter, Replay Routing Information

To implement this attack, attacker may generate false error message or routing loop to gain the access to the communication path between sender and receiver. There are various attack such as rank, version, local reparation, neighbor and DODAG information solicitation are the attacks fall under the category of alteration or spoofing attack. By altering various values, the mentioned attack can create powerful impact on targeting nodes in the network.

4.4 Sybil Attack and Clone ID Attack

In these types of attack, attacker replicates the IDs of legitimate nodes on multiple physical nodes to gain control over large part of network. Attacker creates fake IDs on physical nodes and spreads themselves over network to attract the traffic.

4.5 Sinkhole Attack

In this attack, attacker will attract the neighbor nodes toward particular route/node by broadcasting fake beneficial routing path. The impact of the attack can be more powerful by merging another attacking technique.

4.6 Selective Forwarding Attack

In order to implement this attack, attacker will introduce one node in the network as a faulty node which will either drop the packet or stop propagating the packet to the next node or only forwards selected packets to the next node which will create the

negative impact on neighbor node to select alternative route for the communication. This attack may lead to DoS attack as well. This attack has powerful impact on IoT social domain.

4.7 Hello Flooding Attack

In IoT network, in order to join the network, node needs to broadcast initial HELLO message with single power and data link-layer parameters to introduce the node toward neighbor nodes. The DIO message is used to communicate the DODAG network. By joining the network, attacker can convince the nodes that it is inside their network range and by broadcasting low cost routes, nodes can be convinced to send the message to the attacker. By combining other attacks, the impact of the attacks can be increases.

4.8 Wormhole Attack

By initiating this attack, attacker can disrupt the traffic flow and topology of the network by convincing the legitimate node to pass the messages through the short path which is going through the attacker. When wormhole is combined with another attack like sinkhole, it will create a huge harm to victim.

5 Conclusion

This paper concludes various routing protocols used in the IoT networks along with the security threats and their impact on network. In IoT network, number of various devices is connected with each other through various communication technologies which leads to interoperability and scalability issues as well but security and privacy of data are the issues on which many researchers are still working and trying to find the efficient solutions. In this paper, many threats are covered which can be detected by using various machine learning techniques to implement intrusion detection systems (IDS) but prevention solutions must be designed to prevent these types of attacks. Blockchain technology is another solution to manage the devices and their information to prevent cyber threats.

References

1. Patel K et al (2019) IoT: leading challenges, issues and explication using latest technologies. In: IEEE third international conference on electronics communication and aerospace technology
2. Azer MA, Abbokr A (2017) IoT ethics challenges and legal issues. In: 12th International conference on computer engineering and systems (ICCES)
3. Nasser N et al (2018) Routing in the Internet of Things. In: IEEE global communications conference
4. Maciel R et al (2021) Impact evaluation of DDoS attacks using IoT devices. In: IEEE international systems conference (SysCon)
5. Khan R, Khan SU, Zaheer R, Khan S (2012) Future internet: the Internet of Things architecture, possible applications and key challenges. In: 10th international conference on frontiers of information technology, Islamabad, India
6. Lai C-F, Vasilakos AV, Tsai C-W (2014) Future Internet of Things: open issues and challenges. Wirel Netw 20(8):2201–2217
7. Xu LD Senior Member, IEEE, He W, Li S (2014) Internet of Things in industries: a survey. IEEE Tran Ind Inform 10(4)
8. Prasad AH, Hema Bharat T (2017) Network routing protocols in IoT. Int J Adv Electron Comput Sci 4(4). ISSN 2393-2835.
9. Patel B et al (2014) AODV routing protocol performance in malicious environment. I advanced computing, networking and informatics, vol 2, pp 207–214
10. Sobral JVV et al (2019) Improving the performance of LOADng routing protocol in mobile IoT scenarios. IEEE Access 7
11. Sasidharan D, Jacob L (2018) A framework for the IPv6 based implementation of a reactive routing protocol in ns-3: case study using LOADng. Simul Model Pract Theory 82: 32–54
12. Salman T, Jain R. Networking protocols and standards for Internet of Things. In: Geng H (ed) Internet of Things and data analytics handbook, 1st edn. © 2017 John Wiley & Sons, Inc. Wiley
13. Aarika K et al (2020) Perception layer security in the internet of things. In: International workshop on artificial intelligence & Internet of Things (A2IoT) 9–12 Aug 2020, Leuven, Belgium
14. Vasoya S, Patel N, Ramoliya D, Patel K (2020) Potentials of machine learning for data analysis in IoT: A detailed survey. In: 2020 3rd international conference on intelligent sustainable systems (ICISS), pp 291–296. https://doi.org/10.1109/ICISS49785.2020.9316074
15. Liang L et al. A denial of service attack method for an IoT system. In: 2016 8th international conference on information technology in medicine and education (ITME)
16. Munshi A et al. DDOS attack on IOT devices. In: 2019 IEEE international conference on robotics, automation, artificial-intelligence and Internet-of-Things (RAAICON)
17. Lokulwar PP, Deshmukh HR (2017) Threat analysis and attacks modelling in routing towards IoT. In: 2017 International conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC), pp 721–726. https://doi.org/10.1109/I-SMAC.2017.8058273
18. Wallgren L, Raza S, Voigt T (2013) Routing attacks and countermeasures in the RPL-based internet of things. Int J Distrib Sens Netw 9(8):794326
19. Nawir M, Amir A, Yaakob N, Lynn OB (2016) Internet of Things (IoT): taxonomy of security attacks. In: 2016 3rd International conference on electronic design (ICED), pp 321–326
20. Pongle P, Chavan G (2015) A survey: attacks on RPL and 6LoWPAN in IoT. In: 2015 International conference on pervasive computing (ICPC), pp 1–6

Operating Frequency Prediction of Annular Ring Microstrip Antenna Using Extreme Learning Machine

Kanhaiya Sharma, Ganga Prasad Pandey, Dinesh Kumar Singh, Krishna Chaya Addagarrala, and Shailaja Salagrama

Abstract This article presents, Machine Learning-based proficient and quick, Extreme Learning Machine (ELM) procedure to predict the operating frequency of the Annular ring compact microstrip antenna (ARCMA)in the range of 0.95–5.76 GHz. As per UHF and SHF bands, 100 data samples are collected through simulation using CST by varying the proportion of inner radius to an outer radius of ARCMA's. Arbitrarily 90 samples were utilized to set up the ELM model, and the remaining 10 data samples were utilized to test the model. It was observed that the simulated and predicted operating frequencies of the ARCMA's are exceptionally near one another and can be used to design an annular ring microstrip antenna for UHF and SHF band applications.

