

Lecture Notes on Data Engineering  
and Communications Technologies 57

Jude Hemanth  
Robert Bestak  
Joy long-Zong Chen *Editors*

# Intelligent Data Communication Technologies and Internet of Things

Proceedings of ICICI 2020

 Springer

# **Lecture Notes on Data Engineering and Communications Technologies**

Volume 57

## **Series Editor**

Fatos Xhafa, Technical University of Catalonia, Barcelona, Spain



The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

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

More information about this series at <http://www.springer.com/series/15362>

Jude Hemanth · Robert Bestak ·  
Joy Iong-Zong Chen  
Editors

# Intelligent Data Communication Technologies and Internet of Things

Proceedings of ICICI 2020

 Springer

*Editors*

Jude Hemanth  
Department of Electronics  
and Communication Engineering  
Karunya University  
Coimbatore, Tamil Nadu, India

Robert Bestak  
Czech Technical University  
Prague, Czech Republic

Joy Iong-Zong Chen  
Department of Electrical Engineering  
Dayeh University  
Changhua, Taiwan

ISSN 2367-4512

ISSN 2367-4520 (electronic)

Lecture Notes on Data Engineering and Communications Technologies

ISBN 978-981-15-9508-0

ISBN 978-981-15-9509-7 (eBook)

<https://doi.org/10.1007/978-981-15-9509-7>

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

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

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

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

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*We are honored to dedicate this book to all  
the participants, technical program  
committee members and editors of ICICI  
2020.*

# Foreword

It is with deep satisfaction that I write this Foreword to the Proceedings of the ICICI 2020 held in JCT College of Engineering and Technology, Coimbatore, Tamil Nadu, on August 27–28, 2020.

The conference brought together researchers, academics and professionals from all over the world, experts in Data Communication Technologies and the Internet of Things.

This conference particularly encouraged the interaction of research students and developing academics with the more established academic community in an informal setting to present and discuss new and current work. The papers contributed the most recent scientific knowledge known in the field of data communication and computer networking, communication technologies and its applications like IoT, big data, machine learning, sentiment analysis and cloud computing. Their contributions helped to make the conference as outstanding as it has been. The members of the local organizing committee and their helpers have made a great deal of effort to ensure the success of the day-to-day process of the conference.

We hope that this program will further stimulate research in Intelligent Data Communication Technologies and Internet of Things and provide practitioners with better techniques, algorithms and tools for deployment. We feel honored and privileged to serve the best recent developments in the field of Intelligent Data Communication Technologies and Internet of Things to you through this exciting program.

We thank all guest editors, keynote speakers, technical program committee members, authors and participants for their contributions.

Dr. K. Geetha  
Conference Chair, ICICI 2020

# Preface

This conference Proceedings volume contains the written versions of most of the contributions presented during the conference of ICICI 2020. The conference provided a setting for discussing recent developments in a wide variety of topics including data communication, computer networking, communicational technologies, wireless and adhoc network, cryptography, big data, cloud computing, machine learning, IoT, sentiment analysis and healthcare informatics. The conference has been a good opportunity for participants coming from various destinations to present and discuss topics in their respective research areas.

International Conference on Intelligent Data Communication Technologies and Internet of Things tends to collect the latest research results and applications on Intelligent Data Communication Technologies and Internet of Things. It includes a selection of 70 papers from 313 papers submitted to the conference from universities and industries all over the world. All of the accepted papers were subjected to strict peer-reviewing by 2–4 expert referees. The papers have been selected for this volume because of quality and the relevance to the conference.

The guest editors would like to express our sincere appreciation to all authors for their contributions to this book. We would like to extend our thanks to all the referees for their constructive comments on all papers and keynote speakers; especially, we would like to thank the organizing committee for their hard work. Finally, we would like to thank Springer publications for producing this volume.

Coimbatore, India  
Changhua, Taiwan  
Prague, Czech Republic

Dr. Jude Hemanth  
Dr. Joy Iong-Zong Chen  
Dr. Robert Bestak

# Acknowledgements

ICICI 2020 would like to acknowledge the excellent work of our Conference Steering Committee and keynote speakers for their presentation on August 27–28, 2020. The organizers also wish to acknowledge publicly the valuable services provided by the reviewers.

On behalf of the editors, organizers, authors and readers of this conference, we wish to thank the keynote speakers and the reviewers for their time, hard work and dedication to this conference. The organizers wish to acknowledge the guest editors Dr. Jude Hemanth, Dr. Joy Chen, Dr. Robert Bestak and Conference Chair Dr. K. Geetha for the discussion, suggestion and cooperation to organize the keynote speakers of this conference. The organizers also wish to acknowledge speakers and participants who attended this conference. Many thanks are given to all persons who helped and supported this conference. We would like to acknowledge the contribution made to the organization by its many volunteers. Members have contributed their time, energy and knowledge at a local, regional and international levels.

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

# Contents

<b>A Semi-supervised Learning Approach for Complex Information Networks</b> .....	1
Paraskevas Koukaras, Christos Berberidis, and Christos Tjortjis	
<b>Tackling Counterfeit Medicine Through an Automated System Integrated with QR Code</b> .....	15
A. M. Chandrashekhar, J. Naveen, S. Chethana, and Srihari Charith	
<b>Efficient Gs-IBE for Collusion Resistance on Multi-handler Cloud Environment Using Attribute-Modulator (AM)</b> .....	29
J. Priyanka and M. Ramakrishnan	
<b>Preventing Fake Accounts on Social Media Using Face Recognition Based on Convolutional Neural Network</b> .....	41
Vernika Singh, Raju Shanmugam, and Saatvik Awasthi	
<b>A DNN Based Diagnostic System for Heart Disease with Minimal Feature Set</b> .....	55
Merin David and K. P. Swaraj	
<b>Robust Automated Machine Learning (AutoML) System for Early Stage Hepatic Disease Detection</b> .....	65
Devendra Singh, Pawan Kumar Pant, Himanshu Pant, and Dinesh C. Dobhal	
<b>BS6 Violation Monitoring Based on Exhaust Characteristics Using IoT</b> .....	77
B. S. Pavithra and K. A. Radhakrishna Rao	
<b>Analytical Classification of Sybil Attack Detection Techniques</b> .....	89
Ankita S. Koleshwar, S. S. Sherekar, V. M. Thakare, and Aniruddha Kanhe	
<b>Smart and Autonomous Door Unlocking System Using Face Detection and Recognition</b> .....	99
Amrutha Kalturi, Anjali Agarwal, and N. Neelima	



<b>Frequency Reconfigurable Circular Patch Antenna</b> .....	109
R. S. Janisha, D. Vishnu, and O. Sheeba	
<b>A New Avenue to the Reciprocity Axioms of Multidimensional DHT Through Those of the Multidimensional DFT</b> .....	119
B. N. Madhukar and S. H. Bharathi	
<b>A Deep Learning Approach for Detecting and Classifying Cancer Types</b> .....	135
G. Murugesan, G. Preethi, and S. Yamini	
<b>Automated Surveillance Security System Using Facial Recognition for Homes and Offices</b> .....	143
Gunjan Bharadwaj, Shubham Saini, Anmol, Ajeet Chauhan, and Puneet Kumar	
<b>Strategies for Boosted Learning Using VGG 3 and Deep Neural Network as Baseline Models</b> .....	151
K. S. Gautam, Vishnu Kumar Kaliappan, and M. Akila	
<b>Integrating ICT in Communicative Language Teaching in Bangladesh: Implementation and Challenges</b> .....	169
Nafis Mahmud Khan and Khushboo Kuddus	
<b>Emerging Role of Intelligent Techniques for Effective Detection and Prediction of Mental Disorders</b> .....	185
Priti Rai Jain and S. M. K. Quadri	
<b>Comparison Analysis of Extracting Frequent Itemsets Algorithms Using MapReduce</b> .....	199
Smita Chormunge and Rachana Mehta	
<b>Customer Lookalike Modeling: A Study of Machine Learning Techniques for Customer Lookalike Modeling</b> .....	211
Anna Mariam Chacko, Bhuvanapalli Aditya Pranav, Bommanapalli Vijaya Madhvesh, and A. S. Poornima	
<b>Cold Start in Recommender Systems—A Survey from Domain Perspective</b> .....	223
Rachna Sethi and Monica Mehrotra	
<b>Image Classification Using Machine Learning Techniques for Traffic Signal</b> .....	233
Satthi Reddy, Nishanth, Praharsha, Dattatreya Dash, and N. Rakesh	
<b>Deep Domain Adaptation Approach for Classification of Disaster Images</b> .....	245
Anuradha Khattar and S. M. K. Quadri	

**Automated Intelligent IoT-Based Traffic Lights in Transport Management System** ..... 261  
 Akhilesh Kumar Singh and Manish Raj

**A Document Clustering Approach Using Shared Nearest Neighbour Affinity, TF-IDF and Angular Similarity** ..... 267  
 Mausumi Goswami

**A Privacy Preserving Hybrid Neural-Crypto Computing-Based Image Steganography for Medical Images** ..... 277  
 Tejas Jambhale and M. Sudha

**Low-Power Reconfigurable FFT/IFFT Processor** ..... 291  
 V. Sarada and E. Chitra

**Color Image Watermarking in DCT Domain Using SVR** ..... 301  
 Shabnam Thakur, Rajesh Mehta, and Geeta Kasana

**Leaf Detection Using Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Classifying with SVM Utilizing Claim Dataset** ..... 313  
 Kazi Sultana Farhana Azam, Farhin Farhad Riya, and Shah Tuhin Ahmed

**Prediction of Congestive Heart Failure (CHF) ECG Data Using Machine Learning** ..... 325  
 Manthan S. Naik, Tirth K. Pancholi, and Rathnakar Achary

**Aspect-Based Unsupervised Negative Sentiment Analysis** ..... 335  
 Mainak Ghosh, Kirtika Gupta, and Seba Susan

**A System for Anxiety Prediction and Treatment Using Indian Classical Music Therapy with the Application of Machine Learning** .... 345  
 G. Kruthika, Padmaja Kuruba, and N. D. Dushyantha

**Design of a Water and Oxygen Generator from Atmospheric Pollutant Air Using Internet of Things** ..... 361  
 D. K. Niranjana and N. Rakesh

**Use of LSTM and ARIMAX Algorithms to Analyze Impact of Sentiment Analysis in Stock Market Prediction** ..... 377  
 Archit Sharma, Prakhar Tiwari, Akshat Gupta, and Pardeep Garg

**Design of Medical Image Cryptosystem Triggered by Fusional Chaotic Map** ..... 395  
 Manivannan Doraipandian and Sujarani Rajendran

**User Engagement Recognition Using Transfer Learning and Multi-task Classification** ..... 411  
 Hemant Upadhyay, Yogesh Kamat, Shubham Phansekar, and Varsha Hole

**QoS Aware Multi Mapping Technology in SD-WAN** ..... 421  
 Viswanathan Varsha and C. N. Sminesh

**Auto-Completion of Queries** ..... 435  
 Vidya S. Dandagi and Nandini Sidnal

**Design of CMOS Active Inductors for RFIC Applications: A Review** ... 447  
 Zishani Mishra, T. Prashanth, N. Sanjay, Jagrati Gupta, and Amit Jain

**Optimized Web Service Composition Using Evolutionary Computation Techniques** ..... 457  
 S. Subbulakshmi, K. Ramar, Anvy Elsa Saji, and Geethu Chandran

**Emotion Recognition from Speech Signal Using Deep Learning** ..... 471  
 Mayank Chourasia, Shriya Haral, Srushti Bhatkar, and Smita Kulkarni

**COVID-19 Database Management: A Non-relational Approach (NoSQL and XML)** ..... 483  
 Priya Shah, Rajasi Adurkar, Shreya Desai, Swapnil Kadakia, and Kiran Bhowmick

**Study of Effective Mining Algorithms for Frequent Itemsets** ..... 499  
 P. P. Jashma Suresh, U. Dinesh Acharya, and N. V. Subba Reddy

**A Novel Method for Pathogen Detection by Using Evanescent-Wave-Based Microscopy** ..... 513  
 Vijay A. Kanade

**Analyzing Social Media Data for Better Understanding Students’ Learning Experiences** ..... 523  
 T. Ganesan, Sathigari Anuradha, Attada Harika, Neelisetty Nikitha, and Sunanda Nalajala

**Load Balancing in Cloud Computing Environment: A Broad Perspective** ..... 535  
 Minakshi Sharma, Rajneesh Kumar, and Anurag Jain

**An Energy-Efficient Routing Protocol Using Threshold Hierarchy for Heterogeneous Wireless Sensor Network** ..... 553  
 Ashok Kumar Rai and A. K. Daniel

**Performance Analysis of Fuzzy-Based Relay Selection for Cooperative Wireless Sensor Network** ..... 571  
 Nitin Kumar Jain and Ajay Verma

**Rational Against Irrational Causes of Symptom Recognition Using Data Taxonomy** ..... 583  
 S. M. Meenaatchi and K. Rajeswari

**Multimodal Emotion Analytics for E-Learning** ..... 593  
 J. Sirisha Devi and P. Vijaya Bhaskar Reddy

**Intrusion Detection: Spider Content Analysis to Identify Image-Based Bogus URL Navigation** ..... 603  
 S. Ponmaniraj, Tapas Kumar, and Amit Kumar Goel

**Plant Species Identification Using Convolutional Neural Network** ..... 619  
 Harsha H. Ashturkar and A. S. Bhalchandra

**Twitter Sentiment Analysis Using Supervised Machine Learning** ..... 631  
 Nikhil Yadav, Omkar Kudale, Aditi Rao, Srishti Gupta, and Ajitkumar Shitole

**Classification of Skin Disease Using Traditional Machine Learning and Deep Learning Approach: A Review** ..... 643  
 Honey Janoria, Jasmine Minj, and Pooja Patre

**Spam SMS Filtering Using Support Vector Machines** ..... 653  
 P. Prasanna Bharathi, G. Pavani, K. Krishna Varshitha, and Vaddi Radhesyam

**Design of Open and Short-Circuit Stubs for Filter Applications** ..... 663  
 S. Aiswarya and Sreedevi K. Menon

**Wireless Sensor Networks and Its Application for Agriculture** ..... 673  
 Kirankumar Y. Bendigeri, Jayashree D. Mallapur, and Santosh B. Kumbalavati

**Human Disease Diagnosis Using Machine Learning** ..... 689  
 Sarika Jain, Raushan Kumar Sharma, Vaibhav Aggarwal, and Chandan Kumar

**Soccer Result Prediction Using Deep Learning and Neural Networks** ..... 697  
 Sarika Jain, Ekansh Tiwari, and Prasanjit Sardar

**Ranking Diabetic Mellitus Using Improved PROMETHEE Hesitant Fuzzy for Healthcare Systems** ..... 709  
 K. R. Sekar, S. Yogapriya, N. Senthil Anand, and V. Venkataraman

**Hybrid Recommendation System for Scientific Literature** ..... 725  
 Indraneel Amara, K Sai Pranav, and H. R. Mamatha

**Classification of VEP-Based EEG Signals Using Time and Time-Frequency Domain Features** ..... 737  
 M. Bhuvaneshwari and E. Grace Mary Kanaga

**Performance Analysis of Matching Criteria in Block-Based Motion Estimation for Video Encoding** ..... 749  
 Awanish Kumar Mishra and Narendra Kohli

**Convolutional Recurrent Neural Network Framework for Autonomous Driving Behavioral Model** ..... 761  
 V. A. Vijayakumar, J. Shanthini, and S. Karthick

**Video Enhancement and Low-Resolution Facial Image Reconstruction for Crime Investigation** ..... 773  
Joel Eliza Jacob and S. Saritha

**Facial Emotion Recognition System for Unusual Behaviour Identification and Alert Generation** ..... 789  
P. Sarath Chandran and A. Binu

**Secure Digital Image Watermarking by Using SVD and AES** ..... 805  
K. M. Sahila and Bibu Thomas

**BER Analysis of Power Rotational Interleaver on OFDM-IDMA System Over Powerline** ..... 819  
Priyanka Agarwal, Ashish Pratap, and M. Shukla

**Offline 3D Indoor Navigation Using RSSI** ..... 831  
S. Vivek Sidhaarthan, Anand Mukul, P. Ragul, R. Gokul Krishna, and D. Bharathi

**Meanderline Pattern Wearable Textile Antenna for Position Identification in Military Applications** ..... 847  
Pruthvi Tenginakai, V. Keerthana, Sowmini Srinath, Fauzan Syed, and P. Parimala

**Authentication of Robots Using ECC to Access Cloud-Based Services** ..... 861  
Saurabh Jain and Rajesh Doriya

**Intelligent Web of Things Based on Fuzzy Neural Networks** ..... 871  
Zahraa Sabeeh Amory and Haider Kadam Hoomod

**Author Index** ..... 889

# About the Editors

**Dr. Jude Hemanth** received his B.E. degree in ECE from Bharathiar University in 2002, M.E. degree in Communication Systems from Anna University in 2006 and Ph.D. from Karunya University in 2013. His research areas include computational intelligence and image processing. He has authored more than 100 research papers in reputed SCIE indexed/Scopus indexed international journals and international conferences with leading publishers such as Elsevier, Springer, IEEE, etc. His Cumulative Impact Factor is more than 100. He has authored 1 book with (VDM-Verlag, Germany) and 24 edited books with reputed publishers such as Elsevier, Springer, IET and IOS Press.

**Dr. Robert Bestak** obtained a Ph.D. degree in Computer Science from ENST Paris, France (2003), and M.Sc. degree in Telecommunications from Czech Technical University in Prague, CTU, (1999). Since 2004, he has been an Assistant Professor at the Department of Telecommunication Engineering, Faculty of Electrical Engineering, CTU. His main research interests include 5G networks, cognitive networks and spectrum management. He is the Czech representative in the IFIP TC6 working group, and he serves as Associate Editor of Telecommunication System and Electronic Commerce Research, Springer. Dr. Bestak has served as Steering and Technical Program Committees member for numerous IEEE/IFIP international conferences. He participated in several national and EU founded research projects (FP7-ROCKET, FP7-TROPIC, etc.).

**Dr. Joy Iong-Zong Chen** is currently a Full Professor of the Department of Electrical Engineering Dayeh University at Changhua Taiwan. Prior to joining the Dayeh University, he worked at the Control Data Company (Taiwan) as a Technical Manager from September 1985 to September 1996. His research interests include wireless communications, spread spectrum technical, OFDM systems and wireless sensor networks. He has published a large number of SCI journal papers in the issues addressed physical layer for wireless communication systems. Moreover, he also majors in developing some applications of the IOT (Internet of Thing) techniques, and Dr. Joy I.-Z. Chen owned some patents authorized by the Taiwan Intellectual Property Office (TIPO).

# A Semi-supervised Learning Approach for Complex Information Networks



Paraskevas Koukaras, Christos Berberidis, and Christos Tjortjjs

**Abstract** Information Networks (INs) are abstract representations of real-world interactions among different entities. This paper focuses on a special type of Information Networks, namely Heterogeneous Information Networks (HINs). First, it presents a concise review of the recent work in this field. Then, it proposes a novel method for querying such networks, using a bi-functional machine learning algorithm for clustering and ranking. It performs and elaborates on supervised and unsupervised, proof-of-concept modeling experiments on multi-typed, interconnected data while retaining their semantic importance. The results show that this method yields promising results and can be extended and utilized, using larger, real-world datasets.

**Keywords** Information networks · Machine learning · Supervised learning · Unsupervised learning · Clustering · Ranking · Social media · NoSQL · Graph databases · Big data

## 1 Introduction

During the last decade, a great surge in data generation has been observed resulting from the extensive usage of Social Media (SM) from users around the world. The need for effective warehousing and modeling of this information is evident when trying to capture the vast structural and semantic diversification of these data. Until recently, information was treated as homogeneous, but since the intense rise of SM,

---

P. Koukaras · C. Berberidis · C. Tjortjjs (✉)

The Data Mining and Analytics Research Group, School of Science and Technology, International Hellenic University, 57001 Thessaloniki, Greece

e-mail: [c.tjortjjs@ihu.edu.gr](mailto:c.tjortjjs@ihu.edu.gr)

P. Koukaras

e-mail: [p.koukaras@ihu.edu.gr](mailto:p.koukaras@ihu.edu.gr)

C. Berberidis

e-mail: [c.berberidis@ihu.edu.gr](mailto:c.berberidis@ihu.edu.gr)

has led to the search for new ways to handle this vast interconnected and interacting information. Therefore, the concept of Heterogeneous Information Networks (HIN) has emerged offering new opportunities for representing real-world Information Networks (IN) with the usage of multi-typed and multi-layer nodes and links.

To that end, improvements in IN were proposed offering a transition from homogeneous IN to HIN being accompanied by algorithmic modifications and improvements for all kinds of Machine Learning (ML) tasks. Surveying the state-of-the-art, the proposed approach can generate new insights by combining an experimental with a theoretical implementation offering a novel approach in the study of data modeling, IN and ML methods. This study builds upon existing knowledge on this domain of studies [1, 2].

Initial research is presented improving the state-of-the-art methodologies regarding ML concepts, such as supervised and unsupervised learning. HIN can be modeled utilizing NoSQL databases enforcing use-case tailored IN modeling techniques. The experimentation conducted aims to present and support concrete facts and the reasoning behind the SM domain of studies.

The concept of SM involves very complex interactions of data entities and presents the characteristics of an ecosystem comprising of multiple SM platforms with a wide diversity of functionalities and/or multi-typed data and relations. Work such as [3] offers an overview of the historical concepts regarding social networks and the necessary information for understanding their structure [4].

Outlining the abovementioned concepts, HIN is utilized mitigating any issues arising from real-world data modeling processes. They address the preservation of information integrity and loss of data through information abstraction since they allow two or more different objects to be associated (linked) offering complex structure formation and rich semantics. Since they pose multi-layer information networks utilizing concepts from graph theory, they allow data to be modeled while introducing numerous new opportunities for performing ML tasks; either by updating/adjusting already existing algorithms or conceiving new ones. On the other hand, they may become computationally expensive due to the vast amount of data handling. Therefore, they should be handled with state-of-the-art databases. This paper presents an overview of the literature and preliminary experimentation while presenting a novel method. It is structured as follows: Sect. 2 presents an overview of IN and their evolution. Section 3 defines the problem. Section 4 outlines the process of the proposed novel methodology. Section 5 presents the experiments. Finally, Sect. 6 discusses preliminary results and future work.

## 2 Information Networks and Mining Tasks

The surge of SM usage has led to the evolution of IN. Homogeneous networks and the graph information modeling that they impose, seem inadequate to support the increasing demands of such diversification. This diversification should be present in



nodes and links to correctly schematize and represent real-world networks. In SM, entities continuously interact with one another forcing the identification of various multi-typed components. For that reason, the research frontier evolved introducing HIN, promising to handle and maintain these complex data streams.

HIN was proposed in [5] with various research attempts debuting the following years, greatly enhancing the information retrieval capabilities of extremely rich IN [6]. Data modeling is the most important task that acts as a prerequisite for performing any ML task. Therefore, many algorithms need to be modified or redesigned to work with HIN. Richer data create more opportunities for data mining, but, at the same time, it alleviates the chances for arising issues regarding their handling and management. According to HIN theory, IN structure involve multiple IN that are interconnected, i.e. multilayer networks [7]. The basic inception of HIN addresses the issue of real-world/system data modeling, such as social activities, bio networks characterized by multiple relationships or types of connections [4]. Important role to that end plays the IN analysis process that is performed in vast IN, but in a way, that preserves the information structural integrity and generality [5]. Research domains, such as social network analysis, web mining, graph mining, etc. allow for further elaboration on effective IN analysis [8]. Examples of the basic mining tasks that are performed in HIN are: clustering [9], classification [10], similarity search [11], and more, while more advanced data mining tasks refer to pattern mining, link associations, object and link assumptions [6].

## 2.1 Supervised Learning Methods

In general, supervised ML methods include regression and classification. The difference resides in that regression methods deal with numeric values for feature labeling, while classification is not limited to numeric values for the distinct feature labeling. Very common such algorithms are Decision Trees (DT), Naïve Bayes (NB), Stochastic Gradient Descent (SGD), K-Nearest Neighbours (KNN), Random Forest (RF), Support Vector Machine (SVM), and more [12–14].

These examples need to be modified so that their functionality is adjusted to graph-based IN such as HIN. Classification or regression algorithms split the labeled datasets into two sets, the training and test set respectively, utilizing a process for evaluating performance (e.g.,  $k$ -fold cross-validation) of the algorithms [13]. This is a straightforward process when dataset objects present the same structure that is quite uncommon for real-world datasets. Graph theory comes into place along with linked based object classification introducing nodes that are linked through edges, expanding the capabilities of the aforementioned classification algorithms. Homogeneous IN imposes that the nodes and links have the same characteristics while HIN exposes variations. The structural type of data objects (usually nodes) may be different allowing for simultaneous classification of multi-typed data structures while their labels/attributes are linked (directionally or bi-directionally) with other nodes.

By representing information with HIN and utilizing graphs, generates new opportunities for visualization and prediction of information transfer between multi-typed nodes. Examples of classification tasks involving HIN are, Inductive Classification (IC), that deals with consistent decision making regarding very large datasets, Transductive Classification (TC) dealing with the labeling of non-labeled data, Ranked-based Classification (RC), dealing with ranking and classification simultaneously and more [8].

## 2.2 *Unsupervised Learning Methods*

One of the most common tasks of unsupervised learning is clustering, where algorithms attempt to create groups of objects contained in datasets that exhibit similar characteristics. This is achieved by utilizing a variety of distance measures for determining the distance of various objects from each cluster's center. Common such measures are Manhattan, Euclidean, Cosine or Jaccard distances [15].

Similar to supervised learning methods, metrics are utilized for evaluating the results of any unsupervised learning algorithm that is adjusted to work with SM data. Examples of these are, qualitative measures such as Separateness, which measures the distance between clusters, Cohesiveness that measures the distance in node-to-node scale, or the silhouette index that represents a mixed measure implementing both Cohesiveness and Separateness at the same time [16].

Complex IN, such as HIN, generates other possible issues related with formed communities of objects (nodes). The majority of already implemented methods addressing this topic offer approaches that incorporate subgraphs. Subgraphs are generated for faster and easier handling of such data with lots of research dealing with the optimization and improvement of this mining task [17].