Keywords Microstrip antenna · Extreme learning machine · Annular ring · Compact

1 Introduction

The market for telecommunication services has additionally seen massive upgrades in data speeds, from the GSM and CDMA to 3G, 4G and 5G networks. Due to these

K. Sharma (✉)
Computer Science Department, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India
e-mail: sharmakanhaiya@gmail.com

G. P. Pandey
Pandit Deendayal Energy University, Gandhinagar, Gujarat 387002, India

D. K. Singh
G.L. Bajaj Institute of Technology and Management, Gr. Noida, UP 201301, India

K. C. Addagarrala
General Motors. GM Milford Proving Grounds, Milford, MI 48380, USA

S. Salagrama
University of the Cumberlands, Kentucky 40769, USA

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_45

advancements, the need for compact microstrip antennas is significantly increased. As compared to other geometries of microstrip antenna, ARCMA is preferred to design compact microstrip antenna as it shows fundamental mode at a lower frequency. In the design process of a microstrip antenna, the operating frequency is one of the critical parameters. Using analytical techniques, the analysis of ARCMA was complicated and involved lots of mathematical calculations. Hence, it is a time-consuming process. Existing literature shows that antenna engineers have used different approaches to analyze annular ring microstrip antenna [1]. Recently, an analysis of annular ring antenna is carried out using capacitive coupling matching network [2]. The investigation (demonstrating) of ARCMA is not simple for the most part because of the presence of fringing field and boundary conditions of the shape as compared with rectangular microstrip antenna. In such a manner, numerous endeavors have been committed to creating precise investigating methods, as straightforward as conceivable. Many researchers have investigated ARCMA using regression-based adaptive machine learning approaches like Artificial Neural Network (ANN) [3], Support Vector Machine Regression (SVMR) [4], Gaussian Process Regression (GPR) [5]. ELM is a one hidden layer feedforward Neural network and part of Artificial intelligence(AI). The EML has better generalization performance and is faster than the gradient-based learning algorithms. For a large data set, ELM performs better than Random Forest and SVM [6]. Due to its wonderful performance of EML in terms of efficiency and accuracy, it is applied to solve many real-world problems in all areas of engineering. By using the AI approach, the design process of ARCMA will become simple, adaptive, fast, and accurate. In the present investigation, the authors investigated the operating frequency of ARCMA using the ELM approach.

2 Extreme Learning Machines

ELM is a single hidden layer-based feedforward neural network with the ability to converge much faster than traditional neural networks, and it is three-layer architecture, i.e., the first layer is input,second layer is single hidden layer and the last layer is output. The main focus of ELM is to tune the hidden layer. The generalized output function of ELM is evaluated as

$$f_H(x) = \sum_{i=1}^{H} \beta_i k_i (w_i * x_j + b_i),\ j = 1, \ldots, N \tag{1}$$

where H is number of hidden unit, N is the number of training samples, w is weight vector between input and hidden layer, k is activation function, b is a bias vector and x is input vector. The proposed ELM architecture is shown in Fig. 1. In Fig. 1, four critical parameters of ARCMA, i.e., inner radius (a_i), outer radius (a_o), height h and dielectric constant ϵ_r are given as input and the operating frequency (f_r) is predicted as output.

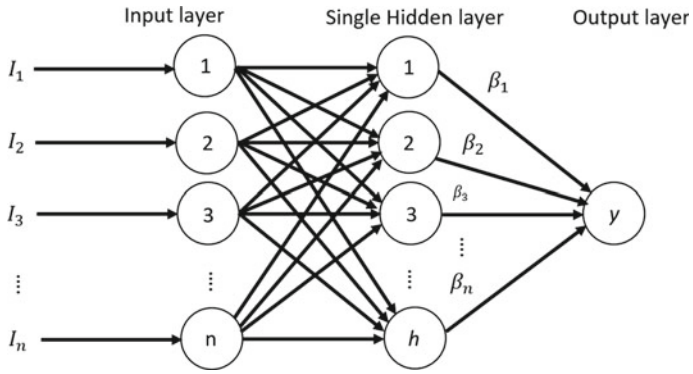


Fig. 1 Proposed ELM model

3 Antenna Design and Simulation

An annular ring patch antenna is designed by loading a circular ring at the center of the circular patch. The size of an annular ring patch antenna is comparatively smaller than a circular microstrip antenna operating at the same frequency [7]. In the Annular ring patch antenna, the typical path length traveled by the current is significantly longer than the corresponding circular patch [7], which leads to the overall size reduction of the antenna. The proposed design structure of ARCMA with front and side view is depicted in Fig. 2. From the figure, it is clear that the ARCMA consists of a circular patch of radius a_0 with one ring slot at the center of radius a_i. The calculation of various parameters of ARCMA and its analysis is studied in [8].

$$f_{nm} = \frac{K_{nm}c}{2\pi\sqrt{e_\mathrm{r}}} \tag{2}$$

where c is the velocity of light, K_{nm} is the roots of characteristic equation give as

$$J_n'(ka_\mathrm{o})Y_n'(ka_\mathrm{i}) - J_n'(ka_\mathrm{i})Y_n'(ka_\mathrm{o}) = 0 \tag{3}$$

where $J_n(x)$ and $Y_n(x)$ are the Bassel function of the first and second order, respectively, and the prime denotes derivatives with respect to k_ρ. In the presence of fringing field, the effective permittivity is give as

$$\epsilon_\mathrm{e} = \frac{1}{2}(\epsilon_\mathrm{e}+1) + \frac{1}{2}(\epsilon_\mathrm{e}-1)\left(1+\frac{10h}{W}\right)^{\frac{-1}{2}} \tag{4}$$

where W is the width of ring and is defined as $W = a_\mathrm{o} - a_\mathrm{i}$; for more accurate result of annular ring $\frac{a_\mathrm{o}-a_\mathrm{i}}{a_\mathrm{o}+a_\mathrm{i}} < 0.35$ and $o = 2 \times a_\mathrm{i}$, should be satisfied. In the presence of

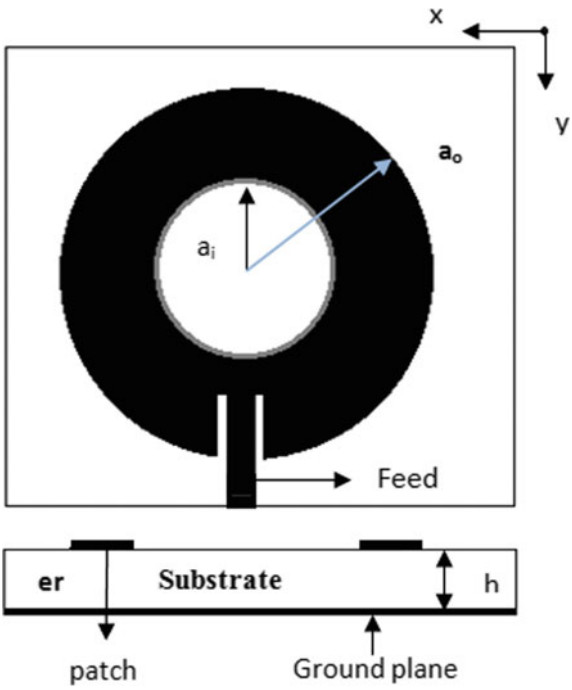


Fig. 2 Front and side view of annular ring microstrip antenna

the fringing field along the curved edge of the ring, the outer and inner radii defined as

$$a'_{\mathrm{o}} = a_{\mathrm{o}} + \left(\frac{W_{\mathrm{e}}(f) - W}{2}\right) \tag{5}$$

$$a'_{\mathrm{i}} = a_{\mathrm{i}} - \left(\frac{W_{\mathrm{e}}(f) - W}{2}\right) \tag{6}$$

where W_{e} is a effective width of the ring and defined as

$$W_{\mathrm{e}}(f) = W + \left(\frac{W_{\mathrm{e}}(0) - W}{1 + (\frac{f}{f_p})^2}\right) \tag{7}$$

where $W_{\mathrm{e}}(0)$ is give as $W_{\mathrm{e}}(0) = \left(\frac{120\pi h}{z_{\mathrm{o}}\sqrt{\epsilon_{\mathrm{e}}}}\right)$ and $f_p = \left(\frac{z_{\mathrm{o}}}{2\mu h}\right)$, here z_{o} is the quasi static characteristic impedance of the microstrip line with width W, μ is the permeability and h is the thickness of the substrate.

By using Eqs. 2–5, the various parameters of ARCMA is calculated to design an antenna. The initial values of $a_{\mathrm{o}} = 25$ mm and $a_{\mathrm{i}} = 1$ mm fixed with $\epsilon_{\mathrm{r}} = 4.3$ and $\tan\delta = 0.025$, feed length = 20.8 mm, feed width = 5 mm, $L_{\mathrm{g}} = W_{\mathrm{g}} = 59.6$ mm and $h = 1.6$ mm to resonate at 1.7 GHz. 100 Data samples were collected through