HIN allows for better handling of data that are packed with a large variety of different attributes/features. The structure representation of these datasets may involve multi-typed nodes and links. All in all, HIN push the research frontiers on information modeling, expanding capabilities for more complex and demanding mining tasks, since they showcase ways for storing and handling very large datasets, like SM data, envisioning the development of even more novel supervised learning methods.

## 3 **Problem Definition**

In real-world situations, data entities are usually objects, individuals, groups, or components, developing a unique way of interacting with one another, on their semantic meaning. The structure that offers a general form of these relationships comprise IN. The commonest IN these days are SM networks, Biological networks, Traffic networks, etc.

In the case of social networks, information about individuals concerning their activities, behaviors, connections, and preferences, offers great opportunities for extensive information analysis. The research community has made great efforts to model this information by extending previous state-of-the-art to match the requirements of the new information era. Although most of the attempts deal with concepts regarding textual information and their classification, clustering or ranking, now, they do pay much attention to IN structures.

Furthermore, most of the approaches on clustering and ranking address the problem of global ranking and clustering on complex IN. This paper aims to highlight ongoing work regarding an approach for effectively discovering knowledge through querying on a graph database on close vicinity of a specific node, thus locally. The experimental part of the approach was implemented in Java while being conceived in a way to be integrated with any multi-model database that supports the functionalities of a NoSQL Database Management Systems (DBMS), e.g. Neo4j [18]. The next section presents the proposed methodology addressing the issue of bi-functional ranking and clustering.

## 4 Proposed Methodology

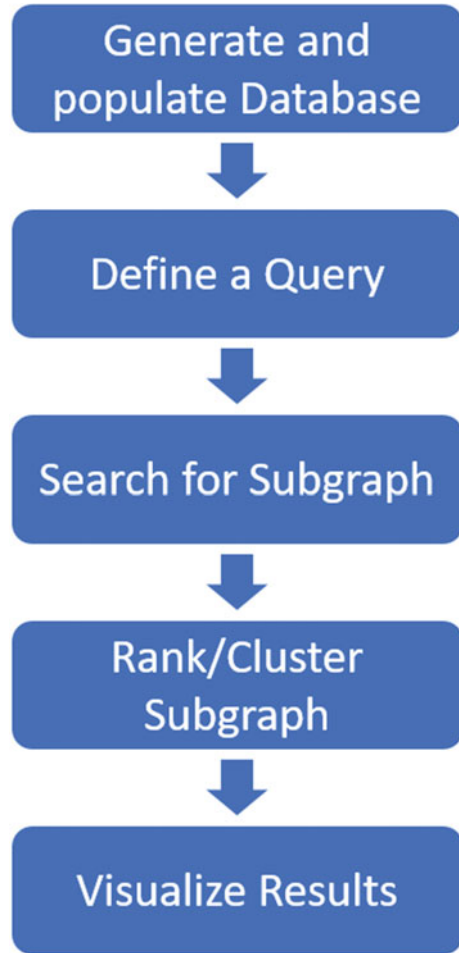
The proposed methodology envisions to improve the understanding of the fields of IN modeling through the utilization of graph theory. The proposed approach requires the employment of a NoSQL DBMS for modeling the rich information contained in IN such as HIN. Then the development, testing, and evaluation of a bi-functional algorithm covering the tasks of clustering and ranking simultaneously such as RankClus [19], takes place.

The state-of-the-art covers the topic of global ranking and clustering while the proposed methodology envisions to cluster and rank multi-typed data entities in close vicinity after a user prompt, utilizing a NoSQL DBMS [18]. The proposed algorithm is considered a bi-functional algorithm belonging to the semi-supervised learning where the human intervention is required for specifying results after the algorithmic execution [8].

Next, the theoretical conception of the proposed algorithm is presented. As stated in previous sections, data modeling is a very important task before any mining task commences. For very large graphs, it is crucial to decide on the network schema to be utilized. Examples of such schemas are Multi-relational network, Bipartite network, Star-schema networks, and more [20]. The steps of the proposed algorithm are presented in Fig. 1.

First, a NoSQL DBMS with data from SM or other multi-typed data is created and populated. Next, a query  $q$  that searches for a node is defined. Around that node, the subgraph  $s$  is determined by applying constraints regarding its size and links. For example, the user can specify a threshold regarding many links or nodes to be included in the query results. This is achieved by executing the KNN algorithm [14] to define and store all nodes in close vicinity from node  $n$ , where  $i$  is a limiter

**Fig. 1** Process flow diagram of the proposed algorithm



for the sum of nearby neighboring nodes. This process results in the subgraph  $s$ . Then, ranking and clustering on objects of  $s$  (including link and node attributes) are performed discovering knowledge about  $q$ , displaying various existent relations by utilizing the RankClus algorithm [19] or a modification of it, highlighting the semi-supervised capabilities of a bi-functional machine-learning algorithm. Next, lists are generated as shown in Table 2 presenting clusters and ranked lists resulting from the algorithm execution. Finally, tests and graph evaluation metrics [21] can take place in the results output.

In this paper, the mathematical formulation of the proposed approach is omitted, since it constitutes part of future work. To that end, this section includes a process flow diagram of ongoing research along with a description of it. Once the experimental results expand and elaborate on the proposed method's capabilities in real-world datasets, a formula is to be generated.

**Table 1** Experimental dataset description

Entries	Description
$I_1, \dots, I_{100}$	$i$ ranges between 1 and 100, representing the pool of Influencers
SM	The SM that each Influencer is associated with. The distinct values are: Facebook, YouTube, Twitter, Instagram, LinkedIn, VK, Google+
Topic	The topic that each influencer is associated with. the distinct values are: Advertising, News, Technology, Games, Shopping, Music, Sports, Movies, Travel, Discussion

## 5 Preliminary Experimentation

For the first steps of the proposed approach, preliminary experimentation is presented on the addressed issue by testing the functionality of RankClus algorithm [19], on a custom-made synthetic dataset, specifically tailored to match the needs and highlighting this study's methodological approach. The aim is to present both the IN structuring prospects utilizing graphs and the ML possibilities. The dataset is labeled as InfluencerDB as it contains entries describing several SM influencers, the topic categories that are involved with, and the SM they use. Testing is conducted with Java while utilizing the IntelliJ IDEA [22].

For the experiments, custom-made Java classes were generated enabling the processing of the InfluencerDB dataset. The workflow is straightforward. First, read the dataset from a.xml file and then use Java classes, implementing functions for running the experimental steps. The results of this experiment produce clusters, each representing SM Topics while ranking the Influencers contained in each cluster, taking into consideration the frequency of SM usage.

### 5.1 Dataset

For this preliminary test of the proposed methodology, a synthetically generated dataset is utilized described by a.xml file containing the necessary entries for experimentation. The dataset contains entries for 100 Influencers, seven SM networks, and 10 Topics, accompanied by the respective Influencer-Topic links and Influencer-SM links. According to Table 1 the choice of SM and Topic entries is based on findings from [1].

This is part of the attempt to model a simple IN utilizing graph concepts (edges, links). The nodes and links were generated in a random way to fulfil the testing requirements and they do not contain any real-world associations. In addition, according to Fig. 2, for ease in visualization purposes<sup>1</sup> a part of the aforementioned multi-layered architecture is distinguished. Therefore, a graph-based IN consisting

<sup>1</sup>Figure 2 presents six Influencers, seven SM networks, and four Topics. There are nine Influencer-Topic links and 14 Influencer-SM links.

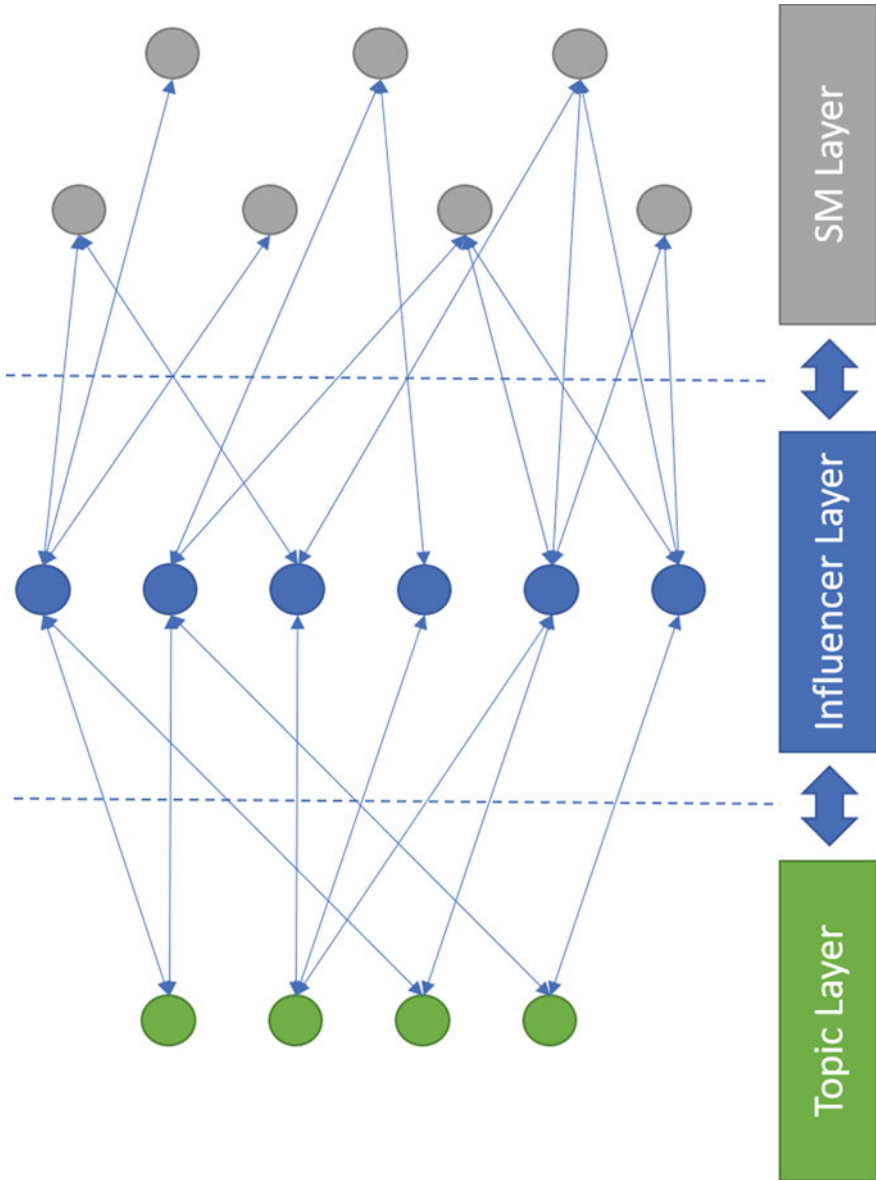


Fig. 2 Synthetic dataset information network

of three layers of information, one for Influencers, one for Topics, and one for SM entities.

In real-world situations, these layers can be numerous presenting in a more detailed way similar IN structures.

## 5.2 Algorithm

This section presents the functionality description of the RankClus [19] algorithm. This study aims to expand on that and use it for the proposed approach for ranking and clustering in SM IN. RankClus implementation converts the input objects utilizing a mixed model for each cluster. This is achieved while calculating the newly appointed positions to the nearest cluster before considering any re-allocation resulting from the identification of points with new attributes. Next, an iterative process takes place until the point that no significant positional updates can be spotted on all of the newly appointed positions of objects. In that way, clustering becomes better, essentially appointing identical objects to the same object group. At the same time, the ranking also improves by generating better attributes for comparison taking place in the next iterations.

The algorithmic process is split into a few consequent and repeating steps:

- First, clusters are generated randomly, meaning that the dataset entries (objects) are placed under a random cluster label.
- Next, the ranking procedure calculates the conditional rank of the clusters generated in the previous step. If any clusters do not contain any objects, then restart.
- Then, mixed model component coefficients are calculated. Another iteration takes place for getting the values of the new objects along with centroids (center points) of each cluster.
- The position of all objects is adjusted and are placed to the nearest cluster in case their distance from the center of their cluster was altered. That way the clusters are re-calculated.

Finally, all the above steps are repeated until reaching a point that further iterations do not impose major alterations (process abiding with the previous points) to cluster formation.

## 5.3 Experimentation

For the experiments, IntelliJ IDEA [19] and Java were utilized. Data were imported and processed. The experimental process is split into subtasks; responsible for implementing and producing preliminary results related to the proposed methodology.

- The most important subtask is the one that generates a cluster instance for performing the actual clustering task. In this subtask, all the internal steps are implemented including iterations for the RankClus algorithm [19].
- Next, a subtask for ranking influencers is implemented while incorporating functions for comparing during the iterative inner steps of the algorithm.
- Two more subtasks are defined for describing and handling all the necessary, processes that associate link objects with their properties (id, source, and destination) and node objects with their properties (id, name) respectively.
- In addition, another subtask is generated for handling the dataset stream.
- Another subtask is utilized for the storage of the output from execution into HashMaps.<sup>2</sup>

Finally, two more subtasks are responsible for handling a ranked influencer and for initializing arrays regarding clustered objects respectively.

## 5.4 Elaboration on Results

This section is dedicated to presenting experimental results. The goal is to offer an overview of the bi-functional process of clustering and ranking. Since the dataset is synthetic, the validity of the results cannot be properly examined at the current state of ongoing research.

The nodes and links for Influencers, Topics, and SM are randomly generated, yet this synthetic dataset matches the research domain, i.e. INs representing social networks. Moreover, the ML tasks under scrutiny (clustering and ranking) enable the discovery of knowledge regarding the importance of a node within a network (ranking) and at the same time appoint it to a group enforcing a categorization of entities (clustering). RankClus [19] is an algorithm that can achieve that, meaning that it enforces clustering and ranking simultaneously through an iterative process, without diminishing the weight of the results for each of the distinct processes. Next, a table is presented, summing up the results after executing the experiment. A cluster for each Topic is generated taking into consideration the entries to the dataset while the ranking of individual Influencers is calculated according to the variety of SM they utilize.

Table 2 presents the results after running the implemented proposed methodology on the artificially generated dataset (Sect. 5.1). The top three ranked SM are presented in descending order as well as the top five ranked Influencers in descending order while being clustered under 10 specific Topics. For example, the Topic Advertising is most common (often observed in the dataset) when linked with Instagram, Facebook, and LinkedIn and to the Influencers  $I_{54}$ ,  $I_{34}$ ,  $I_{98}$ ,  $I_{12}$ ,  $I_{76}$ . Similarly, the Topic Technology is most common (often observed in the dataset) when linked with YouTube,

---

<sup>2</sup><https://docs.oracle.com/javase/8/docs/api/java/util/HashMap.html>.



**Table 2** The output of RankClus algorithm

Cluster	SM <sup>a</sup>	Influencer <sup>b</sup>	Topic
1	Instagram, Facebook, LinkedIn	$I_{54}, I_{34}, I_{98}, I_{12}, I_{76}$	Advertising
2	Twitter, Facebook, YouTube	$I_{23}, I_{99}, I_{49}, I_{57}, I_2$	News
3	YouTube, LinkedIn, Instagram	$I_{85}, I_{39}, I_{93}, I_{75}, I_{11}$	Technology
4	YouTube, VK, Google+	$I_{13}, I_{79}, I_{93}, I_3, I_{87}$	Games
5	Instagram, Facebook, YouTube	$I_{67}, I_{79}, I_{53}, I_{81}, I_{53}$	Shopping
6	YouTube, Facebook, Google+	$I_{23}, I_{16}, I_{33}, I_8, I_{95}$	Music
7	YouTube, Facebook, VK	$I_9, I_{19}, I_{100}, I_{69}, I_{34}$	Sports
8	YouTube, VK, Google+	$I_{97}, I_{71}, I_{45}, I_2, I_{92}$	Movies
9	YouTube, Instagram, VK	$I_{41}, I_{19}, I_{93}, I_9, I_7$	Travel
10	Twitter, LinkedIn, Google+	$I_7, I_{95}, I_{17}, I_9, I_{12}$	Discussion

<sup>a</sup>The Top three ranked SM are presented in a descending order

<sup>b</sup>The Top five ranked Influencers are presented in a descending order

LinkedIn, Instagram, and to the Influencers  $I_{85}, I_{39}, I_{93}, I_{75}, I_{11}$ . It is observed that each Topic forms a cluster containing SM and Influencer objects.

Some of the limitations of that preliminary implementation reside to possible computational issues and convergence speeds in case that very large datasets are to be utilized. Furthermore, the preliminary experimentation involves three types of objects, therefore, on a three-type information network without testing on  $k$ -typed information networks where  $k \geq 3$  which is the case in most of the real-world information networks. In addition, according to RankClus [19] implementation, the quality of the initial clusters define the number of times that the algorithm iterates. Seed objects could be used in case of very large datasets to improve the performance of the proposed approach. Since a user-specified object to be clustered and ranked is referred to, some executions of the proposed approach might perform better (faster) than others due to smaller numbers of distinct values contained in the dataset.

## 6 Discussion and Future Work

This paper presented ongoing work regarding bi-functional ML algorithms while attempting to enhance understanding in the specific domain of IN structures and HIN that may expand to various domains of study [23]. This type of IN is appropriate for modeling information streams generated for example from SM enabling opportunities for SM Analytics [24, 25]. The literature displays that mining tasks can be quite challenging. The experimentation is conducted utilizing stored and synthetic data attempting to evaluate a concrete methodological approach.

The contributions are summarized to the following points.

1. The state-of-the-art approaches the topic of global ranking and clustering. A method for local clustering and ranking is envisioned, modeling the streams of data utilizing graphs, allowing for knowledge discovery through querying on nodes.
2. A baseline for further dealing with bi-functional algorithms is prepared; through experimentation and presentation of necessary concepts associated with this field, the implementation, and testing on complex information structures for expanding ongoing research in data analytics.
3. Current research progress generated the appropriate infrastructure for conducting experiments on real-world SM datasets (live or historical).

Due to the GDPR standards [26], acquiring SM datasets for extensive experimentation on SM may become quite difficult due to the ethical, law, confidentiality, privacy, and more issues that they involve. This study performed preliminary experimentation on an artificially generated dataset that mimics a real-world network. This is done acknowledging that it prepares a solid ground for experimentation that is more thorough, utilizing officially approved, supplied, and accessed real datasets.

Based on the aforementioned points and the limitations discussed in Sect. 5.4, the identified future work comprises of the following tasks as it poses a natural expansion of the proposed approach involving further experimental results and elaboration.

- A. Improve the visualization of the results by incorporating a state-of-the-art graph visualization tool like GraphXR [27].
- B. Create the appropriate data infrastructure for the facilitation of very large datasets such as Neo4j [18]. This will generate a concrete data warehousing structure for further testing on other mining tasks.
- C. Conduct exhaustive testing and validation on the experimentation; expand on datasets that are used by the research or practitioner community aiming to highlight possible use cases and the performance of the novel method.

## References

1. Koukaras P, Tjortjis C, Rousidis D (2020) Social media types: introducing a data driven taxonomy. *Computing* 102(1):295–340
2. Rousidis D, Koukaras P, Tjortjis C (2020) Social media prediction: a literature review. *Multimedia Tools Appl* 79(9):6279–6311
3. Gundecha P, Liu H (2012) Mining social media: a brief introduction. *Tutorials Oper Res* 1
4. Boyd DM, Ellison NB (2007) Social network sites: definition, history, and scholarship. *J Comput-Mediated Commun* 13
5. Han J (2009) Mining heterogeneous information networks by exploring the power of links. *Disc Sci*:13–30
6. Sun Y, Han J (2012) Mining heterogeneous information networks: a structural analysis approach. *SIGKDD Explor* 14(2):20–28
7. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks

8. Shi C, Li Y, Zhang J, Sun Y, Yu PS (2016) A survey of heterogeneous information network analysis. *IEEE Trans Knowl Data Eng* 29(1):17–37
9. Sun Y, Norick B, Han J, Yan X, Yu PS, Yu X (2012) Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. In: *KDD*, pp 1348–1356
10. Kong X, Yu PS, Ding Y, Wild DJ (2012) Meta path-based collective classification in heterogeneous information networks. In: *CIKM*, pp 1567–1571
11. Shi C, Kong X, Yu PS, Xie S, Wu B (2012) Relevance search in heterogeneous networks. In: *EDBT*, pp 180–191
12. Caruana R, Niculescu-Mizil A (2006, June) An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd international conference on machine learning*, pp 161–168
13. Tzirakis P, Tjortjis C (2017) T3C: improving a decision tree classification algorithm’s interval splits on continuous attributes. *Adv Data Anal Classif* 11(2):353–370
14. Lafferty J, McCallum A, Pereira FC (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data
15. Choi SS, Cha SH, Tappert CC (2010) A survey of binary similarity and distance measures. *J Syst Cybern Inform* 8(1):43–48
16. Zafarani R, Abbasi MA, Liu H (2014) *Social media mining: an introduction*. Cambridge University Press
17. Zhou Y, Cheng H, Yu JX (2009) Graph clustering based on structural/attribute similarities. In: *VLDB*, pp 718–729
18. Neo4j. neo4j. <https://neo4j.com/>
19. Sun Y, Han J, Zhao P, Yin Z, Cheng H, Wu T (2009) RankClus: integrating clustering with ranking for heterogeneous information network analysis. In: *EDBT*, pp 565–576
20. Shi, C., & Philip, S. Y. (2017). *Heterogeneous information network analysis and applications*. Springer International Publishing.
21. Hernández JM, Van Mieghem P (2011) *Classification of graph metrics*. Delft University of Technology, Mekelweg, The Netherlands, pp 1–20
22. IntelliJ IDEA. JetBrains. <https://www.jetbrains.com/>
23. Pandian MD (2019) Enhanced network selection and handover schema for heterogeneous wireless networks. *J ISMAC* 1(01):160–171
24. Koukaras P, Tjortjis C (2019) Social media analytics, types and methodology. In: *machine learning paradigms*. Springer, Cham, pp 401–427
25. Koukaras P, Rousidis D, Tjortjis C (2020) Forecasting and prevention mechanisms using social media in health care. In: *Advanced computational intelligence in healthcare*, vol 7. Springer, Berlin, pp 121–137
26. Regulation GDP (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. *Official J Euro Union (OJ)* 59(1–88):294
27. GraphXR. KINEVIZ. <https://www.kineviz.com/>

# Tackling Counterfeit Medicine Through an Automated System Integrated with QR Code



A. M. Chandrashekhar, J. Naveen, S. Chethana, and Srihari Charith

**Abstract** Adulteration and the introduction of counterfeit medicine in the pharmaceutical market is an increasing problem in our country. Official reports from WHO states that nearly 35% of the fake drugs in the world are being imported from India making up to a 4000 crore drug mafia market. The main reason is human intervention at every stage of the stringent process in the current system. This induces inevitable human error which is taken advantage of. The absence of an automated system is the main and the biggest drawback of the currently implemented system. The paper proposes an automated system that provides a unique recognition mechanism by generating machine-readable code for each entity. This generation of the machine-readable code is backed up by encryption and hashing and hence cannot be replicated by a third party.

**Keywords** QR code · Metadata · Counterfeit · Hash value · Leaflet

## 1 Introduction

Absence of a stringent system in the pharmaceutical industry to ensure the legitimacy of the medicine and presence of human intervention at every stage leading to an increase in counterfeit medicine. This problem eradication is considered to be the

---

A. M. Chandrashekhar · J. Naveen (✉) · S. Chethana · S. Charith  
Department of Computer Science and Engineering, JSS Science and Technology University  
(Formerly SJCE-Mysore), Mysore, Karnataka 570006, India  
e-mail: [naveen.jayanna@gmail.com](mailto:naveen.jayanna@gmail.com)

A. M. Chandrashekhar  
e-mail: [amc@sjce.ac.in](mailto:amc@sjce.ac.in)

S. Chethana  
e-mail: [chethana.bilugali@gmail.com](mailto:chethana.bilugali@gmail.com)

S. Charith  
e-mail: [shriharicharithrao@gmail.com](mailto:shriharicharithrao@gmail.com)

problem in focus and to ensure the exhibition of the metadata corresponding to a distinct leaflet of a drug to certify the clarity and to tackle counterfeit medicine and adulteration. The metadata represents the manufacturing date, expiry date, ingredients present, price, etc. This paper proposes a secure and controlled way by:

- Generating machine-readable code for each leaflet.
- Uniquely recognize each of the leaflets.
- Mobile App to retrieve the stored information regarding the leaflet [1].
- Creating cloud storage and retrieval mechanism for the metadata [2].

## 2 Literature

Zhang et al. [3] sets out a novel algorithm to cover and secure the multichannel visual masking QR code. The appearance of the QR code is significantly modified with the algorithm, while preserving the original secret knowledge. Unauthorized users of standard QR code reader can not extract any information from secure QR code. A truth table based decoder is designed for approved users and works with the standard QR code reader. Extensive tests are being performed to determine this method's robustness and effectiveness.

Palshikar et al. [4]: highlights the concept of developing a vascular healthcare network capable of providing options such as clinical management, medical records, wellness prediction, and producing QR code for each medical according to their updated wellness information. The paper proposes to reduce the amount of human interaction needed during a regular health checkup. The limitation of this paper is the lack of security in maintaining the information of the patient stored in the QR code which can be accessed by any scanning application.

Abdulhakeem et al. [5] shows how the use of QR Code technology can help improve user experience, particularly in the retail sector.

Balaji et al. [6] claimed that medical and health-related data were exchanged between two sides: one-part hospital management to be viewed as an admin part and then as a consumer part. This project has the main benefit of creating a safe transaction between two sides, versatile, easy to access, and paper-free coverage.

### 2.1 Existing System

The current system consists of intervention and interaction by humans at each and every stage i.e., right from the manufacturing of the medicine at every organization's registered manufacturing units till the delivery of the packaged medicine to the pharmacy. This human intervention at every stage induces the lack of security measures and human-induced errors are inevitable. This, in turn, leads to the adulteration and the intrusion of counterfeit medicine within the system. Taking into consideration

the health [7] and safety of the consumers this system fails to provide the sense of security which in-turn raises the requirement of an authenticated mechanism.

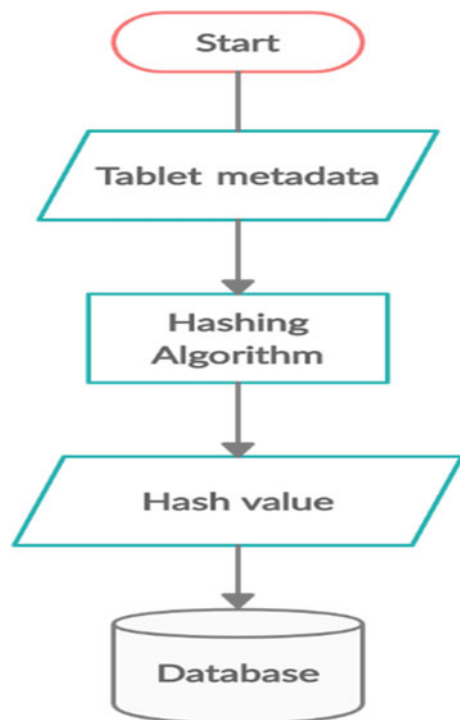
### 3 Proposed Work

The size of each leaflet of the tablet, in general, is very small while the metadata related to it is considerably large. To impart this onto each leaflet hashing is introduced which reduces the metadata into a unique string of characters which is used as an input in the generation of QR code compact in size.

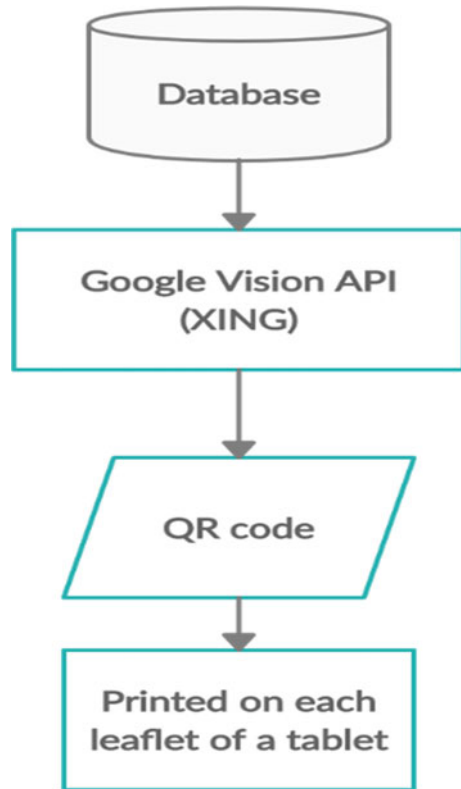
Introduction of the machine-readable code on each leaflet of the tablet enables storing more information in a reduced space. As shown in Fig. 1 the metadata of a particular tablet is converted into its corresponding hash value by using the hashing algorithm and is pushed into the cloud database.

This system also provides an automated mechanism which is created with no human intervention which provides security, transparency, and more feasible (cheap, less time-consuming, and reliable).

**Fig. 1** Storage of hashed value



**Fig. 2** Generation of QR code



The hashed values present in the cloud database acts as input to the creation of the QR code with the help of Google Cloud Vision API and are printed on each corresponding leaflet of the tablet and this flow is shown in Fig. 2.

Integration of the Android Application so that the data is highly accessible even by naive end-users. The main focus of this paper is to introduce a safe and secure mechanism to reduce malfunctioning in the pharmaceutical industry.

The QR code on the leaflet is scanned using a custom-built scanner to check the legitimacy of the medicines and the relative metadata is displayed in the scanner as depicted in Fig. 3.

### Flowcharts

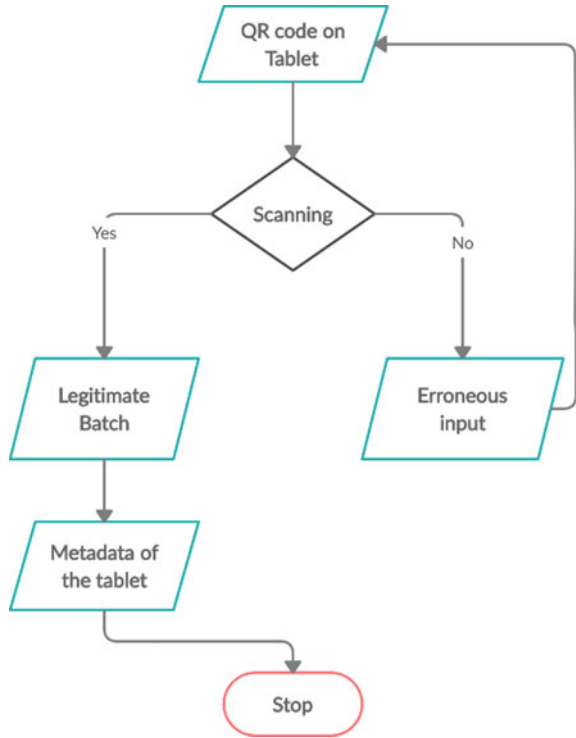
See Figs. 1, 2 and 3.

### The Main Actors

There are two actors in the Android Application. They are—Common public (End-user) and Pharmacy.

There are three actors in the Web Application. They are—Manufacturer, Warehouse, and Pharmacy.

Fig. 3 Retrieval of metadata



Inter-dependency is provided between any actors in the system and thus making the stated system inherently more secure, provides a simple but effective user interface to generate and handle the medicines produced.

### 3.1 Tools and Technologies Used

The development of web application was initiated with Dotnet framework which includes a large class library known as Framework Class Library (FCL) and provides language interoperability (each language can use code written in other languages) across several programming languages.

The development of the Android application [8] started with the Android Studio which provides a Linux Kernel with a level of abstraction between the device and hardware along with that it also provides all the necessary libraries required for the application. It provides an Android Runtime which makes use of Linux core features like memory management and multi-threading, which is intrinsic in the Java language also comprising of the application layer which is the utmost layer. This application is coded in Java where the implementation is carried out with the help of Java Development Kit (JDK) using the software tools given by Java Runtime



Environment (JRE) combining Java Virtual Machine (JVM) and all the necessary libraries where JRE is a part of JDK.

The connection of these applications is interfaced with each other using cloud mainly for data protection, scalability, and flexibility. Amazon Web Services (AWS) [9] is used to create an instance (EC2) [10, 11] which acts as a backend for both the applications. To handle the requests from the android a tomcat server is set up in the EC2 instance [12] which in-turn connects the database.

### 3.2 System Design

System design is the process of determining a system's architecture, components, modules, interfaces, and data to meet defined demands. System design may be seen as applying the system principle to the production of goods.

The manufacturer plays an important role in generating bulk QR code as per requirement, accepts the request for the generation of the tablet, and also verifies the total number of tablets present.

The warehouse plays an important role in mediating between the manufacturer and the pharmacy by accepting the request from the pharmacy and sending those requests to the manufacturer.

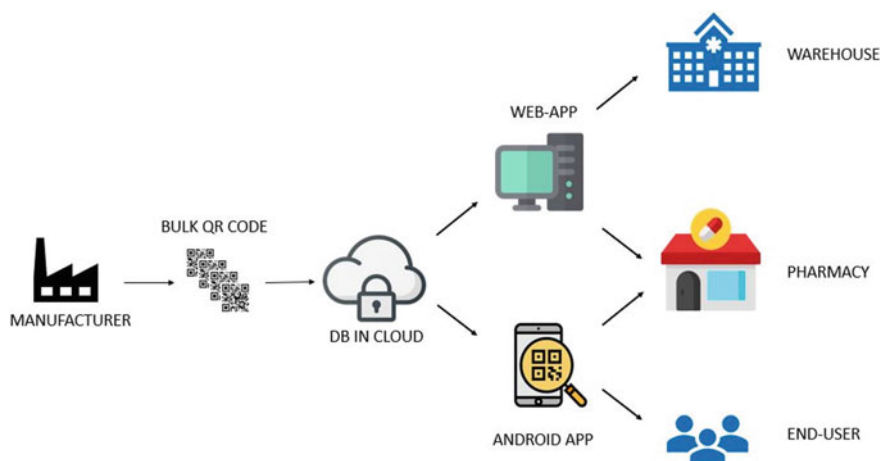
The pharmacy is the face of the customer. He sends the requests to the warehouse based on the requirement from the end-user.

The 3 actors are connected to the cloud and they access their respective data and views by providing unique credentials.

There are 2 different custom-built QR code applications for both end-user and pharmacy holders according to their requirements. The scanner for the end-user encompasses just scanning of the QR code on the leaflet of the tablet and retrieve the necessary information which is a metadata of the tablet that is manufactured date, expiry date, tablet name and also a tinyURL which leads to the tablet's manufacturing company to obtain the extra information regarding the tablet.

The other scanner which is for the pharmacy requires the pharmacy to register and login to use the app to acquire legitimacy of the application and the entire architecture is shown in Fig. 4.

Connection of the android module to the cloud is done by setting up a tomcat server in the cloud which helps to run the queries from the android in the database of the cloud taking into account the IP address of that local machine in the cloud database. With the help of JavaServer Pages, the connection is established between localhost and the target database along with all the queries required for that connection.



**Fig. 4** The architecture of the proposed system

## 4 Experimentation

### Web Application

The job of the manufacturer is to generate the QR codes in bulk which are required by the commoners in the process where the pitstops are warehouses and pharmacies. Taking an instant that a particular company generates 50 different types of tablets having all the metadata mentioned earlier in this paper and making a leaflet embedded with QR codes.

The number of tablets left and the number of them requested will be given. The manufacturer can generate tablets and keep track of the remaining tablets while responding to the requests from the respective warehouse.

The work associated with the warehouse is briefed as a mediator between the manufacturer and the pharmacy where it takes the tablets generated by the manufacturer and ships it to the pharmacies destined.

The warehouse has the necessary details of the tablets which are then manufactured and expiry date and rest of the metadata his hidden from his view to avoid counterfeit.

The warehouse can request the manufacturer for the tablets which are not in stock and are requested by the pharmacy.

Pharmacy is the face for the customer as the customer interacts with the pharmacists alone.

The requirements of the customers are understood by the pharmacies and it is diverted to warehouse and manufacturer alongside.

Pharmacy requests the warehouse for the tablets in need and the warehouse responds to it if it is in stock else it requests the manufacturer for the tablets.

### Android Application

The application of the QR code scanner which is developed has 2 different interfaces:

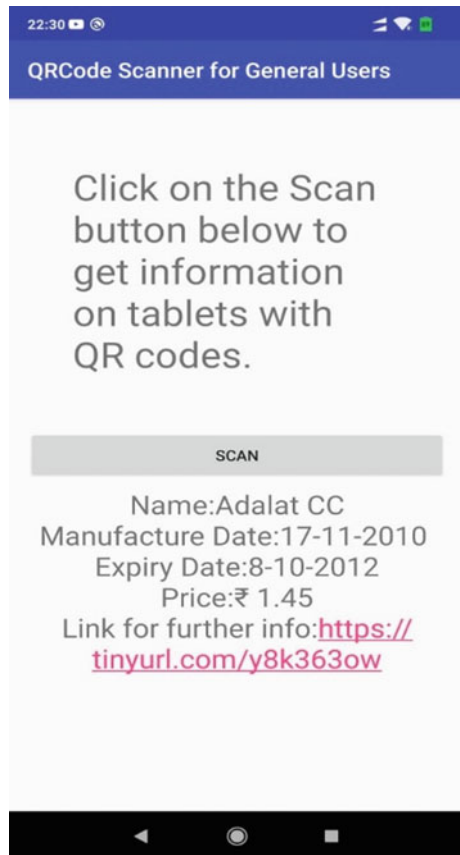
- End-User (The common public)
- Pharmacy

To elaborate the actors and their accessibilities to the application is as follows:

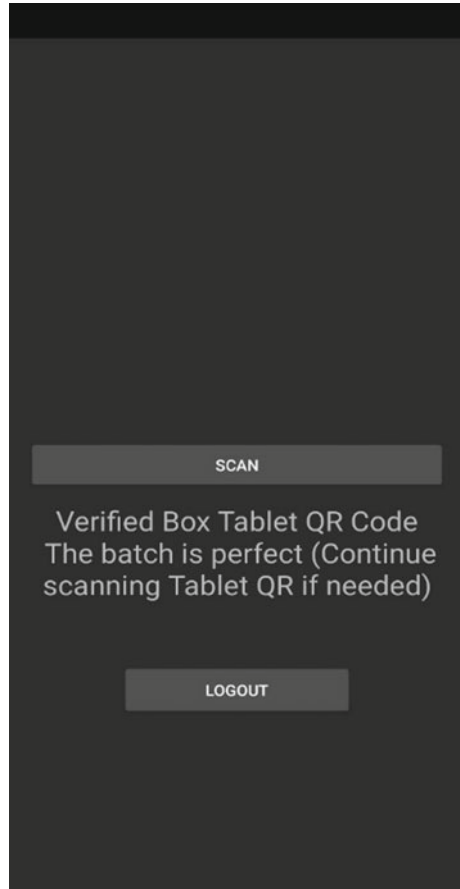
The end-user has a very simple interface for scanning the QR code on the leaflet of the tablet to retrieve the information related to tablet which will be in the form of manufacturing date, expiry date, price, present along with the tiny URL which leads to the website for detailed information regarding the usage and side effects of the tablet which has been scanned by the user as shown in Fig. 5.

The pharmacy has a very important role to play to avoid the counterfeit medicine and adulteration in the market and the final output of a perfect batch scanned by using

Fig. 5 Interface for end-user



**Fig. 6** Output for a legitimate batch



the custom android application is shown in Fig. 6. This is achieved by multi-factor authentication, i.e., Batch, Box, and Tablet.

The manufacturing industry generates tablets in bulk embedded with the unique QR code containing all necessary information regarding the tablet and these tablets are grouped to form boxes which will be shipped to the warehouse.

These boxes will be containing a QR code which is the assimilation of all the hash values of the QR codes present on the leaflet of the tablet which it contains.

When the pharmacy requests for the tablets, these boxes are grouped into a batch which accumulates the QR code which is in-turn generated using the hash values of the QR code present on all the boxes of that batch.

## 4.1 Testing and Results

Testing the software is the process of validating and verifying a software program. The errors are to be identified to fix those errors. Thus, the main objective of software testing is to maintain and deliver a quality product to the client.

### Unit Testing

#### *Unit Testing of Unique QR Code*

The strength of the generation of the QR code lies in the creation of the corresponding hash value. The input for the generation of the QR code is altered to the bare minimum that is, a single character is changed or added which in turn generates a new hash value thus generating a new QR code. This is achieved by using a checksum mechanism to generate each hash value.

The input is parsed into token where the token size is one character and each character's ASCII is used as the input to the checksum. This uniqueness is repeatedly tested with the custom-built android application.

#### *Unit Testing in Android Application*

The login part of the developed application connected to the cloud is unit tested by setting up a Tomcat server running in the Eclipse environment. Camera API integrated with the Google vision API package is mainly used to scan the QR code to retrieve the hash value assigned to it.

#### *Unit Testing in Web Application*

The manufacturer has 2 user interfaces to interact with the warehouse to check if the request has been granted and the change is reflected in the database and to generate the QR code which is verified with those dumped in the cloud database. The manufacturer also has a JAVA UI made with Swing to fabricate the QR code from the hash values obtained with the help of checksum.

The warehouse has 3 interfaces, to interact with both, the manufacturer and the pharmacy. Starting with the first component the warehouse interacts with the manufacturer requesting for the tablets. This is unit tested by checking whether the number of tablets requested is present in the warehouse and if not present the request is done.

The testing of the latter i.e., with the pharmacy, is done by the warehouse indulging in communication by granting the request by checking the number of tablets present in its warehouse. It also has an interface to check the status of both and it is tested against the database for verification.

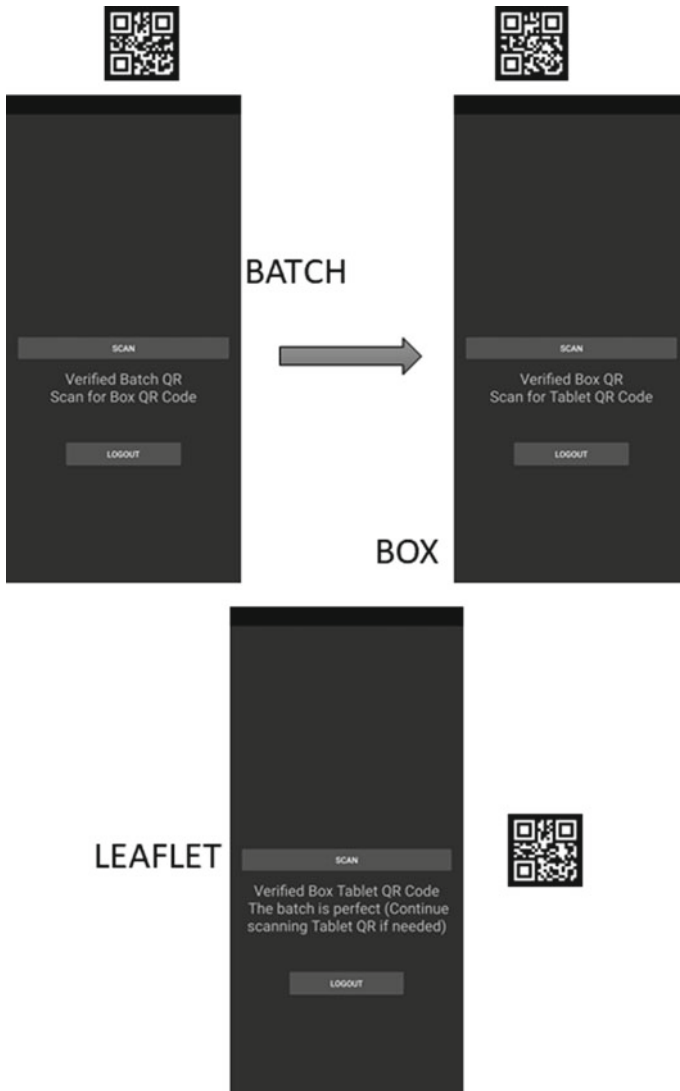
The pharmacy has a simple user interface for requesting the necessary tablets from the warehouse which is done primarily by checking it in its pharmacy and thus unit testing is accomplished.

### Integration Testing

The integration testing in the applications is as follows:

*Integration Testing in Android*

The login and the scanning module in the QR code application developed is integrated and tested by combining the JSP of both the modules and their working is tested for the previous login and scanning test cases. Once all the modules are tested, a valid scanning process is done to check the legitimacy of the application and the process is shown in Fig. 7.



**Fig. 7** A valid scanning process

### *Integration Testing in Web Application*

The 3 actors in the web application are given an increased medium of authentication by having unique credentials and the communication between them is unit tested and integration of all the 3 is successful as imparting between them is tested.

### *Integration Testing with Cloud*

The 2 applications i.e., the android and the web are integrated along with the cloud and the testing between them is considered to be successful as all the modules in the web and android have been unit and integration tested, so the integration testing with the cloud was relatively easy.

## **5 Conclusion**

This paper can be concluded by stating that, QR codes are quickly arriving at high degrees of acceptance. Thus, by creating a secure and automated system for reducing the counterfeit medicine in the market hence making the end-user self-reliant. The strength of the proposed system is an enabled sense of security with the inculcation of hashing as it hides the original metadata from unauthorized users and also there is minimal consumption of system resources. The limitation of this paper corresponds to the reduction of the size of the QR code beyond a specific threshold (tested value =  $0.4 \text{ cm}^2$ ). As the size of the QR code decreases the probability of it being detected also decreases as they are directly proportional to each other. This threshold value can be further reduced by improvement in the camera hardware technology of smartphones.

### **Future Work**

The future adoption of this concept is to generate bulk QR code from a stream of continuous input from multiple sources arriving at a different speed and carrying different types of data (alpha-numerals, bytes, image, audio, etc.) in Real-time.

This android application might help in storing and retrieving data and images related [13] to medicines [14] and their metadata as they are or in the form of a QR code [15] that can be stored and retrieved by using an application similar to the one that has been created and to scale it to the extent of the requirement.

This concept of encryption [16] can be extended to sensitive data [17] such as medical reports and prescription [18] issued and these can be easily transferred between the actors in a secure way [19].

## **References**

1. Czuszynski K, Ruminski J (2014) Interaction with medical data using QR-codes. In: 2014 7th international conference on human system interactions (HSI), Costa da Caparica, pp 182–187. <https://doi.org/10.1109/HSI.2014.6860471>

2. Singh M (2015) Study on cloud computing and cloud database. In: International conference on computing, communication and automation, Noida, pp 708–713. <https://doi.org/10.1109/CCAA.2015.7148468>
3. Zhang J, Li D, Jia J, Sun W, Zhai G (2019) Protection and hiding algorithm of QR code based on multi-channel visual masking. In: 2019 IEEE visual communications and image processing (VCIP). Sydney, Australia, pp 1–4. <https://doi.org/10.1109/VCIP47243.2019.8966044>
4. Palshikar MN, Mane PS, Dhore AS, Kalegore BP, Shinde S (2017) QR code based healthcare system 4(12). e-ISSN: 2348-4470, print-ISSN: 2348-6406
5. Abdulhakeem W, Dugga S (2014) Enhancing user experience using mobile QR-code application. *Int J Comput Inf Technol*
6. Balaji S, Pughazendi N, Praveenkumar SE, Vishal V, Vignesh R (2018) E-medical application using QR code with OTP generation 05(04)
7. Tychalas D, Kakarountas A (2010) Planning and development of an electronic health record client based on the android platform. In: 14th Panhellenic conference on informatics, pp 3–6
8. Jagodić D, Vujičić D, Randić S (2015) Android system for identification of objects based on QR code. In: 2015 23rd telecommunications forum telfor (TELFOR). Belgrade, pp 922–925. <https://doi.org/10.1109/TELFOR.2015.7377616>
9. Mukherjee S (2019) Benefits of AWS in modern cloud. <https://doi.org/10.5281/zenodo.2587217>; Kulkarni G, Sutar R, Gambhir J Cloud computing-Infrastructure as service amazon EC2. *Int J Eng Res Appl (IJERA)* 2(1):117–125
10. Ahmed ST, Chandrashekhar AM, Rahul N (2015) Analysis of security threats to database storage systems. *Int J Adv Res Data Min Cloud Comput (IJARDC)* 3(5)
11. Chandrasekhar AM, Revapgol J, Pattanashetti V (2016) Security issues of big data in networking. *Int J Sci Res Sci Eng Technol (IJSRSET)* 2(1)
12. Koushik P, Takkalakaki J, Chandrashekhar AM (2015) Information security threats, awareness and cognizance. *Int J Tech Res Eng (IJTRE)* 2(9)
13. Chandy A (2019) A review on iot based medical imaging technology for healthcare applications. *J Innov Image Process (JIIP)* 1(01):51–60
14. Deepika P, Sushanth B, Kumar T, Vignesh M Emergency information access using QR code technology in medical field 1(1). ISSN: 2456-3307
15. Patel J, Bhat A, Chavada K A review paper on QR code based android app for healthcare. *Int Res J Eng Technol (IRJET)* 02(07)
16. Arpitha NG, Chandrashekhar AM (2016) Efficient data accessibility in cloud with privacy and authenticity using key aggregation cryptosystem. *Int J Technol Res Eng (IJTRE)* 3(5)
17. Prashanth GM, Chandrashekhar AM, Bulla A (2015) Secured infrastructure for multiple group communication. *Int J Adv Res Inf Commun Eng (IJARICE)* 3(5)
18. Dube S, Ndlovu S, Nyathi T, Sibanda K QR code based patient medical health records transmission: zimbabwean case. National University of Science and Technology, Bulawayo, Bulawayo Province, Zimbabwe
19. Chandrashekhar AM, Sahana K, Yashaswini K (2016) Securing cloud environment using firewall and VPN. *Int J Adv Res Comput Sci Softw Eng (IJARCSSE)* 6(1)



# Efficient Gs-IBE for Collusion Resistance on Multi-handler Cloud Environment Using Attribute-Modulator (AM)



J. Priyanka and M. Ramakrishnan

**Abstract** In the digital world, adequate data has been stored and accessed under a cloud by using various users. The system emanates beneath the user management to resolve the security and privacy risks. In multi-user setup, the attendance to add operations carried out to sign-up and user cancellation from cloud to yield the risk overload. Further to avoid this problem, the paper castoff a Generalized-signcrypt identity-based encryption (Gs-IBE) with a random modulator that has been experimented. The conditional variable  $Am(r)$  been castoff to recognize the collusion in the random value. The  $Am(r)$  perform the distinct job in the random bit generation of each random element selection.

**Keywords** Signcryption · Generalized signcryption · Collusion resistance · Revocation · Random bit · Attribute modulator

## 1 Introduction

Data storage has been most operative in the cloud management, a large volume of data has outsourced/stored, and access under the cloud. These data are used by various users involved in a group activity, where users have been managed by the group manager. These users and data are secured by the default security mechanism provided by the cloud provider [1]. The cloud comes under 3 categories namely private, public, and hybrid. The risk has been reduced in the private cloud when compared to the public and hybrid since the management and data accessing has been under the control of identical [2]. Greatest users derive under the categories of

---

J. Priyanka (✉) · M. Ramakrishnan  
Department of Computer Applications School of Information Technology, Madurai Kamaraj  
University, Madurai, India  
e-mail: [priyankajaya86@gmail.com](mailto:priyankajaya86@gmail.com)

M. Ramakrishnan  
e-mail: [ramkrishod@gmail.com](mailto:ramkrishod@gmail.com)

public and hybrid, the security has been a prime challenge because of the diverse regulators [3]. There have been several authorities work to ensure security, as well as a variety of policy templates discovery, has been developed for cloud data security. Nevertheless, security and privacy issues are there in the cloud data. One of the silent issues of the cloud operation involved user revocation, every user has involved some of the attributes used for security mechanism [4]. When user revocation occurs, the security attribute of the particular user may affect the group user's security. Other issues involved in the multi-user cloud is collusion resistance attack [5]. The cloud might use strong collusion avoidance mechanisms like hash function, integer factorization, and discrete logarithm.