Table 1 Training dataset

ARCMA	a_o (mm)	a_i (mm)	h (mm)	ϵ_r	f_r(GHz)	ARCMA	a_o (mm)	a_i (mm)	h (mm)	ϵ_r	f_r (GHz)
1	10	0.5	1.57	2.33	5.7628	51	25	8.5	2.5	4.5	1.0689
2	10	1	1.57	2.33	5.754	52	25	9	2.5	4.5	1.06
3	10	1.5	1.57	2.33	5.7356	53	25	9.5	2.5	4.5	1.0493
4	10	2	1.57	2.33	5.7059	54	25	10	2.5	4.5	1.0386
5	10	2.5	1.57	2.33	5.6467	55	10	0.5	0.8	3.5	5.1592
6	15	3.5	1.57	2.33	3.7644	56	10	1.5	0.8	3.5	5.124
7	15	4	1.57	2.33	3.7129	57	10	2	0.8	3.5	5.0752
8	15	4.5	1.57	2.33	3.6484	58	10	2.5	0.8	3.5	5.0102
9	15	5	1.57	2.33	3.584	59	15	3	0.8	3.5	3.5949
10	20	5.5	1.57	2.33	2.7022	60	15	3.5	0.8	3.5	3.5457
11	20	6.5	1.57	2.33	2.6282	61	15	4	0.8	3.5	3.4842
12	20	7	1.57	2.33	2.5896	62	15	4.5	0.8	3.5	3.4166
13	20	7.5	1.57	2.33	2.5496	63	15	5	0.8	3.5	3.3612
14	25	8	1.57	2.33	2.0933	64	20	5.5	0.8	3.5	2.5409
15	25	8.5	1.57	2.33	2.0674	65	20	6	0.8	3.5	2.4984
16	25	9	1.57	2.33	2.0449	66	20	6.5	0.8	3.5	2.4635
17	25	9.5	1.57	2.33	2.0189	67	20	7	0.8	3.5	2.4286
18	25	10	1.57	2.33	1.9947	68	20	7.5	0.8	3.5	2.3937
19	10	0.5	1.6	4.3	4.5089	69	25	8	0.8	3.5	1.9221
20	10	1	1.6	4.3	4.503	70	25	9	0.8	3.5	1.8713
21	10	2	1.6	4.3	4.4639	71	25	9.5	0.8	3.5	1.8472
22	10	2.5	1.6	4.3	4.4286	72	25	10	0.8	3.5	1.8218
23	15	3	1.6	4.3	2.8645	73	10	0.5	3.2	12.94	2.4542
24	15	3.5	1.6	4.3	2.8335	74	10	1	3.2	12.94	2.454
25	15	4	1.6	4.3	2.7964	75	10	1.5	3.2	12.94	2.4509
26	15	4.5	1.6	4.3	2.7469	76	10	2	3.2	12.94	2.4452
27	15	5	1.6	4.3	2.7158	77	10	2.5	3.2	12.94	2.4383
28	20	5.5	1.6	4.3	2.0394	78	15	3	3.2	12.94	1.5713
29	20	6	1.6	4.3	2.012	79	15	3.5	3.2	12.94	1.5615
30	20	6.5	1.6	4.3	1.9845	80	15	4	3.2	12.94	1.5525
31	20	7.5	1.6	4.3	1.9322	81	15	4.5	3.2	12.94	1.5372
32	25	8	1.6	4.3	1.5444	82	15	5	3.2	12.94	1.5228
33	25	8.5	1.6	4.3	1.5239	83	20	6	3.2	12.94	1.2043
34	25	9	1.6	4.3	1.5067	84	20	6.5	3.2	12.94	1.1969
35	25	9.5	1.6	4.3	1.4896	85	20	7	3.2	12.94	1.1857
36	25	10	1.6	4.3	1.4702	86	20	7.5	3.2	12.94	1.1759
37	10	0.5	2.5	4.5	2.9203	87	25	8	3.2	12.94	0.9805
38	10	1	2.5	4.5	2.9169	88	25	8.5	3.2	12.94	0.9744
39	10	1.5	2.5	4.5	2.9124	89	25	9.5	3.2	12.94	0.9602
40	10	2	2.5	4.5	2.9039	90	25	10	3.2	12.94	0.95336
41	10	2.5	2.5	4.5	2.8932	91	15	3	1.57	2.33	3.8031
42	15	3	2.5	4.5	2.0193	92	20	6	1.57	2.33	2.6652
43	15	4	2.5	4.5	2.0104	93	10	1.5	1.6	4.3	4.4932

(continued)

Table 1 (continued)

ARCMA	a_o (mm)	a_i (mm)	h (mm)	ϵ_r	f_r (GHz)	ARCMA	a_o (mm)	a_i (mm)	h (mm)	ϵ_r	f_r (GHz)
44	15	4.5	2.5	4.5	1.9953	94	20	7	1.6	4.3	1.9609
45	15	5	2.5	4.5	1.9882	95	15	3.5	2.5	4.5	2.0157
46	20	5.5	2.5	4.5	1.3477	96	20	7.5	2.5	4.5	1.2919
47	20	6	2.5	4.5	1.336	97	10	1	0.8	3.5	5.1511
48	20	6.5	2.5	4.5	1.32	98	25	8.5	0.8	3.5	1.8967
49	20	7	2.5	4.5	1.3062	99	20	5.5	3.2	12.94	1.2144
50	25	8	2.5	4.5	1.0796	100	25	9	3.2	12.94	0.96914

Table 2 Comparison

ARCMA	Simulated	Predicted	Percentage error (%)
91	3.8031	3.8342	0.817753938
92	2.6652	2.6647	0.018760318
93	4.4932	4.4889	0.095700169
94	1.9609	1.96001	0.045387322
95	2.0157	2.0456	1.483355658
96	1.8967	1.8997	0.158169452
97	1.2919	1.2893	0.201253967
98	5.1511	5.1389	0.236842616
99	1.2144	1.2098	0.378787879
100	0.96914	0.9601	0.932785769
| min % error |			0.018760318
| max % error |			1.483355658
| Average % error |			0.436879709

simulation with CST by changing material, the thickness of the substrate, inner and outer radius of the ARCMA in accordance with UHF and SHF band requirements. The details of a data set are tabulated in Table 1.

4 Data Generation and Training

For training and testing, a total of 100 ARCMA's simulated using CST^{TM} 2018 software. In the process of simulation, four ARCMA parameters inner radius a_i, outer radius a_o,height h and dielectric constant ϵ_r are varied at optimized 50ω impedance. The details of the data set are tabulated in the Table 1, in the table, the test data sets are ARCMA 91–100. The division of the dataset for training and testing is random and is 90 and 10%. The proposed model takes less than 1.63 s to train with 98.51% accuracy.

4.1 Result and Discussion

The developed model successfully predicted for test data set with 98.51% accuracy,| Average % error | is 0.436%, | min % error | is 0.0187% and | max % error | is 1.4834%, which is less then 1.5%. Simulated and predicted resonant frequencies are compared, and its result is given in Table 2. From the Table, it is clear that the predicted and simulated operating frequencies are exceptionally near to one another.

4.2 Conclusion

Calculation of the operating frequency for ARCMA by either simulation or measurement analytical methods is very complicated and needs lots of mathematical calculation. By applying the Artificial Intelligence-based Extreme Learning Machine approach, we can speed up the design process of ARCMA by quickly and accurately calculating the resonant frequency. The proposed ELM model was trained in less than 1.63 seconds with 98.5% accuracy and utilized for predicting the resonant frequencies of the ARCMA. The predicted outcome is very close to the simulated one, which is quite good and acceptable by the industries. Hence, the proposed method can be quickly and accurately calculate the operating frequency of ARCMA design an annular ring microstrip antenna for future wireless communication.