In this paper, the user revocation and collusion resistance have been concentrated by using Gs-IBE. The random bit has been generated with the conditional attribute of the attribute modulator (AM) [6]. Generalized signcryption scheme proposed by Han [7] on no account of formal proof, far ahead An Wang [8] established generalized signcryption through formal proof. All data does not require both confidentiality and authenticity but some data essentially accomplish the together. In generalized signcryption, the data has been stored under signcryption, encryption, and signature. In this paper, Id-based generalized signcryption scheme has been proposed, these methods have been combination of Id-based encryption [9] and generalized signcryption, first Id-based signcryption has been introduced by Malone-Lee [10] after the first implementation there has been several literature proposed Id-based signcryption [11–15] and generalized signcryption [16–19]. The proposed scheme can also prove the security of confidentiality under the DBDH and unforgeability under the CDH.

## 2 Preliminary

**Definition 1: Bilinear Map** Let  $G_1$  and  $G_2$  be a two multiplicative cyclic group with the prime order of  $q$  and  $g$  be a generator, the mapping function  $e: G_1 * G_1 = G_2$  said to be bilinear paring with the following satisfying properties.

1. Bilinear: for all  $r, s \in G_1$  and  $a, b \in \mathbb{Z}_q^*$ ,  $e(r^a, s^b) = e(r, s)^{ab}$
2. Non-degeneracy  $e(g, g) \neq 1$
3. Computable.

The algorithm has been efficiently computed  $e(r, s)$  for all  $r, s \in G_1$ .

### 2.1 Complexity Assumption

**Definition 2: Decisional Diffie-Hellman Assumption (DDH)** The element,  $a, b, c \in \{0, 1, 2, \dots, q-1\}$  the  $g$  be the generator, value of  $r \leftarrow g^a$  and  $s \leftarrow g^b$  generated

by the poly ( $x$ ) time algorithm and the value  $t$  has been decided from  $b \leftarrow \{0, 1\}$ , if  $b = 0, t \leftarrow g^c$  (or) if  $b = 1$  if  $t \leftarrow g^{ab}$

$$\Pr[ad(b = 1)|b = 1] - \Pr[ad(b = 1)|b = 0] \quad (1)$$

The DDH has been hard, ( $t, \epsilon$ ) no algorithm able to solve with time ( $t$ ) and specified probability ( $\epsilon$ ).

**Definition 3: Computational Diffie-Hellman Assumption (CDH)** The element  $a, b \in \{0, 1, 2 \dots q - 1\}$  the  $g$  be the generator, value  $r \leftarrow g^a$  and  $s \leftarrow g^b$  generated by the poly ( $x$ ) time algorithm and then compute  $g^{ab}$

$$\Pr[ad(c = 1)] \text{ if and only if } c = g^{ab} \quad (2)$$

The CDH has been hard, ( $t, \epsilon$ ) no algorithm able to solve with time ( $t$ ) and specified probability ( $\epsilon$ ).

## 2.2 Generalized Id-Based Signcryption

A Generalized id based signcryption scheme consists of four algorithms: Setup; Extract; Generalized id based signcrypt and generalized id based unsigncrypt.

**Setup:** Algorithm generates the systems public parameter and master key from the security parameter.

**Extract:** Algorithm will extract the identity of the user; it generates the private-key with the identity and master-key.

**Generalized id-based signcrypt (GIBSC):** with sender identity and private-key as well as the receiver public-key and the message, the algorithm run by the sender and output the ciphertext.

**Generalized id-based unsigncrypt (GIBUSC):** with sender identity and receiver identity as well as the receiver private-key and the ciphertext, the algorithm runs from the receiver side and output the original message or thread if ciphertext has been invalid from the sender and receiver communication.

## 2.3 Security Model

Describe the security model of generalized id based signcryption that has been properly modeled by queries issued by the  $\mathbb{A}$ . The  $\mathbb{A}$  have been issuing the following queries:

- **Private-Key-Extract:**  $\mathbb{A}$  has been submitting the identity to the  $\mathbb{C}$ . The  $\mathbb{C}$  has been responding to the adversary with the private-key of the identity.
- **Signcryption:**  $\mathbb{A}$  has been submitting a sender's and receiver's identities and a message to the  $\mathbb{C}$ , and the  $\mathbb{C}$  has been responding with the ciphertext under the sender's private-key and the receiver's public-key.
- **Unsigncryption:**  $\mathbb{A}$  has been submitting a ciphertext and a receiver's identity to the  $\mathbb{C}$ , and the  $\mathbb{C}$  has been decrypting the ciphertext under the private key of the receiver and verifies the resulting decryption has been valid message and signature pair under the public-key of the decrypted identity. Then, the  $\mathbb{C}$  returns the message.

### 2.3.1 Confidentiality

The proposed model GIBSC security of confidentiality has been indistinguishability of encryptions under adaptive chosen-ciphertext attack (IND-CCA2).

**Game 1:** consider that the following game played by the challenger  $\mathbb{C}$  and the adversary  $\mathbb{A}$ .

**Initialization:** The  $\mathbb{C}$  run the Setup algorithm for the generation of system public parameters  $Sp$  and security-key  $s$  with the input of security parameter  $k$ . Then  $\mathbb{C}$  sends the  $Sp$  to the  $\mathbb{A}$  and preserves the  $s$  secret.

**Step 1:** The adversary  $\mathbb{A}$  makes the above queries adaptively.

**Private key extract Query:**  $\mathbb{A}$  sends the  $Id_d$  to the  $\mathbb{C}$  for the private-key, then the  $\mathbb{C}$  reply the  $D_{id}$  as the private-key corresponding to the  $Id_d$ .

**Step 2:** After the queries, the  $\mathbb{A}$  has chosen the message  $m$  and the challenged two identities  $Id_A, Id_B$ , and submit to the  $\mathbb{C}$ .

**GIBSC Query:** The  $\mathbb{C}$  run the encryption algorithm with message  $m1$  and the identity  $Id_A, Id_B$  and returns the ciphertext to the  $\mathbb{A}$

**Step 3:**  $\mathbb{A}$  submit the identity  $Id_A, Id_B$  and the ciphertext  $\sigma$  to the  $\mathbb{C}$

**GIBUSC Query:** The  $\mathbb{C}$  run the Decryption algorithm with ciphertext  $\sigma$  and the identity  $Id_A, Id_B$  and returns the message  $m$  to the  $\mathbb{A}$

**Challenge:** The  $\mathbb{A}$  submits the two equal length message  $m1$  and  $m2$  and challenged two identities  $Id_A \neq Id_{\square}, Id_B \neq Id_{\square}$  to the  $\mathbb{C}$ . Then  $\mathbb{C}$  choose random bit  $b$  and encrypted message and send to the ciphertext  $\sigma$  to  $\mathbb{A}$

**Guess:** The bit  $b'$  produced by the adversary  $\mathbb{A}$  has been  $b' = b$  then  $\mathbb{A}$  wins the game.

The advantage of adversary  $\mathbf{A}$  has been defined as  $\text{Adv}_{\mathbf{A}}^{\text{IND-CCA2}} = |2 \Pr [b' = b] - 1|$ .

**Definition 4** The GIBSC has been said to be indistinguishability against the Chosen-ciphertext Attack (IND-CCA2) if no polynomially bounded adversary  $\mathbf{A}$  has a non-negligible advantage of winning the game 1.

### 2.3.2 Unforgeability

The proposed model GIBSC security of signature unforgeability has been said to be Existential Unforgeability under Chosen Message Attack (EUF-CMA).

**Game 2:** consider that the following game played between the challenger  $\mathbf{C}$  and the adversary  $\mathbf{A}$ .

**Initialization:** The  $\mathbf{C}$  runs the Setup algorithm for the generation of system public parameters  $\text{Sp}$  and security key  $s$  with the input of security parameter  $k$ . Then  $\mathbf{C}$  sends the  $\text{Sp}$  to the  $\mathbf{A}$  and preserves the  $s$  secret.

**Step 1:** The adversary  $\mathbf{A}$  makes the above quires adaptively like game 1.

**Forgery:** The adversary  $\mathbf{A}$  produces the two challenged Identities  $\text{Id}_A \neq \text{Id}_B$ ,  $\text{Id}_B \neq \text{Id}$ , and the signature  $\sigma$ . The adversary wins the game if  $\sigma$  has been a valid signature on  $m$ ,  $\text{Id}_A$ .

The advantage of adversary  $\mathbf{A}$  has been defined as  $\text{Adv}_{\mathbf{A}}^{\text{EUF-CMA}} = |\Pr[\mathbf{A} \text{ wins}]|$

**Definition 5** The GIBSC has been said to be Existential Unforgeability under Chosen Message Attack (EUF-CMA), if no polynomially bounded adversary  $\mathbf{A}$  has a non-negligible advantage of winning the game 2.

## 3 Implementation of Our Approach

The generalized id based signcrypton (GIBSC) are described as four algorithms:

- **Setup**
- **Extract**
- **Generalized id-based signcrypt (GIBSC)**
- **Generalized id-based Unsigncrypt (GIBUSC).**

In this setup and Extract is the same as the Boyen scheme [12].

PKG	Public key generator
$\mathbf{A}$	Adversary
$\bar{\mathbf{A}}$	Authority
$\mathbf{C}$	Challenger

(continued)

(continued)

PKG	Public key generator
Params	Public parameter
$s$	Random secret
$\sigma$	cipher
DV	Dimensional vector
Am	Attribute modulator

### Setup

Given security parameter  $\varphi$  the PKG choose two groups  $G_a$  and  $G_b$  of prime order  $p$ , a random generator  $g$  of  $G_a$ . the bilinear mapping function  $e: G_a \times G_a \rightarrow G_b$ . The PKG chooses random secret  $s$  and set  $s \in Z_p^*$ . The PKG computes  $g_1 = g^s$  and randomly select  $g_2 \in G_a$  computes  $Z = e(g_1, g_2)$ . The value  $r, s \in G_a$  has been chosen and the vector element  $r = \{r_i\}, s = \{s_i\}$  of the length  $n_r, m_s$  has been picked whose entries are random element from  $G_a$ . The has function has been defined by  $H1: G_b \rightarrow \{0, 1\}^{nm}$  and  $H2: \{0, 1\}^* \rightarrow \{0, 1\}^{nm}$ . The PKG publish System public parameter  $\text{params} = \{G_a, G_b, e, g, g_1, g_2, r^1, s^1, Z, r, s, H1, H2\}$ .

Let  $\text{fun}(\text{Id})$  be the special function where  $\text{Id} \in \{0, 1\}$ . If identity has been vacant  $\text{fun}(\text{Id}) = 0$  otherwise  $\text{fun}(\text{Id}) = 1$ .

Let  $\text{Am}(r)$  be the special function where  $r$  if the value random number does not match previous value  $\text{Am}(r) = 0$  otherwise  $\text{Am}(r) = 1$ .

### Extract

Let  $\text{Id}$  be an identity of the length  $n_u$  and  $\text{Id}[i]$  be the  $i$ th bit of  $\text{Id}$ . define  $U_{\text{Id}} \subseteq \{1, 2, 3, \dots, n_r\}$  be the set of indices  $i$  such that  $\text{Id}[i] = 1$ . The private key of  $d_{\text{Id}}$  has been calculated by the PKG with randomly chosen element  $r_{\text{Id}} \in Z_p^*$  and the function  $\text{Am}(r)$  has been executed.

The PKG calculate the private key as follows:

$$d_{\text{Id}} = (d_{\text{Id}1}, d_{\text{Id}2}) = \left( g_2^s \left( \prod_{i \in U_{\text{Id}}} u_i \right) r_{\text{Id}} g^{r_{\text{Id}}} \right) \quad (3)$$

The sender  $A$  with the identity  $\text{Id}_A$  and the receiver  $B$  with identity  $\text{Id}_B$ , random value the PKG compute the private key as follows:

$$d_A = (d_{\text{Id}_A1}, d_{\text{Id}_A2}) = \left( g_2^s \left( \prod_{i \in U_{\text{Id}_A}} u_i \right) r_{\text{Id}_A} g^{r_{\text{Id}_A}} \right) \quad (4)$$

$$d_B = (d_{\text{Id}_B1}, d_{\text{Id}_B2}) = \left( g_2^s \left( \prod_{i \in U_{\text{Id}_B}} u_i \right) r_{\text{Id}_B} g^{r_{\text{Id}_B}} \right) \quad (5)$$

### GIBSC

The GIBSC works under 3 cases.

Case 1: Encryption

Case 2: Signature

Case 3: Signcryption.

The PKG choose random value  $R \in Z_p^*$  has been executed by the function  $\text{Am}(r)$  and compute  $\sigma 1 = g^R$  and  $c = m \oplus H(w)$  where  $w$  has been calculated as step 1 in each case.

### Case 1: Encryption

1. compute  $w = s^R$
2. compute  $\sigma 2 = d_{A2}$
3. compute  $\sigma 3 = \left( \prod_{i \in U_{\text{Id}_B}} u_i \right)^R$
4. compute  $X = H_s(m, \sigma 1, \sigma 2, \sigma 3, w)$  where  $X$  be the bit string  $X[j]$  be the  $j$ th bit of  $X$ . define  $M \subseteq \{1, 2, 3, \dots, n_s\}$  be the set of indices  $j$  such that  $X[j] = 1$ .
5. compute  $\sigma 4 = d_{A1} \cdot \sigma 3 \cdot \left( m^{\prod_{j \in M} m_j} \right)^R$
6. compute  $\sigma = (\sigma 1, \sigma 2, \sigma 3, \sigma 4, c)$ .

### Case 2: Signature

1. compute  $w = s^R \cdot \text{fun}(\text{Id}_B) = s^R$
2. compute  $\sigma 2 = d_{A2}(\text{Id}_A) = 1$
3. compute  $\sigma 3 = \left( \prod_{i \in U_{\text{Id}_B}} r_i \right)^{R \cdot \text{fun}(\text{Id}_B)} = 1$
4. compute  $X = H_m(m, \sigma 1, \sigma 2, \sigma 3, w)$  where  $X$  be the bit string  $X[j]$  be the  $j$ th bit of  $X$  Define  $M \subseteq \{1, 2, 3, \dots, n_s\}$  be the set of indices  $j$  such that  $X[j] = 1$ .
5. compute  $\sigma 4 = d_{A1} \cdot \sigma 3 \cdot \left( m^{\prod_{j \in M} m_j} \right)^R$
6. compute  $\sigma 5 = C1 \cdot p.l \oplus H(w)$
7. compute  $\sigma = (\sigma 1, \sigma 2, \sigma 3, \sigma 4, \sigma 5) = (\sigma 1, \sigma 2, \sigma 3, \sigma 4, m)$ .

### Case 3: Signcryption

1. compute  $w = s^R \cdot \text{fun}(\text{Id}_B) = s^R$
2. compute  $\sigma 2 = d_{A2}(\text{Id}_A) = 1$
3. compute  $\sigma 3 = \left( \prod_{i \in U_{\text{Id}_B}} r_i \right)^{R \cdot \text{fun}(\text{Id}_B)} = \left( \prod_{i \in U_{\text{Id}_B}} r_i \right)^R$
4. compute  $X = H_s(m, \sigma 1, \sigma 2, \sigma 3, w)$  where  $X$  be the bit string  $X[j]$  be the  $j$ th bit of  $X$ . define  $M \subseteq \{1, 2, 3, \dots, n_s\}$  be the set of indices  $j$  such that  $X[j] = 1$ .
5. compute  $\sigma 4 = d_{A1} \cdot \sigma 3 \cdot \left( m^{\prod_{j \in M} m_j} \right)^R = \left( \prod_{i \in U_{\text{Id}_B}} r_i \right)^R \cdot \left( m^{\prod_{j \in M} m_j} \right)^R$
6. compute  $\sigma = (\sigma 1, \sigma 2, \sigma 3, \sigma 4, c)$

## 4 Security Analysis

**Theorem 1: (Confidentiality)** *The proposed model GIBSC is secure, that the adversary IND-CCA2 has been able to distinguish two valid ciphertexts during the game 1 with running time  $\infty$ , assuming that the DBDHP is  $\epsilon'$  hard in  $G$ . in signcryption model*

$$\epsilon' = \frac{\epsilon}{O(q_{\text{pke}} + q_{\text{sc}} + q_{\text{usc}})(n_r + 1)(n_s + 1)}$$

where  $q_{\text{pke}}$  denoted as private-key extract,  $q_{\text{sc}}$  denoted as signcryption and  $q_{\text{usc}}$  denoted as unsigncryption. For encryption model

$$\epsilon' = \frac{\epsilon}{O(q_{\text{pke}} + q_{\text{enc}} + q_{\text{dec}})(n_r + 1)(n_s + 1)}.$$

**Proof** The algorithm  $\bar{\mathbb{A}}$  has been constructed to solve the DBDH problem, in case of the  $\mathbb{A}$  can break the scheme with the advantage of  $\epsilon$ . The algorithm can solve DBDH problem denoted by, the given  $(g, g^a, g^b, g^c, T)$  to be decided the  $T = e(g, g)^{abc}$ .

**Setup:** The  $\bar{\mathbb{A}}$  randomly chooses two integer  $lr$  and  $ls$  and set  $0 \leq lr \leq q$  and  $0 \leq ls \leq q$ . again  $\bar{\mathbb{A}}$  chooses two integers randomly for the length of  $nr$  and  $ns$  and set  $0 \leq kr \leq n_r$  and  $0 \leq ks \leq n_s$ . The four integers have been selected with DV as follows.

- $x' \in Zlr$  and  $nr - DV (x_1 \dots x_{nr}) \in Zlr$
- $y' \in Zls$  and  $ns - DV (y_1 \dots y_{ns}) \in Zls$
- $z' \in Zq$  and  $nr - DV (z_1 \dots z_{nr}) \in Zq$
- $\omega' \in Zq$  and  $nm - DV (\omega_1 \dots \omega_{ns}) \in Zq$ .

Functions

$$F(\text{Id}) = x' + \sum_{i \in r} xi - lrkr,$$

$$J(\text{Id}) = z' + \sum_{i \in r} zi$$

$$K(s) = x' + \sum_{i \in s} yi - lsk_s,$$

$$L(\text{Id}) = \phi' + \sum_{i \in s} \phi i$$

System Public parameter as follows.

The algorithm Construct the public parameter for any Id and  $X$ .

1.  $g_1 = g^a$  and  $g_2 = g^b$



2.  $r' = g2^{-lrkr+x'}g\omega$ , and  $s_i = g2^{yi}g\omega^i$  ( $1 \leq i \leq nr$ ) for any identity Id,

$$r' \prod_{i \in u_{Id}} r_i = g2^{K(Id)}g2^{L(Id)}$$

3.  $s' = g2^{-lks+y'}g^z$  and  $r_i = g2^{xi}g^{zi}$  ( $1 \leq i \leq ns$ ) for any identity X,

$$s' \prod_{i \in u_{Id}} s_i = g2^{K(X)}g2^{L(X)}.$$

All the system public parameter has been sent to  $\mathbb{A}$

**Phase 1** The Adversary  $\mathbb{A}$  request the polynomial bounded number of queries as follows.

### Extract Queries

The key has been extracted from the identity ID issued by the  $\mathbb{A}$ , the algorithm  $\bar{\mathbb{E}}$  has been computing the key as if and only if  $F(\text{Id}) \neq 0$ . The key has been computed as, The random value  $R$  has been selected  $R_{\text{Id}} \in \mathbb{Z}q^*$  and the private key is

$$d_{\text{Id}} = (d_{\text{Id}1}, d_{\text{Id}2}) = \left( g1^{\frac{J(\text{Id})}{F(\text{Id})}} \left( g2^{F(\text{Id})} g^{J(\text{Id})} \right)^{R_{\text{Id}}} g1^{\frac{1}{F(\text{Id})}} g^{R_{\text{Id}}} \right).$$

### Encryption Query

The  $\bar{\mathbb{E}}$  runs the encryption algorithm for the adversary  $\mathbb{A}$  request towards the plaintext  $m$  for the receiver  $\text{Id}_B$ .

Note: the adversary allowed to do encryption queries at any time.

### Decrypt Query:

The adversary  $\mathbb{A}$  request the decryption,  $\bar{\mathbb{E}}$  to do the decryption algorithm If and only If  $F(\text{Id}_B) \neq 0 \pmod{l}$  then  $\bar{\mathbb{E}}$  run private key extract query and run the encryption query for the adversary  $\mathbb{A}$  request towards the decryption.

Note: the adversary allowed to do decryption query for the ciphertext  $\sigma$  at any time.

### Signature Query

The  $\bar{\mathbb{E}}$  runs the signature algorithm for the adversary  $\mathbb{A}$  request towards the message  $m$  for the sender  $\text{Id}_A$ .  $\bar{\mathbb{E}}$  to do the signature algorithm if and only If  $F(\text{Id}_A) \neq 0 \pmod{l}$  then  $\bar{\mathbb{E}}$  run private key extract query and run the signature query for the adversary  $\mathbb{A}$  request.

### Verify Query

The adversary  $\mathbb{A}$  request the verifier query, the  $\bar{\mathbb{E}}$  runs the verifier algorithm for the adversary  $\mathbb{A}$  request towards the signature/message  $m$  for the receiver  $\text{Id}_B$ .

Note: the adversary allowed to do verifier query for the sign pair  $(m, \sigma)$  at any time.

### Signcrypt Query

The  $\bar{E}$  runs the signcryption algorithm for the adversary  $\mathbb{A}$  request towards the plaintext  $m$  to the sender identity  $\text{Id}_A$  and receiver  $\text{Id}_B$ .  $\bar{E}$  to do the signcrypt algorithm if and only If  $F(\text{Id}_A) \neq 0 \pmod{l}$  then  $\bar{E}$  run private key extract query and run the signcrypt query for the adversary  $\mathbb{A}$  request.

Note: the adversary allowed to do signcrypt query at any time.

### Unsigncrypt

The adversary  $\mathbb{A}$  request the unsigncryption,  $\bar{E}$  to do the unsigncryption if and only If  $F(\text{Id}_B) \neq 0 \pmod{l}$  then  $\bar{E}$ run private key extract query and run the signcryption query for the adversary  $\mathbb{A}$  request.

Note: the adversary allowed to do unsigncrypt query for the signcrypt text  $\sigma$  at any time.

## 5 Performance Analysis

### Computation Cost

Computation cost has been calculated by the number of exponentiations and paring operations used for each algorithm of proposed work. The representation of exponentiations denoted as  $e$  and paring have been denoted as  $p$ . The calculation has been evaluated for Generalized id-based signcrypt (GIBSC) and Generalized id-based unsigncrypt (GIBUSC). The computation cost has also be compared with other implementation.

Implantation scheme	GIBSC	GIBUSC
Lal et al. [16]	$6e + 1p$	$3p + 1e$
Kushwah et al. [17]	$4e$	$2p + 3e$
Yu et al. [18]	$4e + 1p$	$3e + 3p$
Xiaoqin et al. [19]	$6e$	$5p + 2e$
Ours	$7e$	$5p + 2e$

## 6 Conclusion

In this paper, efficient generalized-signcrypt identity-based encryption (Gs-IBE) with a random modulator has been experimented. The performance is analyzed with other previous generalized signcryption, where our scheme has been with relatively higher

cost among other schemes but the proposed scheme has been stronger among the collusion resistance using attribute modulate  $AM(r)$  function.

## References

1. Tian H, Nan F, Jiang H, Chang C-C, Ning J, Huang Y (2019) Public auditing for shared cloud data with efficient and secure group management. *Inf Sci* 472:107–125
2. Ari AAA, Ngangmo OK, Titouna C, Thiare O, Mohamadou A, Gueroui AM (2019) Enabling privacy and security in cloud of things: architecture, applications, security and privacy challenges. *Appl Comput Inform*
3. Priyanka J, Ramakrishna M (2020) Performance analysis of attribute based encryption and cloud health data security. In: 2020 4th international conference on intelligent computing and control systems (ICICCS), IEEE, pp 989–994
4. Thokchom S, Saikia DK (2020) Privacy preserving integrity checking of shared dynamic cloud data with user revocation. *J Inf Secur Appl* 50:102427
5. Patil SM, Purushothama BR (2020) Non-transitive and collusion resistant quorum controlled proxy re-encryption scheme for resource constrained networks. *J Inf Secur Appl* 50:102411
6. Han M, Kim Y (2017) Unpredictable 16 bits LFSR-based true random number generator. In: 2017 international SoC design conference (ISOCC). IEEE, pp 284–285
7. Zheng Y (1997) Digital signcryption or how to achieve cost (signature and encryption)  $\ll$  cost (signature) + cost (encryption). In: Annual international cryptology conference. Springer, Berlin, pp 165–179
8. An Wang X, Yang X, Zhang J (2010) Provable secure generalized signcryption. *J Comput* 5(5):807
9. Shamir A (1984) Identity-based cryptosystems and signature schemes. In: Workshop on the theory and application of cryptographic techniques. Springer, Berlin, pp 47–53
10. Malone-Lee J (2002) Identity-based signcryption. *IACR Cryptol ePrint Arch*:98
11. Barreto PSLM, Libert B, McCullagh N, Quisquater J-J (2005) Efficient and provably-secure identity-based signatures and signcryption from bilinear maps. In: International conference on the theory and application of cryptology and information security. Springer, Berlin, pp 515–532
12. Boyen X (2003) Multipurpose identity-based signcryption. In: Annual international cryptology conference. Springer, Berlin, pp 383–399
13. Chen L, Malone-Lee J (2005) Improved identity-based signcryption. In: International workshop on public key cryptography. Springer, Berlin, pp 362–379
14. Chow SSM, Yiu S-M, Hui LCK, Chow KP (2003) Efficient forward and provably secure ID-based signcryption scheme with public verifiability and public ciphertext authenticity. In: International conference on information security and cryptology. Springer, Berlin, pp 352–369
15. Libert B, Quisquater J-J (2003) A new identity based signcryption scheme from pairings. In: Proceedings 2003 IEEE information theory workshop (cat. no. 03EX674). IEEE, pp 155–158
16. Lal S, Kushwah P (2008) ID based generalized signcryption. *IACR Cryptol ePrint Arch*:84
17. Kushwah P, Lal S (2011) An efficient identity based generalized signcryption scheme. *Theoret Comput Sci* 412(45):6382–6389
18. Yu G, Ma X, Shen Y, Han W (2010) Provable secure identity based generalized signcryption scheme. *Theoret Comput Sci* 411(40–42):3614–3624
19. Shen X, Ming Y, Feng J (2017) Identity based generalized signcryption scheme in the standard model. *Entropy* 19(3):121

# Preventing Fake Accounts on Social Media Using Face Recognition Based on Convolutional Neural Network



Vernika Singh, Raju Shanmugam, and Saatvik Awasthi

**Abstract** In today's world, most people are intensely dependent on online social networks (OSN). People use social sites to find and make friends, to associate with people who share comparable intrigue, trade news, organize the event, exploring passion in an. According to a Facebook review, 5% of monthly active users had fake accounts, and in the last six months, Facebook has deleted 3 billion accounts. According to the Washington Post, Twitter has suspended over 1 billion suspect accounts over a day in recent months, Detection of a fake profile is one of the critical issues these days as people hold fake accounts to slander image, spread fake news, promotes sarcasm that has attracted cyber criminals. There are numerous machine learning methodologies such as supervised learning, SVM-NN, are produced for the effective detection of a fake profile. In this paper, convolution neural networks is proposed with many artificial neural network algorithms like face recognition, prediction, classification and clustering for the efficient identification of account being real or fake and elimination of fake profile account. Furthermore, the study is grounded on the fact of the face-recognizing of the user and performing feature detection and time series prediction. If the user account is detected fake it would not be created.