References

1. Andriesei C (2017), Annular-ring microstrip patch antennas, vol 63, no 67 Buletinul Institutului Politehnic Din Iasi Numarul 4
2. Li Y, Sun S, Jiang L, Tian B (2021) Modeling and analysis of microstrip annular ring antenna with capacitive coupling matching network. Int J RF Microw Comput-Aided Eng 31(2):e22507. Available https://onlinelibrary.wiley.com/doi/abs/10.1002/mmce.22507
3. Khan I, Tian Y-B, Inamullah H, Vllah SU, Rahman, Kamal MM (2018) Design annular ring microstrip antenna based on artificial neural network. In: 2nd IEEE advanced information management, communicates. Electronic and automation control conference (IMCEC), pp 2033–2037
4. Kayabasi A, Akdagli A (2015) A novel method of support vector machine to compute the resonant frequency of annular ring compact microstrip antennas. Cogent Eng 2(1):981944
5. Sharma K, Pandey GP (2019) Predicting resonant frequency of annular ring compact microstrip antenna using various machine learning techniques. In: 2019 IEEE 16th India council international conference (INDICON), pp 1–4
6. Ahmad I, Basheri M, Iqbal MJ, Rahim A (2018) Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. IEEE Access 6:33789–33795
7. Chew W (1982) A broad-band annular-ring microstrip antenna. IEEE Trans Antennas Propag 30(5):918–922
8. Dahele JS, Lee KF (1985) Theory and experiment on the annular-ring microstrip antenna.Annales des Télécommunications 40:508–515

Challenges and Future Perspectives of Low-Power VLSI Circuits: A Study

Paramjeet Chauhan and Saptarshi Gupta

Abstract Low-power technologies, which have taken over the electronics sector, are being studied in this scientific literature. Power dissipation is an important design parameter in VLSI circuits because it predicts the performance of battery-operated devices, which is important in biomedical and communication applications. It gets more difficult to construct high-performance, low-power systems on a chip, as chip size reduces and device density and complexity rise. Furthermore, due to increased design complexity below the 100 nm node, total power management on a device is becoming a severe concern. Leakage current is also important in the power management of low-power VLSI devices. Because leakage and dynamic power consumption account for a large portion of overall power consumption in micro and nanotechnologies, they are becoming more relevant design factors. In order to increase the battery life of portable devices, VLSI circuit design focuses on reducing leakage and dynamic power. The numerous approaches, tactics and power management techniques that may be employed to construct low-power circuit-based systems are addressed in this scientific literature review.

Keywords Low-power · Quantum computing · CMOS · VLSI

1 Introduction

The microelectronics industry has flourished since the invention of the transistor, which provided the foundation for low-power devices. Integrated circuits (IC) have improved circuit performance while simultaneously reducing their size. Power and energy measurement are crucial in every discipline of engineering. Governing bodies and regulators have intensified their efforts to generate new energy consumption

P. Chauhan (✉) · S. Gupta
Department of Electronics and Communication Engineering, SRM Institute of Science & Technology, Delhi NCR Campus, Ghaziabad, Uttar Pradesh, India
e-mail: chauhan.paramjeet@gmail.com

S. Gupta
e-mail: ece.saptarshi@gmail.com

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_46

standards for various types of equipment as a consequence of their increased efforts to create new energy consumption criteria for various types of equipment [1]. Energy is required for lighting, cooking, propulsion and a number of other industrial and domestic purposes. Millions of dollars might be lost in a single hour of complete darkness. Power calculations should be used by engineers who wish to maximize output while minimizing losses. As a result, measurement technologies must be long-lasting, trustworthy and provide enough hardware, software and services to help manufacturers get their products from concept to production lines faster [2].

Electric power is the product of the rate of passage of electric charge (current) and the electric potential (voltage) across the load, to refresh your recollection. In electrical systems, alternating current, direct current or both are employed. Because, it is easily accessible, alternating current is widely used, while electrometallurgical activities and electronic devices rely only on direct current. The type of system and the quantity of power used can help you choose the right power measuring instrument. This is necessary in order to establish the technological idea used in the instrument's design, as well as the inaccuracy of the measuring procedure. The most often utilized electronic gadgets nowadays are portable electronics such as cell phones.

The majority of these gadgets are powered by low-capacity batteries. Every day, these devices are exposed to new and demanding applications that necessitate the use of powerful processors. One of the key concerns of these gadgets' customers is the ability to modify the battery life. As the complexity of applications has risen, so has the relevance of this issue. As a result, the researchers must look for solutions to the issue. Improved battery capacity is one option, but this raises costs, takes up more space, and makes portable electronic devices more cumbersome [1]. As a result, it is necessary to find strategies to reduce power consumption without losing efficiency. As a result, when evaluating this trade-off, power-delay products (PDP, energy) is an appropriate parameter to employ [2]. In addition, powerful processors must be produced and a balance must be struck between the several methods for reducing processor power consumption that have been developed, the most significant of which are supply voltage scaling, frequency scaling, and transistor downscaling [3–5]. The strategies described have both benefits and drawbacks. When the operating voltage of a circuit is reduced, for example, its performance suffers [3]. Current leakage exponential increase, gate control decrease, short channel effects, and greater manufacturing costs have all been connected to downscaling of metal oxide semiconductor field effect transistor (MOSFET) transistors [3, 6–8].

One potential option to overcome these difficulties is to modify current technology and incorporate new technologies such as carbon nanotube (CNT) FET and quantum cellular automata (QCA) [7–10]. Because MOSFET transistors and CMOS technology are the most widely used IC technologies and cannot be quickly replaced, adequate solutions to the aforementioned issues must be found. Another option is to use area-efficient designs to save money on fabrication and power consumption [11]. As a result, the power-area product (PAP) may now be recognized as a legitimate circuit analysis metric. According to the foregoing, using a current approximation computing design strategy for error-tolerant applications that require fast and powerful processors might be a viable option [1, 4, 11]. Approximate computing,

for example, can be used to process photos and videos due to humans' limited visual perceptive abilities [13, 14].

Digital signal processors (DSP) currently account for the majority of the market [13]. Because a major percentage of digital signal processing may be done in an approximate manner with a low proportion of mistakes, this design strategy can be a viable and practical choice in these sorts of processors. Low-power gadgets, which are already sweeping the electronics industry, are actively being researched by researchers. Circuit complexity and speed grow as VLSI technology advances, resulting in significant power consumption. There are two competing constraints in VLSI design: compact size and great performance. The actions of the integrated circuit (IC) designer have contributed in the removal of these constraints.

As a result of the many design constraints, power efficiency has become more crucial. The most advanced battery-powered devices can perform tasks that need a large number of calculations. The most striking feature of Moore's Law is that it has evolved into a universal predictor of semiconductor industry growth. Power dissipation is an important design parameter in VLSI circuits because it is used to evaluate the performance of battery-operated devices, which is important in biomedical and communication applications. As chip size reduces and chip density and complexity rise, creating greater performance low-power consumption systems on a semiconductor becomes increasingly difficult. Because leakage and dynamic power consumption account for a considerable amount of overall power consumption in submicron technologies, these design factors are becoming more important. Leakage and dynamic power reduction are increasingly important goals of VLSI circuit design in order to extend the battery life of portable devices.

The increased usage of electronic equipment in recent years, there has been a greater requirement to verify that any equipment connected to the ac mains has a line current harmonic content that is within permissible limits. Input phase currents are routinely shaped such that they are sinusoidal and in phase with the input phase voltages using power factor correction (PFC) circuits. Sinusoidal line currents are available in many different forms and sizes. Power management is a solution for suppressing speed variations and lowering the power consumption of power-aware systems by regulating supply voltages in a static and dynamic manner. As previously stated, power management may effectively reduce sub-threshold currents by controlling supply voltage, and numerous suggestions have been proposed in this regard thus far. The subsections that follow present a quick summary of power management challenges that DRAM designers have run across, as well as a variety of logic designers' suggestions for power management solutions, particularly for SoCs.

As a result, the IC's power unit grows in size. The need for low-power gadgets is increasing as the number of battery-powered sophisticated working life-saving equipment such as peacemakers and other implanted medical tools grows. Researchers are putting forth more effort than ever before to develop low-power components and design methodologies. In high-power applications, on the other hand, each increase in temperature doubles the rate of silicon failure. With the development of deep sub-micron nodes and nanotechnologies, reducing power consumption has become a logical step.

2 Need for Low-Power Design

Power efficiency is one of the most important concerns for 4G-5G operators. Customers are constantly on the lookout for mobile gadgets that are smaller, trimmer and more appealing. Recent advances in ultra-low-power circuit design have lowered power consumption in sleep mode to picowatts, nanowatts in active mode and ultra-low duty cycle functioning [1]. These developments enable for the use of ambient vibrations rather than batteries as a power source [2–4]. As a result, there is a strong demand to create environmental energy harvesting devices [5–8]. One of the goals of this study is to find a way to fix low currents while yet retaining high efficiency. The objective is to increase the efficiency of both energy conversion and power harvesting [9, 10]. Because CMOS-based devices have lately become increasingly resistant to scaling, Moore's law's continuous scaling down of their size over the preceding several decades has resulted in a slew of new and difficult issues [11–13]. One of the most major challenges faced by transistor-based circuits is power consumption from leakage current due to greater threshold voltage and lower supply voltage [14].

High degrees of silicon integration are required in today's processes, yet sophisticated processes require more power indulgence by definition. As a result, when it comes to low-power devices, design is crucial. In low-power devices, power dissipation has become an essential component. Because of the increasing use of portable electronic devices and the examination of microelectronic technology, VLSI circuit designs have emerged. VLSI technology's rising circuit complexity and speed results in a large rise in power consumption. Transition activity causes energy dissipation in low-power CMOS VLSI circuits by charging and discharging internal node capacitances, which is one of the key variables that determines dynamic power dissipation. Design techniques must be optimized at all stages to save power, area, and improve speed.

When it comes to mobility, the biggest stumbling obstacle appears to be a bigger dissipation. As a result, the entire system's power consumption must be assessed. Even though it is critical to maximize run time while minimizing battery weight, life, and space constraints, reducing overall power consumption in such devices is also critical. As a result, while building a system on chip for portable devices, low-power design is the most important issue to consider. Mobile users typically demand more features and a longer battery life in exchange for a lower price. Longer talk and standby times are a crucial feature for mobile phones, according to over 70% of consumers.

The two most essential parameters in power dissipation are performance and area. Down to 100 nm, power utilization and on-chip power management are key challenges due of the clear better quality. In terms of power optimization, lower package prices and longer battery life are critical. Leakage current is crucial in power management since reduced power consumption is a fundamental issue in high-performance digital and microprocessor systems. The entire power dissipation of integrated circuits is limited by leakage current. The sole conditions for a good semiconductor are low-power consumption and power dissipation calculations.

3 Sources of Power Dissipation

Power dissipation is the rate at which energy is squandered over time in an electrical or electronic system, and it is separated into two types: peak power and average power. Peak power refers to the device's maximum instantaneous power during a certain amount of time, and it has an influence on its dependability, leading in a glitch and, as a result, an increased chance of system failure. The average or maximum instantaneous power is the subject of power minimization schemes. In CMOS circuits, power dissipation is caused by static, short-circuit, and dynamic power. In CMOS circuits, static, short-circuit, and dynamic power all contribute to the overall power dissipation.