**Keywords** Social media analytics · Online social media · Face recognition · Convolutional neural networks · Age prediction

---

V. Singh (✉) · S. Awasthi  
Galgotias University, Greater Noida, India  
e-mail: [vernikasingsh8@gmail.com](mailto:vernikasingsh8@gmail.com)

S. Awasthi  
e-mail: [saatvikawasthi1998@gmail.com](mailto:saatvikawasthi1998@gmail.com)

R. Shanmugam  
School of Computing Science and Engineering, Galgotias University, Greater Noida, India  
e-mail: [dr.sraju@yahoo.com](mailto:dr.sraju@yahoo.com)

# 1 Introduction

There are increasing numbers of people in the current generation with social online networks like Facebook, Twitter, Instagram, LinkedIn, and Google+ [1]. Social networks allow people with common interests or collaborative reasons. It provides them with access to numerous services for example messaging, posting comments on their cyber-walls which are public, commenting on other users' profiles post and exchanging the masking of identity for malicious purposes has become progressively prevalent over the last few years. People rely heavily on OSNs [2] to remain in contact, to organize trade news, activities and perhaps even e-business and to create and share individual personal profiles, email, images, recordings, audios, and videos, and to find and make friends that have attracted cybercriminals interest in carrying out a variety of malignant activities. Government associations use (OSNs) as a forum for effectively providing government-driven services to people and educating and informing them about different situations. This heavy utilization of social networks results in immense measures of data being disseminated and organizations use social networks to promote, advertise, and support their business online. Fake profile accounts show that people do not represent them as a real person. Such an account is manually opened by a person after that actions are automated by bot. Fake profile account is categorized into a Sybil account and duplicate account. A duplicate account applies to a user's account and maintained by the user other than their main account. Fake accounts are classed into user classified (reported) or unauthorized (unwanted) groups of accounts. User malicious accounts records show individual profiles made by a client from a company or non-human element. Alternatively, undesirable accounts are however user identities that are configured to be used for violation of security and privacy.

The social networking site database for Facebook, records a statistic of 4.8% for duplicate accounts, the number of user-misclassified accounts is 2.4% and the number of unauthorized accounts is 1.5% [3]. In 2019, Facebook announced the deletion of 2.3 billion fake profile accounts. This is almost twice, as many as 1.2 billion accounts are withdrawn in the first quarter of 2018. The Facebook compliance report shows that as much as 5 million of its monthly active users are fake and that there is a growing number of attacks. Facebook is estimated to have more than two billion monthly active users and one billion active users each day in the major online social network reports. Accordingly, only 5% of its active monthly users are false in Facebook reports. It is very convenient today to make false accounts [4]. Nowadays, fake profile accounts can be purchased on the web at an extremely cheaper cost furthermore, it can be delivered to the client using publicly supporting., Now it is easier to purchase followers online from Twitter and Instagram. The goal behind Sybil account formation is to defame someone else's image, digital terrorism, terrorist propaganda; fear-based oppressor publicity, campaigns for radicalization, distribution of pornography, fraud and misinformation, popularity shaping, division of opinions, identity insecurity. In this paper, the promptly accessible and designed methodologies are evaluated, that is utilized for the fruitful detection of identifying

fake accounts on Facebook. Utilizing AI models, human-created detection is accomplished, by observing the attitude and the needs of people by examining their experiences. Detection is accomplished by observing the attitude and the needs of people by examining their communication and interaction with each other. Then there is a need to give a fake account to our machine learning model to allow the algorithm to comprehend what a fake account is. Convolution neural networks (CNN) and feature classification algorithms are being used. Our methodology is to differentiate on the reality between true user accounts and fake accounts by checking them at the time of creation and not allowing the fake profiles to be created.

## 2 Problem Identification

### Online Social Network (OSN)

Social media play a vital role in our life as 45% of the population around the world spend at least one hour daily on social networks to share news, post pictures, tweeting, commenting, liking, etc. [1]. Companies use the online social network to advertise and promote their business online. Government Organization use social media as a platform to deliver government services to citizens in an efficient manner and to educate and inform them in various ways. The highly dependent nature of billions of people over social networks has attracted the interest of cybercriminals to carry out malicious activities.

Computer-mediated communication (CMC) is a basic portion of our everyday lives, as a result, it has ended up essential for the populace, it is never going stalwart, as the society is never returning to a more physical environment, to considering characters on CMC stages like social media and gatherings. The idea of identity is expressly distinguished in this consider because it has to be overhauled logically as a critical theme of the social media wrangle about. As expressed prior, a few analysts have examined this issue, but determined 3-discusses are required since it is connected to “self” and “self” as demonstrated overnight. Inquire about concerned with the identity, picture and online gatherings have numerous possibilities and are especially meriting of thorough consider owing to social organizing designs predominant over thousands of a long time. Amid the final era, as well as numerous more youthful people effectively lead complicated genuine organizing, which may not reflect their current or offline presence. It has ended up a matter of concern. As of late, the specified viewpoint is a fair one in setting up untrue characters in social media. However, in another neighborhood, faux-accounts are made in advanced organizing and reasonably considers were done has moreover procured.

### Fake Profile

Fake profiles are created to defame someone’s image by using an individual name and pictures, terrorist propaganda, share fake news, distribution of pornography. It has a high impact on younger age who are highly dependent on social networks.

Most people try to portray an image of themselves who they are not and try to communicate with their dear ones and hate ones to harm them mentally which leads to a serious problem such as depression or suicide. Fake accounts can be categorized into duplicate account which is maintained by a user in addition to their principal account whereas false account is further broken down into user misclassified accounts which represent the user personal account created as a business account and undesirable accounts which are created to carry out malicious activities.

Social networking site, Facebook has estimated that 4.8% profile are double accounts, 2.3% profiles are unsolicited, and 1.6% are undesirable accounts [3]. The Facebook company has started removing 2.2 billion accounts that were detected fake in the Q1 of 2019. That is double the number detected in Q4 of 2018 where 1.2 billion fake profiles were deleted. According to Facebook's Enforcement Report, it is 6% monthly active users are fake and the increase is due to the rise in automated attacks [5]. The creation of fake accounts, buying Twitter and Instagram followers along with putting likes online on posts is very easy nowadays. It can be bought online at a very less cost and can be given to the customer via crowdsourcing services.

The reason behind the creation of fake accounts is mostly among these i.e. either to defame another person, terrorist propaganda, spreading rumors and false news, influencing popularity, polarizing opinions, identity theft, radicalization campaigns, cyberbully, dissemination of pornography and fraud, online extremism.

### **Approach to Solve**

The paper finds out the methods and the measures currently available for the detection of counterfeit accounts using machine learning technologies. Detection is carried out by learning the behaviors of individuals and requires a detailed analysis of the activities of social media interactions with other accounts. To identify other fake accounts and to predict fake accounts for potential use, this machine learning algorithm is training with the latest collection of identified fake profiles. Neural networks and algorithms are used for classification. Our solution uses distinct characteristics of the platform and restricts a platform to use an identity over the so-called internet.

## **3 Literature Survey**

This section summarizes some of the related work done in the field of detecting fake profile accounts using machine learning models.

### **3.1 Sarah Khaled, Neamat El-Tazi and Hoda M. O. Mokhtar**

In 2018 proposed a new algorithm to detect fake accounts and bots in Twitter. Classification algorithms of machine learning have been utilized, They focus on technologies adopted to detect fake accounts and bots on Twitter. Classification algorithms of machine learning have been utilized to choose real or fake target accounts, those techniques were support vector machine (SVM), Neural Network (NN). They proposed a new algorithm support vector machine-neural networks for successful detection of fake accounts. Both approaches adopt techniques of machine learning which highlight collection and methodologies of data reduction. It was also noted that the correlation collection records quality is dynamic among the other optimization algorithms, as redundancy is eliminated. The new algorithm classified 98% of the account of training dataset using fewer features [4].

### **3.2 Mudasir Ahmad wania, Nancy Agarwala and SuraiyaJabinb Syed ZeeshanHussainb**

In 2018 mainly focused on fake profile detection using sentiments. The study is done on the post of real account user and fake profile user and similar emotions they use. The experiment is done on Facebook user profile post. In this paper, the author mainly focuses on fake profile detection using sentiments. The study is done on the post of real account user and fake profile user and similar emotions they use. The experiment is done on Facebook user profile post. Data are trained for 12 emotions, including 8 basic emotions, positive and negativeness, by the use of machine training techniques consequently, outliers are removed using noise removal technique. To train the detection model, many learning algorithms have been used including Support Vector Machine (SVM), multilayer perceptron, J Rip, and Random Forest. The result shows that in the posts of unverified accounts three types of feeling, anxiety, shock and faith are found least. For all three measures, precision, estimation and AUROC, Random forest provides the best result [5].

Estée Van Der Walt and Jan Eloff

Described the detection of fake identities of humans vs bots using machine learning models. Numerous fake accounts are enhanced with features used to detect bot accounts, and the collection of features has been extended to different supervised learning models. This paper focuses on the detection of fake identities of humans vs bots using machine learning models. Numerous fake accounts are enhanced with features used to detect bot accounts, and the collection of features has been extended to different supervised learning models. The highlights of human and machine accounts are indistinguishable. For occasion, the title, it illustrates that the traits utilized to recognize programmer accounts fizzled to recognize human account



points of interest appropriately. The effects of qualified computer models are predictive 49.75% of the best  $F1$  performance. This is due to the fact that human beings are distinct from both in terms of behavior and characteristic, which cannot be modeled in the same way [6].

Gayathri A., Radhika S. and Mrs. Jayalakshmi S. L.

In 2018, explained Identification of fake accounts in media application by using support vector machines and neural networks. Problem definition: Identification of fake accounts in media application by using support vector machines and neural networks. In this report, they reflect a profound learning pipeline for identifying fake accounts rather than utilizing presumptions. It classifies the Sybil account cluster whether it is made by the same individual. The method starts by selecting a profile, at that point extricating the fitting characteristics and passing them to a proficient classificatory that categorizes the account as untrue alongside the input. Future work: utilizing more complex calculations. The other work line is to mimic multi-models utilizing the chosen highlights of other malware-based methods [7].

Author—year	Objective	Techniques used	Accuracy (%)
Gayathri et al. [1]	Detecting t ake accounts in media application	Support vector machine and deep neural networks	–
Mudasir Ahmad Wania et al. [5]	Analysis of real and fake users in facebook based on emotions	Narve Bayes, JRip random forest	Random forest
Şimşek et al. [8]	Detecting f ake twitter accounts	Artificial neural networks	–
Siordia and Moctezuma [8]	Features combination for the detection of malicious Twitter accounts	Feature extraction classification, random forest	94
Ttwari [9]	Analysis and detection of fake profile	Honest region, network nodes, network edge, benign nodes	–
Gupta and Kaushal [7]	Detecting fake user accounts in facebook	Data mining	79

## 4 Techniques Used in Literature Survey

### Support Vector Machine-Neural Networks

SVM-NN is applied to maximize classification accuracy because it achieves maximum accuracy using a reduced number of features is implemented on the provided dataset by performing feature reduction by splitting of data testing and training data using 8 cross folds. Neural networks are created by developing neurons and forming a model which are trained and used to predict results. The prediction accuracy is counted separately by using formula.

$$\text{Accuracy} = \text{number of detected accounts} / \text{total number of accounts} * 100$$

### Random Forest

Random forest is one of a classification algorithm that is unsupervised in nature. The fundamental concept is the selection of random samples from the provided dataset to create a decision tree followed by result prediction from each tree through voting. Finally, select the most voted results as the final prediction result. It is an ensemble method that achieves the highest accuracy as it reduced the overfitted data in sentiment analysis to identify true and sybilaccounts [5].

### Artificial Neural Networks

Artificial Neural networks system framework is computational processing systems that incorporate various interconnected computational nodes that work in a distributed manner to accumulate data from the input to optimize the final output. In implementations of ANN, a connection link is an actual number and the output is determined through a nonlinear input function in each neuron. The objective of the ANN is solving problems like a human brain for example in image recognition, pictures containing dogs can be detected by examining pictures manually marked as “dogs,” “no dogs” or by using the results to recognize dog in other pictures This occurs without prior knowledge that dogs have hair, ears or dog-like heads, for example, Characteristics are created by examples they identified automatically. ANN is used for classification, pattern recognition, clustering, regression analysis, prediction, social networks and even in activities that earlier only humans can do like a painting [9].

### Feature Detection

A convolution neural network has a special architecture in which complex data characteristics are detected feature is referred to as an “interesting” portion of an in general, image is processed as the first-pixel operation has been performed and each pixel is examined to determine whether a function exists on that pixel. When this belongs to a larger algorithm, the algorithm typically only scans the image in the function field. As a prerequisite for integrated function detection, the Gaussian

kernel usually smoothest the image input to the size display and calculates one or more feature images that are often expressed concerning the local image derivative.

### **Feature Extraction**

It is a process of reducing random variables from high dimensional datasets (data with more than 10 dimensions). It is broken down into two more parts that are Feature selection which is a simple approach to find subsets of input variables or attributes and feature extraction. Dimensionality reduction along with feature extraction is applied as a pre-processing step using techniques Linear Discriminant Analysis, K-Nearest Neighbor to reduce. Feature reduction is required to store time and storage [8].

### **Classification**

Classification is a complex phenomenon introducing the definition of classes according to the characteristics of the image following the selection of features such as color, texture and multi-temporal data. The feature dataset obtained is trained with supervised or unsupervised learning algorithms. Then, various classification techniques like extraction, segmentation are applied to the trained dataset to get appropriate results. At last, the classification technique is applied to all pixels by using pixel classification or per-field classification. Image classification covers all unique features of an image and an important factor in the digital environment [3].

## **5 Proposed Methodology**

The proposed algorithm utilizes the convolutional neural network for face recognition [10] and has an age classifier [11]. The features detected from CNN are classified and compared with existing data in the data warehouse for the data comparison.

Convolutional Neural Networks (CNNs) is a category of deep neural networks, most commonly used for visual imaging processing. They consist of neurons that optimize themselves through learning. CNN is also called space invariant artificial neural networks (SIANN). CNN is used for image recognition, video analysis, image classification, time series forecasting, recommendation systems, natural language processing, medical an input image can be taken by each neuron and operated based on numerous ANN. The only significant distinction between current ANNs and CNNs is that they are primarily used inside objects in the field of pattern detection. CNN comprises of 3 types of layers. The layers are convolutional layers, pooling layers and fully connected layers CNN's fundamental functionality can be explained as:

- The pixel values for the image act as the input and form the input layer.
- The convolutive layer determines the output of the neurons and the linear system is intended to apply CNN's "elemental" activation function to the activation output produced by the previous.

- The pooling layer just downsamples the contribution along with the spatial dimensionality and decreases the number of cases., hence reducing computational times.
- The fully connected layers then perform the indistinguishable tasks as generic ANNs and attempt to generate category results from classifying activations
- CNN makes developing network architecture simpler.

## 6 Proposed Algorithm

The approach of this system for fake profile detection is too limiting for each user to one and only one account on a particular platform. The system utilized facial recognition and age classification using CNN to create a unique facial print id of the account creator to identify him/her. This helps us to uniquely identify the customer by bind the facial print to the account eliminating the possibility for a user to create a new fake account. The facial data is utilized to identify two different data i.e. facial features and then the age prediction. This eliminates the possibility of creating new fake accounts by any user.

The input data after face detection and preprocessing are supplied to the two CNNs for Face Recognition and Age Classification. The data processed by the convolutional network by passing through various layers of convolution, ReLU, pooling/sampling and finally classified through a fully connected layer.

The process to detect a fake profile involves the following steps data collection, optimization, face detection, face recognition, age classification, profile association, profile detection. These steps are explained further in the section.

### Data Collection

The data is collected from the user. This is collected from the social media platform and then passed to the system. This step collects two types of data, firstly the profile details like name, age and facial data from the sensors for face recognition. This data is used to process and identify whether the user that is trying to create an account is genuine or not.

### Optimization

The raw data collected in the collection phase is optimized into the format required. The optimization is one of the important steps before the processing of the data as it prepares the data and enhances the data so that during the processing of data the algorithm can produce better results.

### Face Detection

Detection involves the detection of the facial data from the collected data and identifies the points required to be processed and specific to the face rather than the whole picture. The detection phase involves facial detection by identifying the points that are of use to us and eliminating the rest unnecessary points. This is done through

template matching function. It defines a standard template for all the faces and where the different features can be identified independently like eyes, nose, mouth, contour, etc.

### **Face Recognition**

Face Recognition [12] is performed using the Convolutional Neural Network. The data is preprocessed before feeding to the convolutional network. CNN utilizes various hidden layers of convolution, ReLU, and pooling. These layers are arranged in some fashion repeatedly to form the network. After passing through the various hidden layers the output is put to the fully connected layer of the classifier for classification. The data from the output layer is put to comparison by the dataset in the data warehouse for profile detection.

### **Age Classification**

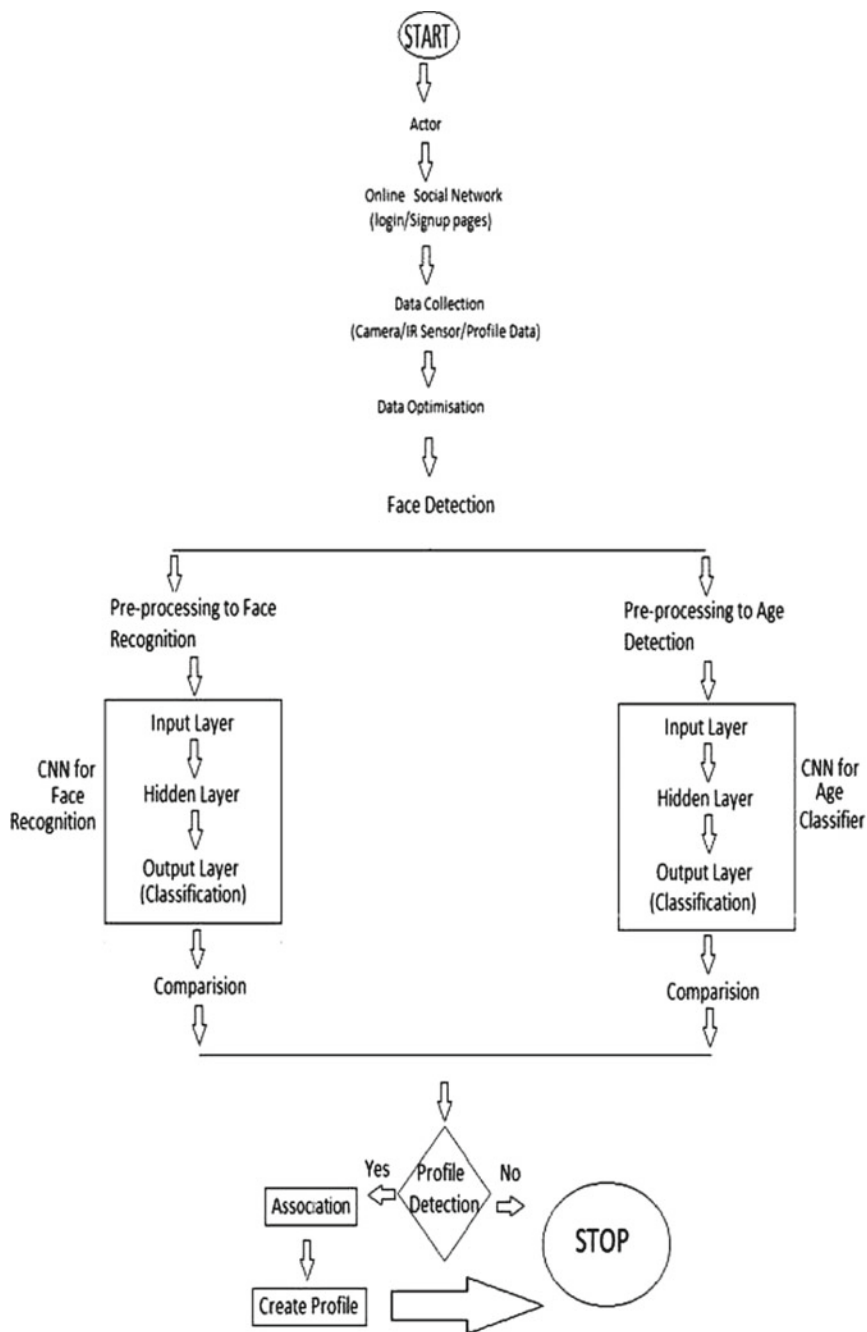
Age classification involves identifying the age of the user using CNN to get near about the age of the user. A different CNN is used for the age classifier. The age predicted by the classifier is given as an estimated range. If the input of the user lies between the range the profile is allowed for creation. This is an extra parameter just to verify the data input by the user and the verification facial data matches the data input by the user at the time of creating the profile [13].

### **Profile Detection**

This involves the main comparison of the data from the data warehouse that utilizes snowflake schema to store the data. The data after the process of classification is passed to this phase. The data is compared with the data in feature data in the data warehouse. If a match is found, then the profile is flagged as fake and the account is not created. If the face is unique, then the age detection is data is utilized to compare the age detected by the algorithm with the data entered by the user. If the age matches near about with a minor difference then the profile is created else it is detected as a fake verification. In the end, if both the tests are accepted then the profile is created and the facial data is moved to the data warehouse for mapping.

### **Profile Association**

The data is associated with the user profile once it is created [4]. The data is added to the data warehouse creating a unique identification of every use and also maintains the data to which all platforms the user is signed up to. The data warehouse utilizes are snowflake schema to store the data by identifying the user by its facial patterns. This eliminates the possibility of creating a fake profile of the user on a platform he/she is not utilizing.



## 7 Future Scope

This paper proposes a method that utilizes CNN for face recognition and age classification to eliminate profile. Further implementation can be done using CNN and the accuracy and success of the method can be identified. The paper only proposes the idea of limiting each person to have only one profile on a social media platform to prevent the creation of a fake profile by mapping the creator's facial signature and profile data verification of age.

## 8 Conclusion

In this paper, the identified problem of fake accounts on the online social media platform is tried to solved by the usage of neural networks and user profiling. The paper proposes a system that is expected to prevent the creation of fake profiles as compared to the previous system that utilized machine-neural networks that eliminate fake profiles. In our knowledge, did not have any such system that has taken this approach for fake profile elimination as proposed in the system. The accuracy is not acclaimed as it has to experiment in the future implementation of this system.

## References

1. Gayathri A, Radhika S, Jayalakshmi Detecting fake accounts in media application using machine learning. *Int J Adv Netw Appl (IJANA)*
2. Sarah Khaled A, El-Tazi N, Mokhtar HMO (2018) Detecting fake accounts on social media. *IEEE Int Conf Big Data 6*:101–110
3. Hudson B, Voter BR (2016) Profile characteristics of fake twitter accounts. *Big Data Soc*
4. Yang Z et al (2014) Uncovering social network sybils in the wild. *Trans Knowl Discov Data (TKDD)* 8(1)
5. Mudasir Ahmad Wania B, Agarwala N, Hussainb SJSZ (2018) Analyzing real and fake users in facebook network based on emotions. In: 11th international conference of communication system and networks
6. Van Der Walt E, Eloff J (2018) Using machine learning to detect fake identities: bots versus humans. *IEEE Access*
7. Gupta, Kaushal Towards detecting fake user accounts in facebook. *Indira Gandhi Delhi Technical University for Women, Delhi, India*
8. Şimşek M, Yilmaz O, Kahriman AH, Sabah L (2018) Detecting fake twitter accounts with using artificial neural networks. 2018 artificial intelligence studies. In: David I, Siordia OS, Moctezuma D (eds) *Features combination for the detection of malicious Twitter accounts*. *IEEE international autumn meeting on power, electronics and computing*
9. Tiwari V (2017) Analysis and detection of fake profile over social network. In: *International conference on computing, communication and automation*
10. Oloyede MO, Hancke GP, Myburgh HC (2018) Improving face recognition systems using a new image enhancement technique, hybrid features and the convolutional neural network. *IEEE Access* 6:75181–75191. <https://doi.org/10.1109/ACCESS.2018.2883748>

11. Rafique I, Hamid A, Naseer S, Asad M, Awais M, Yasir T (2019) Age and gender prediction using deep convolutional neural networks. In: 2019 international conference on innovative computing (ICIC). <https://doi.org/10.1109/ICIC48496.2019.8966704>
12. Yin X, Liu X (2018) Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Trans Image Process* 27(2):964–975. <https://doi.org/10.1109/TIP.2017.2765830>
13. Chen S, Zhang C, Dong M, Le J, Rao M (2017) Using ranking-CNN for age estimation. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). <https://doi.org/10.1109/CVPR.2017.86>



# A DNN Based Diagnostic System for Heart Disease with Minimal Feature Set



Merin David and K. P. Swaraj

**Abstract** Cardiovascular diseases are one of the prime reasons for individual deaths across the globe, claiming millions of lives every year. Cardiovascular disease diagnosis is a critical challenge in the health care field as it has a lot of risk factors associated with it. Moreover, neural networks are ideal in making significant clinical decisions from the huge health-care data produced by the hospitals. This work is an effective method to find significant features and use Deep Neural Networks to build a cardiovascular disease diagnosis system. The proposed system is developed by using a well known dataset called Cleveland dataset of the UCI Repository. The model was introduced with different combinations of features and a deep neural network. The performance of the different subsets of features were evaluated and it was also investigated for a single gender. The proposed system with a minimal feature set helps in the early diagnosis of heart disease and aids the cardiologist in making clinical decisions.