The reverse biased current generates the power consumed by a device due to transistor leakage, also known as static power. The sources of reverse biased current are sub-threshold leakage, gate oxide tunnelling leakage, diffusion area, and substrate. The algebraic total of all biased current and leakage current is the static power. Short-circuit power dissipation is influenced by device supply voltage, transistor threshold voltage, and input rise and fall times. A slow rise time input contributes about 20% of the short-circuit power loss in a COMS inverter [2]. Static power dissipation is reduced more efficiently when the input and gate output transitions are balanced, but gates with quick switching rates utilize more short-circuit power. Due to oxide gates and metal footprints, dynamic power is the biggest amount of power required by a device during the parasitic capacitance charging and discharging process; it is the maximum amount of power consumed when compared to overall power consumption [4]. As a result, dynamic power dissipation is more important than static or leaky power reduction. A general description of a CMOS logic gate for computing switching power includes drain capacitance, connection capacitance, and input capacitance.

4 Methodologies of Low-Power Consumption

Circuit speed has traditionally been utilized as a performance metric by VLSI designers. In fact, in numerous portable applications, power considerations have been the most important design requirement. The major goal of these programmes was to get as much battery life as possible while using as little power as possible. Low-power design is also necessary in high-end systems with high integration density to minimize power consumption and enhance operating speed. To reduce power dissipation in digital systems, low-power techniques should be used throughout the design process, from system to process level. It is critical to understand how electricity is dispersed.

4.1 Process Technology Reduces Power Consumption

- Reducing the device's supply voltage is one of the most effective techniques to minimize power usage. The drawback of this method is that when VDD approaches the threshold voltage, delay may increase significantly. Devices must be sized appropriately to overcome this problem. Some of the advantages of scaling are as follows:
- Enhance the device's capabilities
- Reduce junction and geometric capacitances.
- Improvements in interconnect technology
- Integration density is high.

4.2 Circuit/logic Design Reduces Power Consumption

- More static circuits are used than dynamic circuits.
- Optimized algorithm reduces switching activity.
- Time and bus loading should be optimized.
- Smart circuit methods reduce the number of devices in a circuit.
- Custom design has the potential to increase the power.
- VDD is reduced in non-critical circuits, and appropriate transistor size is achieved.
- Multi-VT circuits are used.
- Sequential circuits re-encoding.

4.3 Using an Architectural Model to Reduce Power Consumption

- Techniques for managing power, such as turning off unneeded blocks
- Pipelining, parallelism, and other architectures
- By allowing specific blocks, memory partitioning may be achieved.
- Reduction in the number of buses on the road across the world
- Minimization of instruction sets for simpler decoding and execution.

4.4 Quantum Computing

One of the developing technologies that can be utilized to replace CMOS technology is quantum-dot cellular technology (QC). Because of its incredibly low-power dissipation, high working frequency, and tiny size, it has gotten a lot of attention in recent years. QC was developed as a result of a search for new technologies, and it offers appealing features such as low energy consumption and high cell density

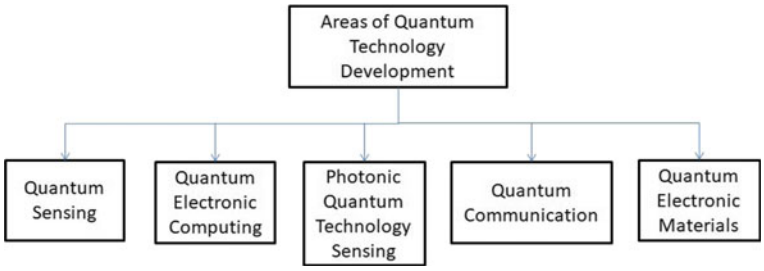


Fig. 1 Areas of quantum technology development

[15]. Existing CMOS logic consumes more energy and takes up more space than QC systems [16, 17]. Quantum cells, which are made up of four quantum dots and two electrons, are used to build circuits in QC. Indium arsenide (InAs) and gallium arsenide (GaAs) are used to make quantum dots, which are nanoscale semiconductor structures (GaAs). In QC implementation, information is transferred through the transmission of a polarization state rather than current. Because of its immediate use in quantum computing, this novel technique sparked a lot of curiosity among academics. QC logic circuits such as multipliers, adders, reversible ALU, divider, decoder, and memory circuits have been the subject of various studies [18–25]. Compared to CMOS components, several of these designs provide benefits such as quicker processing and lower size.

Mass manufacture of ultra-small QC technology, on the other hand, is extremely challenging. Furthermore, QC technology has a high risk of error. Bridge, displacement, misalignment, and cell omission faults, as well as stuck-at-fault, are likely to occur in the gates and interconnections, resulting in a significant error rate when compared to standard CMOS technology. Due to increased proliferation and need for more concurrency, computing systems have evolved over time, from parallel and distributed computing to the most recent technologies such as cloud computing. With the introduction of quantum computing, a paradigm shift in computing methods has been observed, with exponential performance increases, issue optimization, and the solvability of a few previously intractable problems.

Quantum computing is a term that encompasses breakthroughs in a variety of fields, including manufacturing, artificial intelligence, machine learning, and data science. Computer scientists and researchers have focused their efforts on moving away from traditional computing techniques and towards quantum computing specifically as shown in Fig. 1.

4.5 Impact of Power Dissipation

However, when electrical devices are turned on and off, the chip temperature rises in a regular pattern. Heat sinks are used to disperse the heat generated by power

dissipation. The packaging has a higher thermal resistance than the heat sink. Heat is attracted to the heat sink as a result. The rate of heat transmission to the environment must be greater than the rate of heat creation in order to efficiently remove heat.