## 1 Introduction

Any malfunction occurred in the heart is called heart disease. Diameter decrease and subsequent artery blockage is the most widespread cause of heart failure. Heart disease is observed as the most common death cause and it is difficult to diagnose the heart disease due to several associated complicating factors and hence, the disease must be handled carefully. Otherwise it may affect the heart or even result in death.

Currently, the most common diagnosis method by medical practitioners is angiography. This is also considered as the most accurate and appropriate method [5]. But, it has many side effects and is very costly. Moreover, analyzing many risk factors makes it more difficult for doctors. This arises the need for non-invasive methods.

---

M. David (✉) · K. P. Swaraj  
Government Engineering College, Thrissur, Kerala 680009, India  
e-mail: [merin.davidjohn@gmail.com](mailto:merin.davidjohn@gmail.com)

K. P. Swaraj  
e-mail: [swarajkp@gmail.com](mailto:swarajkp@gmail.com)

The conventional non-invasive methods depend upon analyzing the patients medical history, relevant symptoms, and physical examination report. However, these methods are often erroneous [6, 7]. Thus, there is a need to develop systems that diagnose heart diseases by eliminating these issues.

Many learning based decision support systems have been suggested for the detection of heart disease previously. However, these help to direct the focus on pre-processing of features and improving the classifier accuracy only. Also, deep neural networks have been showing better performance in health-care field be it disease prediction, patient similarity learning [2], diet recommendation [3], dosage prediction [4] or any other type of medical learning, recommendation, classification or prediction. In this work, a deep neural network system is proposed with the significant focus on the refinement of features in order to obtain good performance.

The main contributions of this work are as follows:

1. An effective system for heart disease detection with lesser number of features.
2. This work determines the significance of the neural network depth on its performance. Generally, it is assumed that shallow networks perform better than deep ones when working with small datasets. But, this paper shows that deep networks too exhibit good performance provided unnecessary features are eliminated and not made too deep.
3. The performance of the system for single gender data i.e. either male or female patients, was also investigated.

## 2 Related Work

Adequate work in the fields directly related to this work are present. Attribute Weighted Artificial Immune System (AWAIS) by Sahan et al. [8] depends on weighting parameters according to their significance. These are then used to calculate Euclidean distances. The performance of AWAIS was investigated for the Statlog Heart Disease dataset and also another medical dataset for diabetes, Pima Indian Diabetes dataset.

A classification model was proposed with the Statlog Disease dataset again by Tomar et al. [10]. They analyzed the performance of feature selection based LSTSVM model for different training-test ratios and the results indicated that LSTSVM model with 11 features performed well.

Paul et al. [9] proposed a fuzzy decision support system (FDSS) using genetic algorithms (GA). The model finds out knowledge as decision rules. The proposed system also reduces the number of tests needed as it selects the optimal set of attributes and diagnoses the disease early.

Wiharto et al. [11] implemented a hybrid learning model for diagnosis of coronary heart disease. The system used feature selection process based on tiered multivariate analysis, and logistic regression. Classification was performed using multi-layer perceptrons.

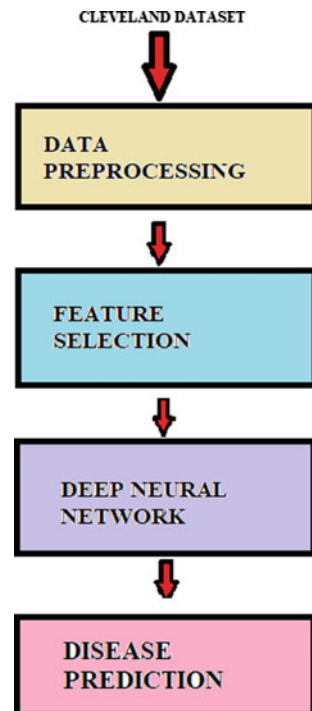
Machine learning techniques were used in the work proposed by Mohan et al. [12] to process raw data and provide a novel work in the area of heart disease. The proposed approach was a hybrid HRFLM that combined Random Forest (RF) and Linear Method (LM). The dataset used was the Cleveland dataset from the Heart Disease dataset. All 13 features of the dataset were used in building of the model.

### 3 Proposed System

The proposed work consists of developing the DNN model with different feature sets and conducting a comparative study on the same in terms of the evaluation metrics. In this study, Python and the Keras library were used to build the deep neural network.

The proposed approach (Fig. 1) begins with a pre-processing phase. It includes handling of missing values, categorical values and the normalization process performed on feature vectors. Sci-kit learn (sk-learn) library has been used for this purpose. Feature extraction is then performed on the data. The reduced data is applied to neural network for model building and updating weights. The network is evaluated for its performance using K-Fold validation.

Fig. 1 System architecture



### ***3.1 Dataset Description and Data Pre-processing Module***

The Heart Disease dataset, freely available at UCI, contains the databases from Hungary, Switzerland, V. A. Long Beach and Cleveland. The database contains 76 attributes, but usually a subset of 14 of them are selected where 13 were feature vectors and the remaining 1 was class label. The details of these feature columns have been clearly described in Table 1. Out of the 13 feature vectors, there were 2 personal features namely sex and age of the patient and the remaining were attributes pertaining to the disease. The confidential features like names, social security numbers were removed from the database for obvious reasons and hence, are not considered.

This work uses Cleveland database only as the remaining three databases had numerous missing values. The Cleveland database had 303 patient records out of which 6 records had columns with missing values. So, those 6 were manually removed from the database and the remaining 297 records were used by the pre-processing module. Hence, the dataset had 297 rows and 14 columns.

The last class variable field suggests the extent of disease in the patient. It has values from 0 to 4 depicting degrees of severity of the disease. Most works with the Cleveland database have focused on distinguishing presence (values 1, 2, 3, 4) from absence (value 0) [12, 14, 15]. So, the multi-class problem was converted into a binary classification problem by mapping all 'num' values greater than '1' to simply '1' to indicate presence of heart disease. Hence, there were 137 records showing the value of '1' i.e the presence of heart disease, and the remaining 160 had the value of '0', i.e the absence of heart disease.

There are columns like 'sex', 'cp' etc. that contain numbers of no specific order of preference, called categorical attributes. The data in these columns usually denote a category or value of the category. They cannot be fed into the model directly and hence, should be One Hot Encoded. In One Hot Encoding, the column which contains categorical data are split into as many columns equal to the number of categories present in that column. Each column will contain either a '0' or a '1'.

### ***3.2 Feature Selection Module***

The Cleveland dataset, as already said, had 13 feature columns and 1 target column. The deep neural network model was built with all of these features. Also, the results of the subset of features were compared. In this work, only six sets of features were selected for the evaluation of the performance as proposed by [13].

There were 201 male samples and 96 female samples in the dataset. So, model building was also performed on single gender data, i.e all male samples or all female samples were taken separately and the model was built for studying the comparisons with these feature sets, as 'sex' is the common personal attribute for all feature sets.

**Table 1** Description of the dataset [16]

Attribute	Description	Type
Age	Age in years	Number
sex	Sex (Value 1: male, Value 0: female)	Categorical
cp	Chest pain type (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)	Categorical
trestbps	Resting blood pressure in mm Hg on admission to the hospital	Number
chol	Serum cholestrol in mg/dl	Number
fbs	Fasting blood sugar greater than 120mg/dl (Value 1: true, Value 0: false)	Categorical
restecg	Resting electrocardiographic results (Value 0: normal, Value 1: having ST-T wave abnormality i.e. T wave inversions and/or ST elevation or depression of greater than 0.05 mV, Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria)	Categorical
thalach	Maximum heart rate achieved	Number
exang	Exercise induced angina (Value 1: yes, Value 0: no)	Categorical
oldpeak	ST depression induced by exercise relative to rest	Number
slope	The slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)	Categorical
ca	Number of major vessels (0-3) Colored by flouroscopy	Number
thal	Status of the heart (Value 3: normal, 6: fixed defect, 7: reversable defect)	Categorical
num	Diagnosis of heart disease i.e. Angiographic disease status (Value 0: less than 50% diameter narrowing, Value 1: greater than 50% diameter narrowing)	Categorical

### 3.3 Prediction Module with Deep Neural Networks

A neural network consist of weighted interconnections of an input layer, zero or more hidden layers and an output layer. Deep neural networks are neural networks that have more than one hidden layer.

This work uses a feed forward deep neural model that had 3 hidden layers with 13, 8 and 4 neurons respectively. The output layer had a single neuron. All layers (input layer doesn't use any activation function by default) used sigmoid activation function.

Adam was used as the model optimizer and loss of the model was calculated using binary cross-entropy formula given by:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=0}^N (y \cdot \log(\hat{y}_i) + (1 - y) \cdot \log(1 - \hat{y}_i)) \tag{1}$$

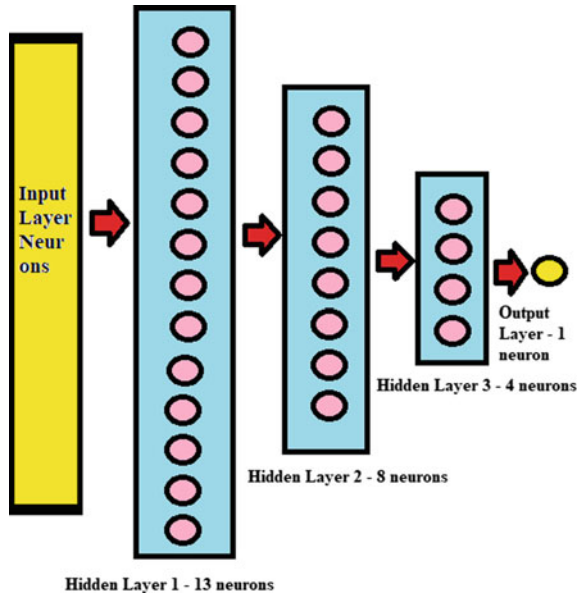
where  $y$  is the set of target values,  $\hat{y}$  is the set of predicted values and  $N$  is the number of values in the set. The block diagram of the model is given in Fig. 2. The dataset was split into 9 splits after leaving the last 9 rows for real-time predictions. So, there are 32 samples in each split. The model was fit with all the features and also the selected feature sets. For studying the performance of these feature sets on single gender data, 3 splits were taken and hence, there were 67 and 32 samples in each split for males-only and females-only data respectively.

### 4 Evaluation Metrics and Results

The standard evaluation metrics such as accuracy, precision, recall and F1-measure [17] have been considered as this is a binary classification problem.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Fig. 2 Deep neural network model



$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

$$\text{F1-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

where TP, TN, FP and FN are True-Positives, True-Negatives, False-Positives and False-Negatives of the confusion matrix respectively. Since the model was run for 9 splits, the above metrics for each fold was determined and averaged to obtain the performance metrics.

The comparison of the different feature sets based on performance metrics after K-fold validation are shown in Table 2 and Fig. 3. Feature set named FS3 achieves highest values for Accuracy, Precision, Recall and F1-measure. FS1 (7-features) and FS3 (8-features) clearly outperforms the model built with the full feature set as it has higher values for all the performance metrics. FS4 (6-features) also achieves better values for Accuracy, Precision and F1-measure when compared to the model built with the 13-feature set but it's Recall value is lesser than the 13-feature set.

**Table 2** Comparison of different feature sets

Feature sets	Accuracy	Recall	Precision	F1-measure
All features	81.6	78.28	79.49	78.6
FS1	82.3	80.94	80.68	80.11
FS2	78.89	81.08	71.43	75.95
FS3	84.38	83.14	83.58	82.57
FS4	81.95	76.51	83.5	79.04
FS5	79.86	77.47	78.61	76.99
FS6	80.56	72.22	80.42	75.13

**Table 3** Comparison of different feature sets for single gender data

Feature sets	Accuracy for males-only data	Accuracy for females-only data
All features	73.13	85.42
FS1	77.11	86.46
FS2	75.12	84.38
FS3	75.29	88.54
FS4	74.13	81.25
FS5	78.11	81.25
FS6	77.86	84.38

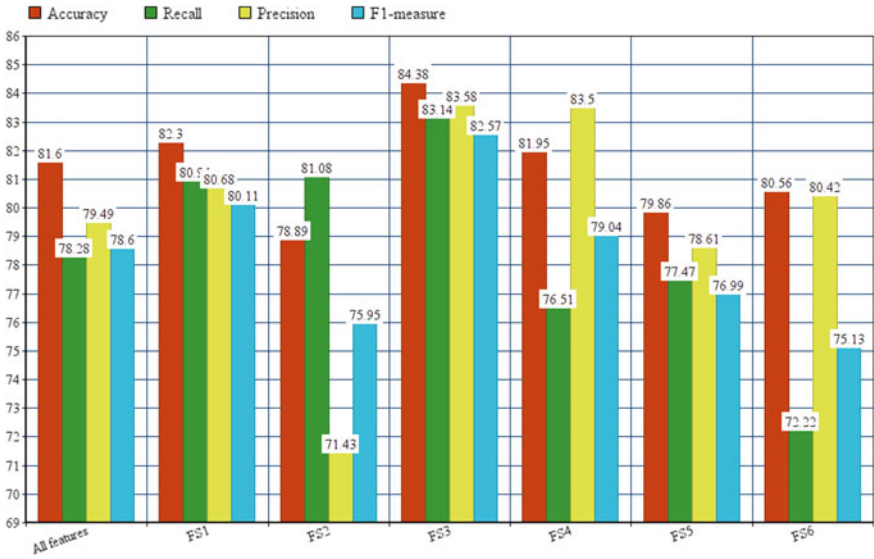


Fig. 3 Graphical illustration of the comparison of different feature sets

The comparison of accuracies for different feature sets on single gender data after K-fold validation is given in Table 3 and Fig. 4. For feature sets FS1 (7-features) and FS3 (8-features), accuracy for disease diagnosis with males-only data and females-only data is clearly higher than the males-only data and females-only data for full features model respectively. Therefore, all the 13-features are not necessary for the deep neural system and the relevant feature sets are FS1 = {sex, cp, fbs, restecg, oldpeak, ca, thal} and FS3 = {sex, cp, fbs, thalach, exang, slope, ca, thal}.



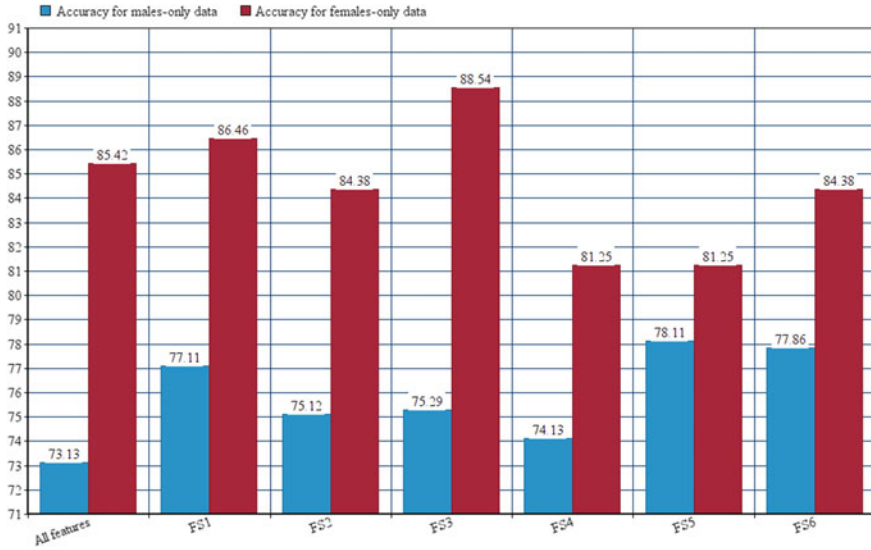


Fig. 4 Graphical illustration of the comparison of different feature sets for single gender data

## 5 Conclusion

From this research work it is concluded that the heart disease prediction is an important health-care field as the death rates should be effectively controlled. If the disease is detected in early stages and all safety measures are adopted as soon as possible, it will help in saving human lives in the long run and early diagnosis of heart abnormalities.

In this work, the performance of a deep neural based heart disease diagnosis system is evaluated. The neural network had three hidden layers with 13, 8 and 4 neurons respectively. Feature selection was employed and six feature sets were selected to train the model. The proposed system with minimal feature sets FS1 and FS3 achieves a better performance when compared with the model employing all the 13-features. This system can be effectively installed in clinical decision support systems at hospitals for aiding medical practitioners in making important clinical decisions.

Limitations of this system is that the dataset is very small. This study can be further extended to real-world datasets with atleast 1000 patient records for multiple classes and its performance could be tested after a relevant outlier detection process. Moreover, the study was conducted with the available data and so, further investigation is required on why there is a huge variation in the performance between the two genders.

## References

1. Bashar Abul (2019) Survey on evolving deep learning neural network architectures. *J Artif Intell* 1(02):73–82
2. Suo Q et al (2018) Deep patient similarity learning for personalized healthcare. *IEEE Trans Nanobiosci* 17(3):219–227
3. Manoharan S (2020) Patient diet recommendation system using K clique and deep learning classifiers. *J Artif Intell* 2(02):121–130
4. Norouzi Kandalan R et al (2020) Dose prediction with deep learning for prostate cancer radiation therapy: model adaptation to different treatment planning practices. *arXiv* (2020): arXiv-2006
5. Arabasadi Z et al (2017) Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm. *Comput Methods Prog Biomed* 141:19–26
6. Yan H et al (2006) A multilayer perceptron-based medical decision support system for heart disease diagnosis. *Expert Syst Appl* 30(2):272–281
7. Vanisree K, Jyothi S (2011) Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks. *Int J Comput Appl* 19(6):6–12
8. Şahan S et al (2005) The medical applications of attribute weighted artificial immune system (AWAIS): diagnosis of heart and diabetes diseases. In: *International conference on artificial immune systems*. Springer, Berlin, Heidelberg
9. Paul AK et al (2016) Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease. In: *2016 5th International conference on informatics, electronics and vision (ICIEV)*, IEEE
10. Tomar D, Agarwal Sonali (2014) Feature selection based least square twin support vector machine for diagnosis of heart disease. *Int J Bio-Sci Bio-Technol* 6(2):69–82
11. Wiharto W, Kusnanto H, Herianto Herianto (2017) Hybrid system of tiered multivariate analysis and artificial neural network for coronary heart disease diagnosis. *Int J Electr Comput Eng* 7(2):1023
12. Mohan S, Thirumalai C, Srivastava G (2019) Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access* 7:81542–81554
13. Latha CB, Jeeva SC (2019) Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inf Med Unlocked* 16:100203
14. Bo J, Tang YC, Zhang Y-Q (2007) Support vector machines with genetic fuzzy feature transformation for biomedical data classification. *Inf Sci* 177(2):476–489
15. Khare S, Gupta D (2016) Association rule analysis in cardiovascular disease. In: *2016 Second international conference on cognitive computing and information processing (CCIP)*. IEEE
16. Detrano R, Dua D, Graff C (2019) UCI machine learning repository. University of California, School of Information and Computer Science, Irvine, CA. <http://archive.ics.uci.edu/ml>
17. Shung KP, Accuracy, precision, recall or F1?. *Towards Data Science Inc.*, Canada. <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

# Robust Automated Machine Learning (AutoML) System for Early Stage Hepatic Disease Detection



Devendra Singh, Pawan Kumar Pant, Himanshu Pant,  
and Dinesh C. Dobhal

**Abstract** Hepatic or Liver disease cause illness because of the perturbation of liver function. The liver performs many critical functions and if it gets injured or diseased, cause significant damage or even death. Failure to take care of early-stage hepatic issues will further deteriorate the organ to more complicated scenarios. Therefore the need to diagnose the condition early so that the condition has maximum potential for successful treatment. AutoML is the next coming wave of machine learning. In the market there are various commercial and open-source tools are available based on AutoML. Auto-WEKA package is such a tool and it is based on AutoML technique, which could be utilized by both experts, and non-experts group of users. The main usage of AutoML is to help the developer by automating the model selection and hyper-parameter optimization. Non-experts utilize AutoML with minimal code or without writing a single line of code. Tools that bases on AutoML make machine learning pipeline building effortless. In this paper, the implementation of Auto-WEKA for hepatic disease detection have been used.

**Keywords** AutoML · Auto-WEKA · Hepatic disease · Model selection · Hyper-parameter optimization

---

D. Singh (✉) · P. K. Pant · H. Pant · D. C. Dobhal  
Department of Computer Science and Engineering, Graphic Era Hill University-Bhimtal Campus,  
Bhimtal, Uttarakhand, India  
e-mail: [devendrasuno@gmail.com](mailto:devendrasuno@gmail.com)

P. K. Pant  
e-mail: [pantpawan@gmail.com](mailto:pantpawan@gmail.com)

H. Pant  
e-mail: [himpant7@gmail.com](mailto:himpant7@gmail.com)

D. C. Dobhal  
e-mail: [dineshdobhal@gmail.com](mailto:dineshdobhal@gmail.com)

# 1 Introduction

Early-stage hepatic disease identification is a critical task in the patients because it may function normally even when the hepatic is partially damaged. Symptoms sometimes come to light when it's too late. Hepatic patients' survival rate might increase when the diagnosis is performed at an early stage. Hepatic is one of the core parts of the human body and its main responsibility is to filter blood before sending to the rest of the body parts. Hepatic harm is one of the best deadliest ailments on the planet. Hepatic failures among Indians pose a high-risk rate. Earlier, alcohol and other obesity-associated diseases usually caused by hepatitis B and C may now be considered the most common causes of liver disease. There has been a paradigm shift in the dynamics of liver cirrhosis and approximately 10 lakh new patients are diagnosed in India each year. By 2025, India will potentially become the world's capital for hepatic infections. With the sedentary routine, too much alcohol consumption and smoking are the main cause of hepatic infection in India. There are more than 100 types of hepatic infections in the human body. Therefore, a hepatic patients screening system is developed that will enhance disease diagnosis in the medical field. Such screening systems can allow doctors to make specific decisions on the hepatic disease of the patient using automated diagnosis tools. Hepatic disease diagnosis at an initial phase is significant for a better cure. Prediction of diseases in the starting phases is a challenging task for medical researchers. Hepatic disease is a broad term covering all the potential problems causing the liver to fail to perform its designated functions. There are several types of hepatic diseases: virus-caused diseases like hepatitis A, hepatitis B, and drug-caused hepatitis C, Diseases, toxins, or too much alcohol. Types include fatty hepatitis and cirrhosis. Symptoms of the hepatic disease can vary, but frequently include swelling of the abdomen and legs, easy bruising, changes in the colour of the stool and urine, and jaundice, or skin and eye yellowing. Symptoms don't occur often. Our research paper provides hepatic disease diagnosis with the help of latest and very useful technique called automatic machine learning.

Machine learning has provided some significant breakthroughs in various diverse fields. Fields like healthcare, financial services, retail, transportation, marketing, insurance and many more. Nowadays, Machine learning becomes the need for the organization. If they are not using machine learning, they are missing the current industry and their customers need. The main problem associated with traditional machine learning is that the whole process is human dependent, but the problem is that not all business has the resources to invest in an experience machine learning team. Therefore, AutoML may be the answer to such kind of situations. WEKA contains an implementation of mostly used machine learning algorithms. In the current era of technology, user can easily use machine learning with the help of available machine learning algorithms and other feature selection techniques with the help of open-source packages like WEKA [1] and PyBrain [2]. WEKA provides many valuable resources for data pre-processing, classification, regression, clustering, visualization and much more. The main advantage of WEKA is that it is open-source software.

Automated machine learning has mainly used for automating the whole process of machine learning pipeline. Auto-WEKA package based on AutoML. Auto-WEKA is a package of WEKA data mining tool that performs algorithms selection from an available set of algorithms and hyper-parameter optimization. Hyper-parameter optimization plays a vital role in model selection. The best algorithm from available algorithms may be used but without the best-tuned parameters, its optimal performance cannot be achieved. AutoML provides us with the power that without the skilled data scientists an efficient machine-learning model can be built.

The main aim of automated machine learning is to make machine learning easy, reduces human error as before the work has done manually, improving models performance by fine-tuning. Thus the automated machine is a new era and a very demanding topic for research. Many firms investing a lot of amount in the research area of improving existing automated machine learning and have also used this in our paper to take advantage of automated machine learning. The main objective of this research is to build a program for the screening of hepatic diseases using classification algorithms for machine learning to distinguish hepatic patients from healthy people. An integrated machine-learning model is developed using AutoML, Auto-WEKA, Auto-Keras and Auto-Sklearn techniques to support the medicinal community for the diagnosis of hepatic disease among patients. AutoML is defined as a combination of machine learning and automation. Doctors and medical practitioners can readily utilize this screening model as a screening tool for hepatic disease.

### ***1.1 AutoML Versus Traditional Machine Learning***

AutoML concepts are varied. According to [3], for example, AutoML is designed to reduce demand for data scientists and enable domain experts to create ML applications automatically without much requirement. Awareness of statistics and the ML. In [3], AutoML is specified as an Automation and ML combination. Automated Machine Learning offers tools and processes for making Machine Learning accessible to non-machine Learning experts, increasing the performance of Machine Learning and speeding up Machine Learning. There are many tools available in the market for AutoML platform. With the help of AutoML, the time-consuming part of the AI application development lifecycle can be automated. In traditional machine learning, more attention has to pay for manual work. For example, in traditional machine learning, a model selection with best-tuned parameters a repetitive task has to be performed manually which can be a very time-consuming process. With the help of AutoML, the process can be automated which can save both time and resource.

## 1.2 Advantages of AutoML

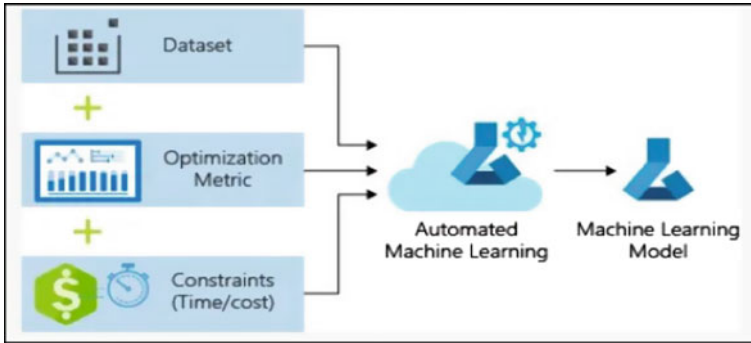
- Automated machine learning provides the solution by automating some or all of ML's moves. It helps the seeker to incorporate supervised learning, which includes pattern recognition from the labelled data.
- Automated ML is responsible for model consistency and accuracy. Chances of a mistake or mistake happening are reduced. AutoML thus provides a higher degree of satisfaction levels.
- The data processing time is decreased so developers can utilize this time on other aspects of machine learning like investing more time on optimization of functions.
- AutoML is cost-effective by automating tasks. The number of resources is being reduced such as the number of developers in a project.
- AutoML selects and extracts the features of a given dataset along with best-tuned hyperparameters.

## 1.3 Risks Associated with AutoML

- When the data comes from various sources and is available in different formats, AutoML can not be consistent with different datasets.
- With AutoML's support, only some part of the problem can be solved. Mainly focuses on selecting and improving model performance.
- AutoML is computationally expensive in terms of hardware resources and time it needs for a particular problem. It needed more time and resources to tune the hyperparameter than the other parameters.
- The time is also a critical part of AutoML. Every problem is time-bound in the sense to achieve the best accuracy and sometimes more time can overfit the model.

## 2 AutoML Pipeline

A standard model of machine learning follows four processes, reading data from the data source, pre-processing, and hyperparameter optimization to the prediction of results. All the steps are performed and controlled manually. AutoML mainly focuses on two points: the collection of data and prediction. In AutoML intermediate steps are automated. AutoML provides us with optimized models that are ready for prediction. AutoML pipeline is an automatic method of applying machine learning to real-world issues. It automates everything from data selection to development of a machine learning model. Human input and expertise in the conventional machine learning model are needed at different stages of machine learning. With the help of AutoML, the productivity of experts can be improved by automating repetitive machine learning tasks which can also reduce the human errors that arise in the manual process (Fig. 1).



**Fig. 1** Auto train models [4]

## 2.1 Benefits of AutoML

For data scientists, AutoML is not a substitute. It cannot do what a data scientist can do in terms of analytical thought, predictive engineering or knowing the meaning and weaknesses of its projects [3]. Before AutoML, experienced data scientist spends lots of time on manual tasks such as feature selection and hyper-parameter tuning instead of other deep levels of analysis. AutoML makes it possible to run the automated process in the background during that time data scientists can focus on complex tasks. So can be said that AutoML is not the replacement of data scientists. For new data scientists, AutoML makes it possible to make their projects ready on time and also their fundamental skills such as algorithm selection and tuning of hyper-parameters. Various other categories of users can take benefits of AutoML for example healthcare domain, finance domain, marketing domain, banking domain, etc.

## 3 AutoML Tools

### 3.1 Auto-WEKA

Auto-WEKA package is used for selecting a learning algorithm and setting the values of hyperparameter. Auto-WEKA does all the processes in a fully automated manner. This can be used by non-experts to define the machine learning algorithms and the hyperparameter values for their desired application. It provides a simple and easy GUI interface, which can be utilized by non-experts with minimal machine learning.

### **3.2 *Auto-Keras***

Auto-Keras is an open-source software based on automated machine learning. Auto-Keras search the correct model architecture and hyper-parameters for the deep learning models automatically. With help of the Auto-Keras and with few lines of code, a powerful deep learning model can be created.

### **3.3 *Auto-Sklearn***

Auto-Sklearn is a package of scikit-learn based applications for automated machine learning. Scikit-learn is a free software library for machine learning in the python programming language. Auto-Sklearn fits well in small and medium-sized datasets but is not ideal for a complex deep learning system capable of managing a large number of data sets.

### **3.4 *Auto-Pytorch***

Facebook manages Pytorch library. Auto-Pytorch is used to automate the time consuming manual process in machine learning. Searching the appropriate architecture and hyper-parameter values for training the deep neural network is crucial to achieving the best performance.

### **3.5 *Amazon-Lex***

Amazon Lex is the deep learning technology-based web service provided by Amazon Alexa. It is a web service that allows clients to use text and voice to create conversational interfaces into any program. Developers to quickly build Chat-bots with minimal effort mainly use it. Chat-bots mainly used for conversational interactions to help firms to engage the customers. This comes with functionalities such as automatic speech recognition (AST) to translate from voice to text and natural language understanding (NLU).

### **3.6 *Tpot***

The Tree-Based Pipeline Optimization Tool (TPOT) is an open-source data science python tool for automation. TPOT uses the scikit-learn library used in the python



as its ML menu. TPOT uses a genetic algorithm for finding the best parameter and model ensembles. TPOT tries a pipeline, tests its output, and then changes parts of the pipeline by a random search for better algorithms.

### ***3.7 H2O Auto-ML***

H2O is an open-source machine-learning framework that is distributed in memory. H2O contains an automatic machine learning module component called H2OAutoML which is used primarily to automate the machine learning workflow. H2OAutoML is used to automatically train and tune several models within a given time limit.

### ***3.8 RapidMiner***

With the help of RapidMiner, automated machine learning can be performed. The main advantage of automated machine learning is that reduced time and effort for building a model can be used. With the help of RapidMiner auto model, predictive models can be built within minutes. Thus the need for technical expertise can be eliminated. Only the data have to be uploaded and specify the outcomes, the Auto Model part will do the rest other things with high-value insights. RapidMiner Auto Model works as a fully automated manner from data exploration to modelling to production.

### ***3.9 Cloud Auto-ML***

Cloud AutoML is a suite of machine learning products provided by Google that allows developers with limited machine learning knowledge to train high-quality models that are specific to their business needs by leveraging Google's state-of-the-art transfer learning and Neural Architecture Search. Cloud AutoML offers an easy to use GUI to practice, test, develop and model. The suite contains many technology tools for different tasks. Google's AutoML is paid software and to use it in our project, have to pay as compared to open-source alternatives. AutoML Vision is a package of Cloud AutoML suite used for training the model on images. Cost of AutoML vision depends on time and number of images used by that model for prediction.

## 4 Experimental Setup and Results

In recent research works, researchers developed numerous hepatic disease detection models for early-stage diagnosis. Neural network models for disease predictions are one of the exceptional medical aids for doctors, including a diagnostic support system, an expert system, an intelligent diagnostic system and a hybrid system for the detection and screening of hepatic diseases. Christopher [5] proposed a diagnostic system for detecting medical conditions, taking into account six benchmarks: hepatic disorder, heart disease, diabetes, breast cancer, hepatitis and lymph (Fig. 2).

Patients with the Hepatic disease have been increasing because of high-level consumption of alcohol. In our experiment, data have been collected from local hospitals. Patients with hepatic disease dataset have been used from Kaggle [4] website. Collection of data from the primary data source is 4872. The collection of data set from a secondary source is 15,347. So with the help of total 20,219 datasets, an automated machine learning system have been built which can be useful to evaluate prediction algorithms and it is an initiative to reduce the burden of doctors. The quest for the right algorithm and hyper-parameter optimization is a time-consuming process if wanted to create a program that is based on conventional machine-learning. It is a manual and repetitive process. The best solution exists for this scenario is that Auto-WEKA is used, a package of WEKA software for automated machine learning. For a given problem, Auto-WEKA simultaneously selects learning algorithms and hyper-parameters appropriate to the problem for achieving improved performance. So in all our experiments, Configuration means that it has selected multiple learning algorithms and different hyper-parameters then applying multiple combinations of these to achieve results. The general form of the confusion matrix is presented in Table 1.

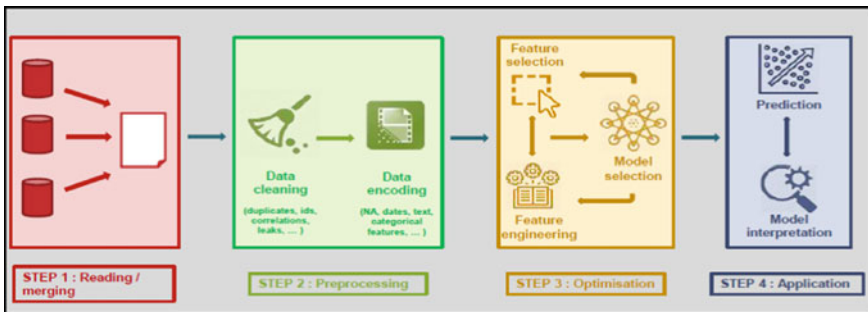


Fig. 2 Processes in AutoML [6]

Table 1 Confusion matrix

	Predicted: no	Predicted: yes
Actual: no	TN	FP
Actual: yes	FN	TP

**Table 2** Confusion matrix—Exp. 1

Patient	Healthy	Classified as
9323	3208	Patient = liver disease
4298	3390	Healthy = non-liver disease

$$\text{Accuracy} = (\text{TP} + \text{TN})/\text{Total} = (9323 + 3390)/(12713+20219) = 0.6287 \text{ (62.87\%)}$$

```

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
          0.772   0.413   0.652     0.772   0.707     0.366  0.731   0.702   non liver disease
          0.587   0.228   0.721     0.587   0.647     0.366  0.731   0.763   liver disease
Weighted Avg.  0.680   0.320   0.686     0.680   0.677     0.366  0.731   0.732

Temporary run directories:
C:\Users\User\AppData\Local\Temp\autoweka181477834081768825\
    
```

**Fig. 3** Detailed accuracy by class—35 min results

```

Classifier classifier = AbstractClassifier.forName("weka.classifiers.functions.SGD", new String[]{"-F", "0", "-L", "0.0029730794654194573", "-R",
"2.4672854956636744E-12", "-H"});
classifier.buildClassifier(instances);

Correctly Classified Instances      12713      62.8743 %
Incorrectly Classified Instances    7506      37.1257 %
Kappa statistic                    0.2575
Mean absolute error                 0.3713
Root mean squared error             0.6093
Relative absolute error             74.2515 %
Root relative squared error        121.8618 %
Total Number of Instances          20219
    
```

**Fig. 4** Accuracy with other parameters—35 min results

### 4.1 Experimental Setup-1:

Initially, a basic AutoML configuration setup has been used. A time limit of 35 min has been set for Auto-WEKA to find the right algorithms and hyper-parameters. For this time duration, Auto-WEKA tried 1059 configurations to get the best result. In this experiment as seen in Fig. 4, auto-WEKA tool got the best accuracy when choosing stochastic gradient descent (SGD) algorithm. The confusion matrix in Table 2 summarizes their classification results of data in the form of rows and columns (Fig. 3).

### 4.2 Experimental Setup-2

Auto-WEKA package focus on automating the steps of model selection and optimization of hyper-parameter. A time limit of 45 min has been set for Auto-WEKA to find the right algorithms and hyperparameters. For this time duration, Auto-WEKA tried

```

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.766    0.425    0.643     0.766   0.699     0.348  0.730    0.698    non liver disease
      0.575    0.234    0.711     0.575   0.636     0.348  0.730    0.761    liver disease
Weighted Avg.  0.671    0.329    0.677     0.671   0.668     0.348  0.730    0.730

Temporary run directories:|
C:\Users\User\AppData\Local\Temp\autoweka5188351282878020452\
    
```

Fig. 5 Detailed accuracy by class—45 min results

1493 configurations to get the best result. In this experiment as seen in Fig. 6, auto-WEKA tool got the best accuracy when choosing the logistic regression algorithm. It is advisable to give more time to Auto-WEKA for better results (Fig. 5).

The confusion matrix in Table 3 summarizes their classification results of data in the form of rows and columns.

```

Classifier classifier = AbstractClassifier.forName("weka.classifiers.functions.Logistic", new String[]{"-R", "8.441855801802037"});
classifier.buildClassifier(instances);

Correctly Classified Instances      13559      67.0659 %
Incorrectly Classified Instances    6660      32.9341 %
Kappa statistic                    0.3413
Mean absolute error                 0.4288
Root mean squared error             0.4559
Relative absolute error             85.7613 %
Root relative squared error         91.1757 %
Total Number of Instances          20219
    
```

Fig. 6 Accuracy with other parameters—45 min results

```

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.994    0.036    0.965     0.994   0.979     0.959  0.999    0.998    non liver disease
      0.964    0.006    0.994     0.964   0.979     0.959  0.999    0.998    liver disease
Weighted Avg.  0.979    0.021    0.979     0.979   0.979     0.959  0.999    0.998

Temporary run directories:
C:\Users\User\AppData\Local\Temp\autoweka2566127104543696565\
    
```

Fig. 7 Detailed accuracy by class—60 min results

Table 3 Confusion matrix—Exp. 2

Patient	Healthy	Classified as
7748	2362	Patient = liver disease
4298	5811	Healthy = non-liver disease

$$\text{Accuracy} = (TP + TN) / \text{Total} = (7748 + 5811) / (13,559 / 20,219) = 0.6706 \text{ (67.06\%)}$$

```
Classifier classifier = AbstractClassifier.forName("weka.classifiers.trees.RandomForest", new String[]{"-I", "10", "-K", "0", "-depth", "0"});
classifier.buildClassifier(instances);

Correctly Classified Instances      1975      97.9042 %
Incorrectly Classified Instances    424      2.0958 %
Kappa statistic                    0.9581
Mean absolute error                 0.1557
Root mean squared error             0.2049
Relative absolute error             31.1377 %
Root relative squared error         40.9761 %
Total Number of Instances          20219
```

Fig. 8 Accuracy with other parameters—60 min results

### 4.3 Experimental Setup-3:

A time limit of 60 min has been set for Auto-WEKA to find the right algorithms and hyper-parameters. For this time duration, Auto-WEKA tried 1958 configurations to get the best result. In this experiment, as seen in Fig. 8, auto-WEKA tool got the best accuracy when choosing random forest algorithm. The time has also been increased to 90, 100, 150 min but found that the results are almost similar to experiment 1 and 2. If the time increased more than 60 min the accuracy was not increasing. This is an example of overfitting and found that to achieve optimal values, it is advisable to give sufficient time to Auto-WEKA but there should be a threshold to avoid overfitting (Fig. 7).

The confusion matrix in Table 4 summarizes the classification results of data in the form of rows and columns. The following terminologies are useful to determine the accuracy score of the model. Precision, Recall and *F1* score are calculated as per formulas in 1, 2 and 3 respectively. All the calculation for the experiments is summarized in Table 5.

$$\text{Precision} = (\text{True Positive}) / (\text{True Positive} + \text{False Positive}) \tag{1}$$

Table 4 Confusion matrix—Exp. 3

Patient	Healthy	Classified as
10,049	61	Patient = liver disease
363	9746	Healthy = non-liver disease

$$\text{Accuracy} = (\text{TP} + \text{TN}) / \text{Total} = (10,049 + 9746) / (19,795 / 20,219) = 0.9790 \text{ (97.90\%)}$$

Table 5 Precision, recall, *F1* score and accuracy scores of experiments

Exp. No.	Precision	Recall	<i>F1</i> score	Accuracy (%)
1	0.744	0.685	0.713	62.87
2	0.766	0.643	0.699	67.06
3	0.994	0.965	0.979	97.90

$$\text{Recall} = (\text{True Positive})/(\text{True Positive} + \text{False Negative}) \quad (2)$$

$$F1 \text{ Score} = 2 * (\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \quad (3)$$

## 5 Conclusion

AutoML is still an active area of research. Automated machine learning (AutoML) can be used for classification, regression and configuration of neural networks. In our experiments, the time constraint is found to be an important factor in autoML to achieve higher accuracy. This is our proposed properly configured AutoML model according to our need. In [7], there are various frameworks available to work upon. In autoML, timing is found to be a critical factor to achieve good accuracy and one should try AutoML for different timings and then choose a timeslot which provides higher accuracy. Globally, the demand for Data Scientists is increasing year by year. Many companies facing problems for hiring talented data scientists because they are expensive. AutoML provides a way to bridge the gap in the data science industry. AutoML saves time for highly skilled experts so that the time would be utilized in other areas of the project.

## References

1. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The WEKA data mining software: an update. *ACM SIGKDD Explor Newsl* 11(1):10–18
2. Schaul T, Bayer J, Wierstra D, Sun Y, Felder M, Sehnke F, Rückstieß T, Schmidhuber J (2010) PyBrain. *JMLR*
3. Yao Q, Wang M, Chen Y, Dai W, Yi-Qi H, Yu-Feng L, Wei-Wei T, Qiang Y, Yang Y Taking human out of learning applications: a survey on automated machine learning. [arXiv:1810.13306](https://arxiv.org/abs/1810.13306)
4. <https://www.kaggle.com/>
5. Christopher N New automatic diagnosis of hepatic status using bayesian classification
6. [https://www.slideshare.net/AxeldeRomblay?utm\\_campaign=profiletracking&utm\\_medium=sssite&utm\\_source=ssslideview](https://www.slideshare.net/AxeldeRomblay?utm_campaign=profiletracking&utm_medium=sssite&utm_source=ssslideview)
7. Zoller M, Huber MF Benchmark and survey of automated machine learning frameworks. [arXiv: Learning](https://arxiv.org/abs/1810.13306)

# BS6 Violation Monitoring Based on Exhaust Characteristics Using IoT



B. S. Pavithra and K. A. Radhakrishna Rao

**Abstract** An IoT-based continuous emission tracking and warning system are proposed which consists of Arduino ATMEGA328 processor, a Wi-Fi ESP8266 interface, gas sensor MQ135, LCD module and BLYNK application. It is placed at the exhaust of the vehicle to continuously monitor the exhaust characteristic data and is collected in the cloud server for further processing and alerting the vehicle owner. The prototype is monitored for the emission characteristics of BS6 standards. The BS6 standard is taken as reference as the existing BS4 standards have been withdrawn since April 2020. The objective of the paper is to introduce a vehicular emission monitoring and warning system and to ensure the pollutants limits are within BS6 standards. The prototype is capable of detecting and monitoring the pollutants such as CO, CO<sub>2</sub> and NO<sub>x</sub> which are majorly found in the exhaust fumes. The measured data is shared through LCD and a BLYNK notification to the vehicle owner to continuously monitor the exhaust. This system is real-time, Portable, low cost and it provides a good interface with the application resulting in controlling the emission especially in the urban areas.

**Keywords** BS6 violation · Emission · Arduino · WIFI ESP8266 · MQ135 · BLYNK · LCD module

## 1 Introduction

Among the environmental pollutions, air pollution is one of the most crucial pollutions which is rapidly increasing due to the human urbanization which has a direct

---

B. S. Pavithra (✉) · K. A. Radhakrishna Rao  
Department of Electronics and Communication, PES College of Engineering, Mandya,  
Karnataka, India  
e-mail: [pavithrabs09@gmail.com](mailto:pavithrabs09@gmail.com)

K. A. Radhakrishna Rao  
e-mail: [karkrao@gmail.com](mailto:karkrao@gmail.com)

hazardous impact on health and environment. Hence air quality monitoring is extremely important in our day to day living. Air pollution from motor vehicles, power plants, industry, and households account for 70% of the pollution in the world with the composition of the quantity of carbon monoxide (CO) 99%, Carbon dioxide 89%, and nitrogen oxides (NOx) 73% and other particulates that include lead, sulphur oxide and dust particles [1]. Hence with the increase in the digital miniaturization makes the devices smarter and the evolution of the internet of things and cloud computing has made the lives better [2].

Emission of hazardous gases to the environment is mainly from the vehicles due to the incomplete combustion of the fuels in the engine of a vehicle leads to an increase in the pollution and unfavourably affecting the environment. The greenhouse effect occurs due to the increase in air pollution. Emission of these gases from the vehicles cannot be completely avoided but, to be controlled. The main purpose of pollution monitoring is not only to monitor the vehicle condition based on the exhaust and provide the exhaust data to the user but also to make the effort to improve the environment condition to reduce the Global warming. Monitoring the vehicles to check whether they are adhering to the BS6 norms of the government which helps to identify the vehicles which emit more pollutants to the environment along with the fraudulent activities by the vehicle owners and the fuel stations.

Emissions from many air pollutants like industries, vehicles and household activities have various negative effects on human health and the environment which could even lead to natural calamities. BS-VI standard vehicles include enhanced On board diagnostics (OBD) requirements for all vehicle classes, for two and three-wheeled vehicles, and the introduction of emission limits on nitrogen oxides which is the major reductions compared to BS4 standards. Also the sulphur content in the fuels.

## 2 Related Work

In this paper, a low price radio frequency identification (RFID) technology system is developed to detect the emission from the vehicles along with the notification system. This system, in turn, helps in monitoring the vehicle engine health from the emission values captured. This system works on the concept that when the vehicles stop in the traffic signals in the urban areas, reliable reading of air ratio from a vehicle is taken through RFID, which detects the vehicle emission level. So from the values obtained in the real-time the vehicle emission levels can be monitored and controlled effectively [3].

To reduce environmental pollution due to the emission of carbon particles from automobiles, a chemical sensor is used to keep track of the emission level using Wireless Sensor Networks [WSN] technique. From the results of the experiment conducted, it is concluded that by continuous monitoring of the carbon particles emission helps in controlling the air pollution from vehicles [4].

Using Wireless Inspection and Notification system (WINS) with the concept of ZigBee and GSM, a system is developed which is used to monitor the vehicles that



are violating the emission standards. In this paper, mature wireless communication techniques like a radiofrequency module (ZigBee) and GSM are adopted to collect and commute the emission information from vehicles [5].

In this paper [6] the main aim is to reduce the emission of harmful gases like carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO<sub>2</sub>), hydrocarbon (HC), and sulphur dioxide (SO<sub>2</sub>) from the road vehicles. Here the system is developed in stages. At first stage catalyst based accurate monitoring optic absorption spectroscopy based gas sensor and ADXL210 accelerometer chip system is developed. It monitors the emission from the engines at different states and conditions.

This paper [7] mainly focus on reducing the cost of the expensive sensors used to measure the harmful gases emitted and matter concentration from the combustion engines. Control system of EGR (Exhaust Gas Recirculation) is included along with Emission monitoring system in the proposed system which involves the temperature, rotational speed and airflow sensor which has a highly flexible electronic valve in operation to measure the accurate concentration of gases.

The paper [2] explains with the increase in the digital miniaturization makes the devices smarter and the evolution of the internet of things and cloud computing has made the lives. This improvisation in the devices is useful in the control of different environmental hazardous.

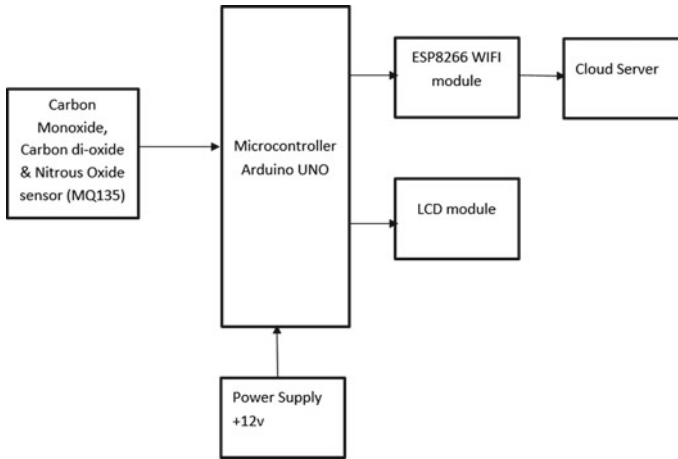
In the paper [8], Automation of the greenhouse environment is explained using the Internet of things. The environment of a greenhouse is made of a wall of transparent roof maintaining the good climatic condition, which ensures the plant's growth requiring the soil moisture, sunlight, temperature and humidity. This technique helps in the improving of the environmental pollution which in turn keeps the human health.

### 3 Methodology

The MQ135 gas sensor is an air quality sensor used for detecting the NO<sub>x</sub>, smoke, CO, CO<sub>2</sub>. Fast response and high sensitivity made its usage vast. It has a stable long life and simple drive circuit. It requires power supply ranging from 2.5 to 5.0 V. Hardware connections are made as follows: V<sub>cc</sub>–5 V, GND–GND, and A0–Analog0 (Fig. 1).

The microcontroller board used is the ATmega328P. It has 14 digital which can be used as both input and output pins and out of 14 pins 6 can be used as Pulse width modulation (PWM) outputs, 6 analog inputs, a 16 MHz quartz crystal, a Universal serial bus (USB) connection, a power jack, an ICSP header and a reset button. Flash memory of 32 KB with Operating voltage is 5v and clock speed is up to 16 MHz. The Uno can be programmed with the Arduino Software. It communicates using the original STK500 protocol. The ATmega328 supports I2C Two-way interfacing (TWI) and Serial peripheral interfacing (SPI) communication.

ESP8266 is a low cost, compact and powerful WiFi module used for adding WiFi functionality via a UART serial connection. It requires 3.3 V power. A Liquid Crystal Display is used to display the exhaust values from the vehicles which cross



**Fig. 1** The proposed system block diagram

the threshold value. Blynk server is used for communications between mobile phones and Hardware used in the system.

The process flow of the system is explained in Fig. 4. After the device initialization, exhaust gases are sensed through a sensor and processed and compared in the microcontroller. Exhaust gases level is compared with the threshold values which is detected keeping BS6 constraint during testing if the exhaust value is crossed then the data is transferred and stored in the server and also displayed on the LCD module along with a BLYNK notification to the user mobile phone. The process is real-time and continuous (Fig. 2).

## 4 Implementation

The system comprises a sensing module with MQ-135 sensor. The sensor is used to sense the Carbon Monoxide, Carbon dioxide and Nitrous oxide gases exhaust from the vehicle. The microcontroller used in the proposed system is Arduino-UNO processor. The data is received from the sensor MQ-135, the Arduino-UNO processor processes and compares the CO, CO<sub>2</sub> and NO<sub>x</sub> values with the threshold values which are taken based on the testing results of BS6 engine vehicle using BS6 fuel. If the emission is above the threshold value, then the data is passed to the cloud server via the ESP8266 WIFI module. The ESP8266 WIFI module is used for communication between the system and the Cloud server using BLYNK application (Fig. 3).