5 Conclusion

This scholarly literature provides a thorough examination of low-power circuit-based system concepts, strategies, and power management techniques. One of the most significant criteria for next generation computer and communication device operators is power efficiency (i.e. 4G-5G). Power dissipation, on the other hand, has emerged as a serious issue in the design of very large scale integrated (VLSI) circuits, notably for the manufacture of smaller integrated circuits (ICs) and portable electronic devices. Several process technologies, including as circuits, gates, and register-transfer (RT), have garnered a lot of attention in order to decrease power dissipation. According to previous experience with the old method, system-level design optimization has the biggest impact on power usage. Furthermore, as design complexity decreases below the hundreds of nanometre node, overall power management of a device becomes a major difficulty. The power management of low-power VLSI devices is affected by leakage current. As a result, as compared to other techniques, system-level power optimization solutions can save a significant amount of energy. Quantum dot computing has recently been identified as one of the forthcoming technologies that might eventually replace CMOS. It has received a lot of attention in recent years because to its extremely low-power dissipation, high operating frequency, and small size.

References

1. Chandrakasan AP, Sheng S, Brodersen RW (1992) Low-power CMOS digital design. IEEE J Solid-State Circuits 27:473–484
2. Rabaey J, Pedram M (1996) Introduction. In: Low power design methodologies, 1st edn. Kluwer, New York, pp 5–16
3. Chandrakasan AP, Mehra R, Potkonjak M, Rabaey J, Brodersen RW (1995) Optimizing power using transformations. In: IEEE transactions On CAD, pp 13–32
4. Rabaey J, Pedram M (1996) Algorithm and architectural level methodologies. In: Low power design methodologies 1st edn. Kluwer, New York, pp 335–340
5. Zhang YX, Lu SL, Mao BQ (2004) A low-power design methodology clock-gating. Microelectron Comput 21:23–26
6. Yaman Çakmak I, Toms W, Navaridas J (2016) Cyclic power-gating as an alternative to voltage and frequency scaling. IEEE Comput Archit Lett 15:77–80
7. Shin I, Kim J-J, Shin Y (2015) Aggressive voltage scaling through fast correction of multiple errors with seamless pipeline operation. In: IEEE transactions on circuits and systems I, vol 62, Issue 2, pp 468–477
8. Benini L, De Micheli G, Macii E (2002) Designing low-power circuits: practical recipes. In: IEEE circuits and systems magazine, vol 1, pp 6–25

9. Han J, Member, IEEE, Zhang Y, Huang S, Chen M, Zeng X (2016) An area-efficient error-resilient ultra-low-power subthreshold ECG processor. In: IEEE transactions on circuits and systems-II, vol 2
10. Chen Z, Shott J, Plummer J (1994) CMOS technology scaling for low voltage low power applications. In: ISLPE-98: IEEE international symposium on low power electronics. San Diego, CA, pp 56–57
11. Bernstein K, Cavin RK, Porod W, Seabaugh A, Welser J (2010) Device and architecture outlook for beyond CMOS switches. Proc IEEE 98(12):2169–2184
12. Bondyopadhyay PK (2002) Moore's law governs the silicon revolution. Proc IEEE 88:78–81
13. Haron NZ, Hamdioui S (2008) Why is CMOS scaling coming to an END? In: Proceedings of the 2008 3rd international design and test workshop, Monastir, Tunisia, pp 98–103
14. Dennard RH, Gaensslen FH, Yu H-N, Leo Rideovt V, Bassous E, Leblanc AR (2007) Design of ion-implanted MOSFET's with very small physical dimensions. In: IEEE solid-state circuits society newsletter, vol 12, pp 38–50
15. International technology roadmap for semiconductors, Process integration devices and structures (PIDS) (2011) http://www.itrs.net/Links/2011ITRS/Home2011.htm
16. Lent CS, Tougaw PD, Porod W, Bernstein GH (1993) Quantum cellular automata. Nanotechnology 4(1):49–57
17. Snider GL, Orlov AO, Amlani I et al (1999) Quantum-dot cellular automata: review and recent experiments (invited). J Appl Phys 85(8):4283–4285
18. Kim SW, Swartzlander EE (2009) Parallel multipliers for quantum-dot cellular automata. In: Proceedings of the 2009 IEEE nanotechnology materials and devices conference, Traverse City, MI, USA, pp 68–72
19. Kim SW, Swartzlander EE (2010) Multipliers with coplanar crossings for quantum-dot cellular automata. In: Proceedings of the 10th IEEE international conference on nanotechnology, Seoul, Republic of Korea, pp 953–957
20. Balali M, Rezai A, Balali H, Rabiei F, Emadi S (2017) Towards coplanar quantum-dot cellular automata adders based on efficient three-input XOR gate. Results Phys 7:1989–1995
21. Sasamal TN, Singh, AK, Mohan A (2016) Efficient design of reversible ALU in quantum-dot cellular automata. Optik 127(15):6172–6182
22. Sasamal TN, Singh AK, Ghanekar U (2016) Design of non-restoring binary array divider in majority logic-based QCA. Electron Lett 52(24):2001–2003
23. Kianpour M, Nadooshan RS (2011) A novel modular decoder implementation in quantum-dot cellular automata (QCA). In: Proceedings of the 2011 international conference on nanoscience, technology and societal implications (NSTSI), Bhubaneswar, India, pp 1–5
24. Kianpour M, Nadooshan RS (2016) A novel quantum dot cellular automata X-bit × 32-bit SRAM. IEEE Trans Very Large Scale Integr (VLSI) Syst 24(3):827–836
25. Chougule PP, Sen B, Dongale TD (2017) Realization of processing in-memory computing architecture using quantum dot cellular automata. Microprocess Microsyst 52:49–58