Exhaust smoke of the vehicle is passed to MQ135 sensors. The sensors pass the level of Carbon monoxide, Carbon dioxide and Nitrous oxide particles from the vehicle emission. The values received by the microcontroller and the microcontroller compares the values of the carbon monoxide, Carbon dioxide and Nitrous oxide

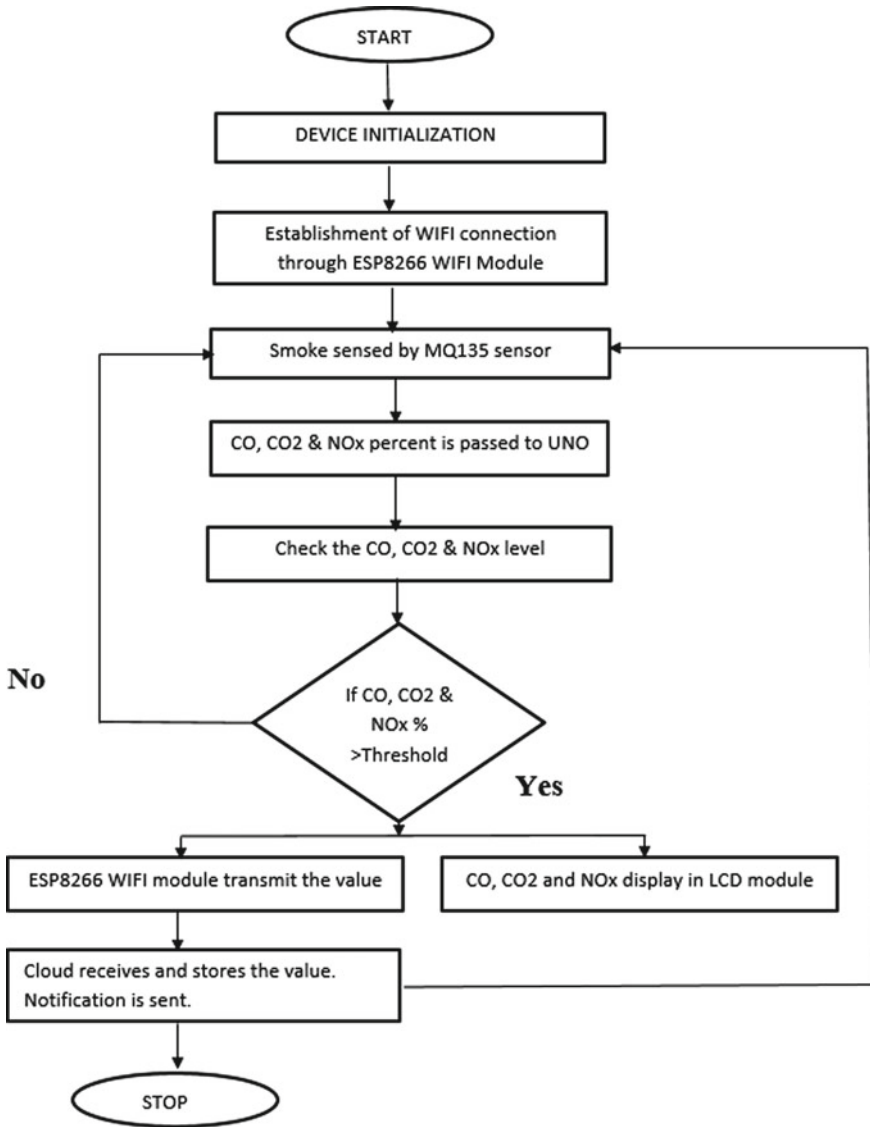


Fig. 2 The proposed system flow chart

received from the sensors with the threshold values. The system receives the emission values from the sensor node. It compares the emission values with the BS6 emission standards set as threshold values (testing results). If the exhaust value exceeds the threshold value then the value is transmitted through the wireless mode and stored in the cloud server. A message notification will be sent to the vehicle owner as a warning if the emission level crosses the threshold value along with the LCD of the

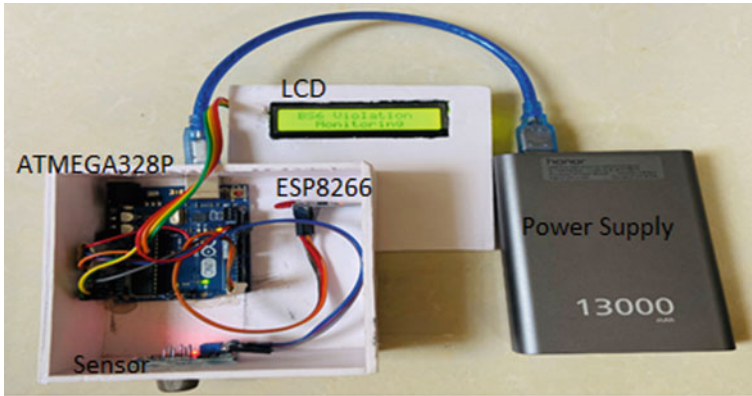


Fig. 3 Implementation setup

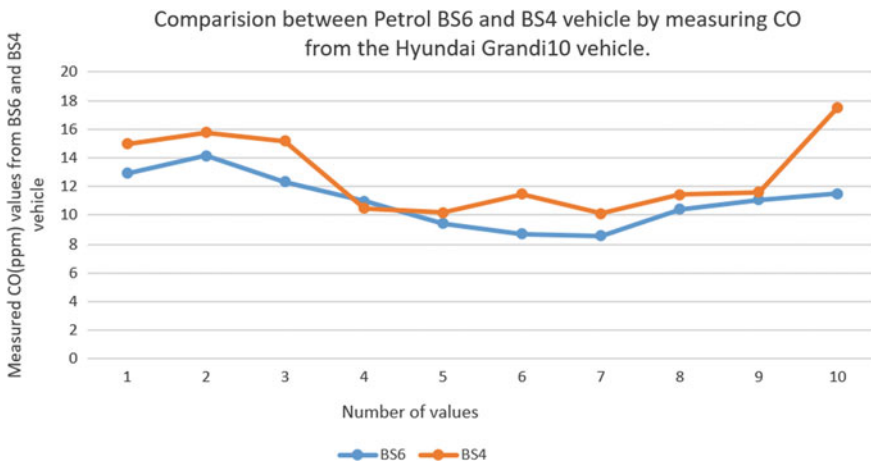
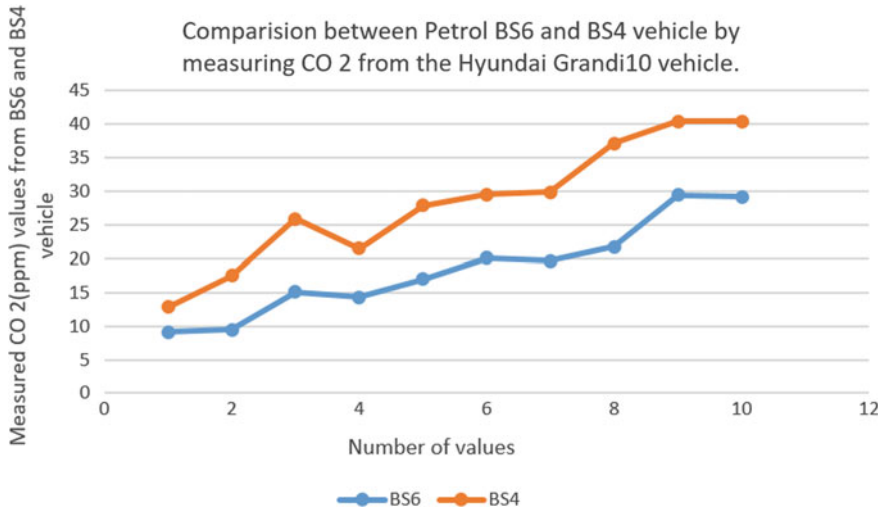


Fig. 4 Graph of CO with 10 samples tested from the Hyundai Grandi10 petrol vehicle

values. If there is any fraud occurs without the knowledge of the vehicle owner to the fuel, could be detected from the exhaust characteristics report stored in the server and report to the concerned authority for further action. The system is portable, less time-consuming and real-time.

### 5 Results and Discussion

Having a true objective to ensure the proper evaluation of the data obtained from the testing, many trials were carried out. For the trails, values are observed with a



**Fig. 5** Graph of CO<sub>2</sub> with 10 samples tested from the Hyundai Grandi10 petrol vehicle

different time period of the vehicle. To check the difference between the emission values from the vehicles having the constraint of BS4 and BS6 vehicle. The vehicle used for the analysis is Hyundai Grandi10 Petrol and diesel model. Emission values are taken from the vehicles in two different cases and are plotted as shown in Figs. 4 and 5. Firstly, testing is conducted with petrol BS4 and BS6 vehicles and next is diesel BS4 and BS6 vehicles. In the graphs, the  $x$ -axis is marked with the number of times testing is conducted for CO, CO<sub>2</sub> and NO<sub>x</sub> particles. The  $y$ -axis is marked with the gases in parts per million (ppm).

10 trails were conducted for the petrol BS4 vehicle and the 10 different values were plotted in the graph for each gas. A similar procedure is conducted with BS6 petrol vehicle. 10 trails were conducted for the diesel BS4 vehicle and the 10 different values were plotted in the graph for each gas. Similarly, the procedure is conducted with a BS6 diesel vehicle.

From the graph, Carbon monoxide gas emission values are captured and analyzed from the vehicle. By comparing the values collected from two different standard vehicles, it is observed that with the usage of the BS6 standard petrol vehicles pollution can be controlled than the BS4 standard petrol vehicles (Fig. 4).

From the graph, Carbon dioxide gas emission values are captured and analyzed from the vehicle. By comparing the values collected from two different standard vehicles, it is observed that with the usage of the BS6 standard petrol vehicles pollution can be controlled than the BS4 standard petrol vehicles (Figs. 5 and 6).

From the graph, Nitrous oxide gas emission values are captured and analyzed from the vehicle. By comparing the values collected from two different standard vehicles, it is observed that with the usage of the BS6 standard petrol vehicles pollution can be controlled than the BS4 standard petrol vehicles (Figs. 6 and 7).

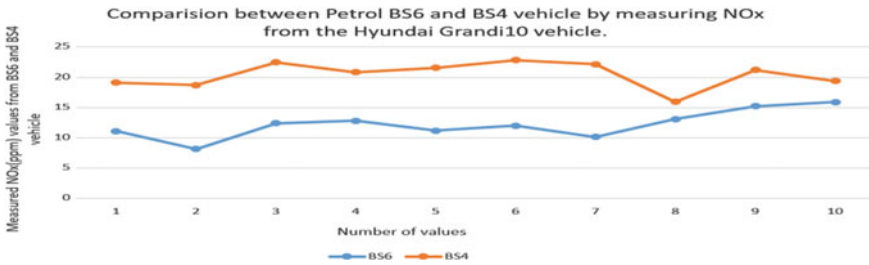


Fig. 6 Graph of NOx with 10 samples tested from the Hyundai Grandi10 petrol vehicle

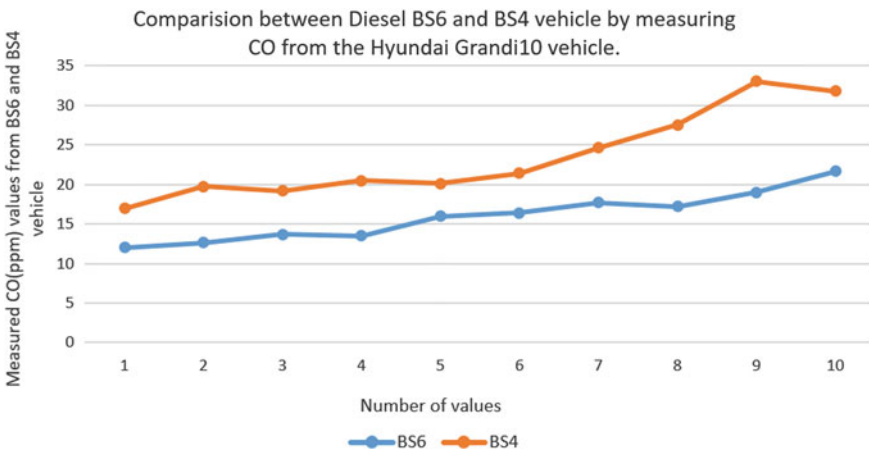


Fig. 7 Graph of CO with 10 samples tested from the Hyundai Grandi10 diesel vehicle

From the graph, Carbon monoxide gas emission values are captured and analyzed from the vehicle. By comparing the values collected from two different standard vehicles, it is observed that with the usage of the BS6 standard diesel vehicles pollution can be controlled than the BS4 standard diesel vehicles (Figs. 7 and 8).

From the graph, Carbon dioxide gas emission values are captured and analyzed from the vehicle. By comparing the values collected from two different standard vehicles, it is observed that with the usage of the BS6 standard diesel vehicles pollution can be controlled than the BS4 standard diesel vehicles (Figs. 8 and 9).

From the graph, Nitrous oxide gas emission values are captured and analyzed from the vehicle. By comparing the values collected from two different standard vehicles, it is observed that with the usage of the BS6 standard diesel vehicles pollution can be controlled than the BS4 standard diesel vehicles (Fig. 9).

Threshold values of each gas are set in the software to have a check on the emission level from the vehicles. The emission threshold value is set by collecting the 100 sample values of the two different standards (BS4 and BS6) petrol and diesel vehicle emission, by averaging and adding +50 ppm as a variance value to it. Vehicle taken

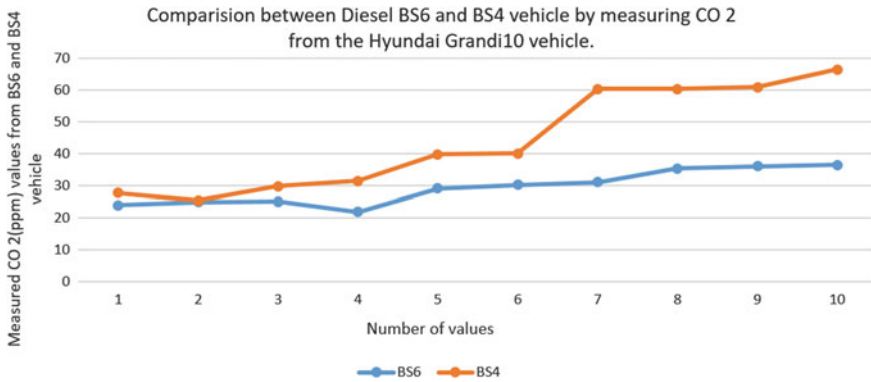


Fig. 8 Graph of CO<sub>2</sub> with 10 samples tested from the Hyundai Grandi10 diesel vehicle

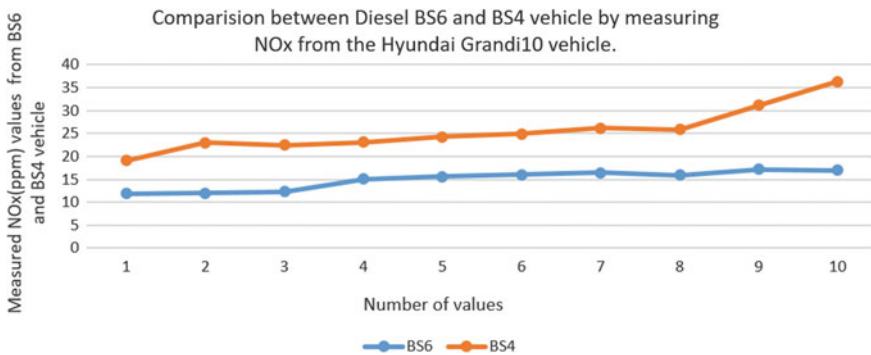


Fig. 9 Graph of NO<sub>x</sub> with 10 samples tested from the Hyundai Grandi10 diesel vehicle

for testing is BS4 Hyundai Grandi10 with BS4 Petrol and Diesel fuels and BS6 Hyundai Grandi10 with BS6 Petrol and Diesel fuels. Whenever the emission value crosses the threshold value, a message notification is sent to the vehicle owner mobile through the BYLNK application, also an LCD message is displayed.

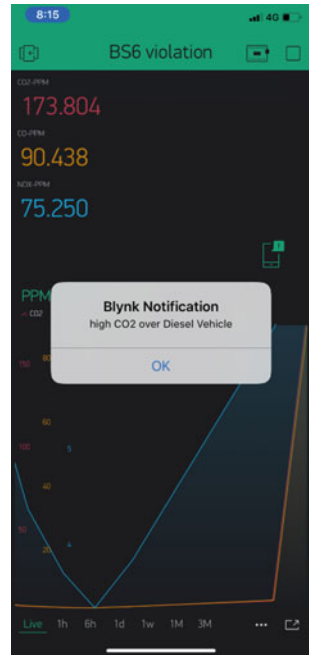
Figures 10 and 11 shows the warning notification in the mobile phone and Figs. 12 and 13 shows the warning message displayed in the LCD showing the gases value.

Above figure depicts the warning notification sent to the vehicle user mobile phone when the emission value of the CO<sub>2</sub> gas increases beyond the threshold value set indicating the increased level of the emission crossing the standard limits set (Fig. 10).

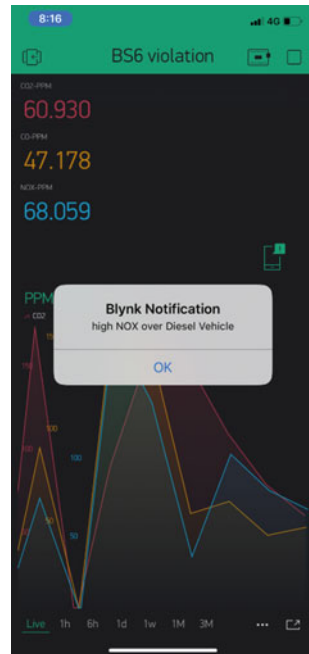
Above figure shows the warning notification sent to the vehicle user mobile phone when the emission value of the NO<sub>x</sub> gas increases beyond the threshold value set indicating the increased level of the emission (Fig. 11).

Along with the BLYNK notification to the user mobile phones, a warning message is displayed in LCD to alert the vehicle user regarding the emission of the gases

**Fig. 10** Warning notification to the vehicle user mobile phone through BLYNK application for CO<sub>2</sub> gas



**Fig. 11** Warning Notification to the vehicle user mobile phone through BLYNK application for NO<sub>x</sub> gas







**Fig. 12** Warning message displayed in LCD for CO<sub>2</sub> gas



**Fig. 13** Warning message displayed in LCD for NO<sub>x</sub> gas

crossing the threshold value set. The above figure shows the Carbon dioxide gas emission value as it crosses the threshold value (Fig. 12).

Along with the BLYNK notification to the user mobile phones, a warning message is displayed in LCD to alert the vehicle user regarding the emission of the gases crossing the threshold value set. Above figure shows the Nitrous oxide gas emission value as it crosses the threshold value (Fig. 13).

## 6 Conclusions

The system can be dynamically and continuously used for vehicle emission testing. This system will overcome the drawback of the static vehicle emission center by frequent monitoring of the exhaust from vehicles which helps in minimizing the release of pollutants to the environment.

This method helps in tracking and controlling the vehicles with high emission and with the monitoring of the vehicles whether it adheres to BS6 norms. The fuel

fraudulent activities from the stations can be avoided on tracking the exhaust data stored in the server.

**Acknowledgements** I heartily thank my project guide, for his support and guidance during my project work to develop the system.

Also I thank all teaching staffs of the department for their constant encouragement, support and guidance in the work.

## References

1. Ervin K (2011) Design of carbon monoxide (Co) gas disposal device detector in motor vehicles. Semarang
2. Sivaganesan D (2019) Design and development Ai-enabled edge computing for intelligent-Iot applications. *J Trends Comput Sci Smart Technol (TCSST)* 1(02):84–94
3. Kachavimath AV (2015) Control of vehicle effluence through internet of things and android. *IJCST* 3(5):148–153
4. Dakshayini M, Kiran Kumar S (2014) Monitoring of air pollution using wireless sensor networks. *IJCTA* 5(3):1213–1218
5. Akash SJ, Prasad K (2016) Design and implementation of web based vehicle emission monitoring and notification system using ZigBee and GSM. *IJRET* 3(5):201–204
6. Wei C, Zhuang Z, Ewald H, Al-Shamma'a AI (2012) Harmful exhaust emissions monitoring of road vehicle engine. In: IEEE on 13 Feb 2012 at 2012 fifth international conference on intelligent computation technology and automation
7. Graba M, Bieniek A (2013) Monitoring system of harmful substances emission at compression ignition engine with exhaust gas recirculation. In: IEEE on 18 July 2013 at 2013 international symposium on electrodynamic and mechatronic systems (SELM)
8. Raj JS, Ananthi JV (2019) Automation using Iot in greenhouse environment. *J Inf Technol* 1(01):38–47
9. Leelaram T, Ramya T (2005) RFID based vehicle emissions in cities on internet of things. *IJRMEET* 3(2):1–8

# Analytical Classification of Sybil Attack Detection Techniques



Ankita S. Koleshwar, S. S. Sherekar, V. M. Thakare, and Aniruddha Kanhe

**Abstract** Secured network is an important issue in today's era. The various defense techniques have been developed to protect the network from the attacks which hampers the performance of the network. The attacks like Blackhole, Identity Spoofing, Sybil, etc. are observed. In such types of attacks, the Sybil is the destructive attack. In Sybil attack, attacker misleads the other nodes by viewing the wrong identity of the users who are alert from the nodes in the network. In a Sybil attack, a node illegitimately claims various identities. Sybil attack intimidates network in routing, fair resource allocation, data aggregation and misbehavior detection. This paper focuses on a systematic study of Sybil attack and gives the critical analysis of the various parameters which affects the performance along with the advantages and limitations.

**Keywords** MCC · Security issues · Classification of Sybil attack · Advantages · Limitations · Countermeasures

## 1 Introduction

The “mobile cloud computing” concept was proposed after “cloud computing” in 2007. Mobile cloud computing works based on both data storage and the data where the process undergoes outside the mobile device. Nowadays, the environment of mobile cloud computing is too conspicuous to attackers.

Today, the mobile cloud computing (MCC) provides a plenty of services in many sensitive fields such as online transactions, business sector and banking,

---

A. S. Koleshwar (✉) · S. S. Sherekar · V. M. Thakare  
Department of Computer Science and Engineering, SGB Amravati University, Amravati, India  
e-mail: [ankita.koleshwar12@gmail.com](mailto:ankita.koleshwar12@gmail.com)

A. Kanhe  
Department of Electronics and Communication Engineering, National Institute of Technology  
Puducherry, Puducherry, India

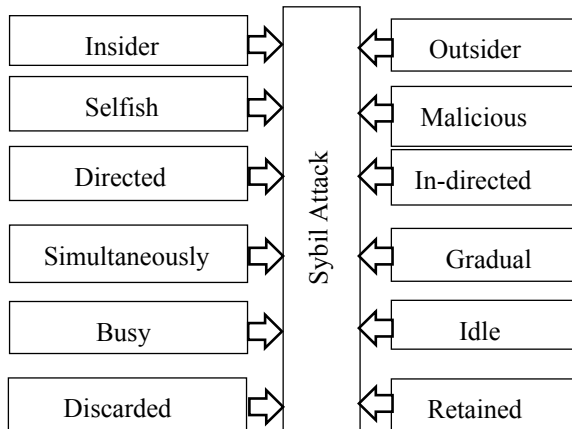
which demands the high-security-related vulnerabilities through the mobile cloud computing platform [1]. Nowadays, a single surfing data is also not secured on the Internet, and so many people suffer from severe attacks and security breaches. Sybil attack is one of the destructive attacks among various attacks [2]. In the Sybil attack, the attacker creates one or more identity on a single physical device to control one or more identities [3]. In this attack, an attacker generates a huge amount of fake identities and can be said as Sybil identities, and since all the Sybil identities are restricted by the attacker, he can unkindly plan a valuable number of false statements into the system and challenge it, by creating the decisions useful for him. Basically, Sybil attacks smash the trust mechanism at the back of the systems.

Security plays a major role in MCC; there are many introduced techniques to detect any defense against the Sybil attack. The main motive behind the threats in networking is because of its open structural design and its vulnerability to various attacks. Hence, a malicious user could create multiple fake IDs to achieve sensitive information from users. Due to Sybil attack, [4] OSN, mobile networks and wireless sensor network are affected in the manner of honesty. Hence, the study on various attacks is necessary to improve security and protect the users. Rest of the paper is organized in the following manner: Sect. 2 describes the related work, Sect. 3 is about discussion, and Sect. 4 concludes the paper.

### 1.1 Classification of Sybil Attack

In peer-to-peer system, the first proposed attack is Sybil attack [5, 6]. The attacker changes the status of the system of the P2P network by constructing an enormous amount of forged ids. The taxonomy of various kinds of Sybil attacks is conferred, and the ability of the attacker is achieved by numerous characteristics which are given in Fig. 1.

Fig. 1 Classification of Sybil attack



- Insider versus Outsider: The efficiency of initiating Sybil attack is achieved easily by both attackers.
- Selfish versus Malicious: The attacker which provides the forged information for own profit is a selfish attacker, and the attackers which are challenging the system are malicious attackers.
- Directed versus In-directed Communications: The intruder can simply correspond with the real node by employing a Sybil id or utilizing the actual id to communicate with other nodes and to retrieve the direction for Sybil information by such authentic id.
- Obtaining Sybil ids, Simultaneously versus Gradually: The intruder may acquire the Sybil ids concurrently or may increasingly create them step by step.
- Busy versus Idle: Scattered system may contribute by all Sybil ids, or some of them may not work, while remaining is in an active state.
- Discarded versus Retained: Handling of old Sybil identities are very important for the attacker. After receiving a Sybil node, one could recognize another by monitoring the maintained communication with an expected node and the detected Sybil node.

## 2 Related Work

In related work, the Sybil node detection is given. A Sybil node uses various fake identities at the same time, and it will hamper the network traffic, and data packet will never achieve the destination. In this paper, different methods of detection of the Sybil attack node are analyzed.

In [7], Sybil attack detection mechanism is given which is based on RSS to protect the system against the Sybil attacks. There is no requirement of extra hardware, and it performs on the MAC layer using 802.11 protocol. In [8], it also gives the RSS-based Sybil attack detection technique with the validation of node which exactly categorizes the Sybil id with HTTP. MAC is used to authenticate the node. Legitimate nodes get permitted to come into the network if the node is authenticated. In paper [9], in MANET, to prevent the Sybil attack, calculation of the efficiency of available authentication methods is given. This paper gives the authentication model for MANET which develops the hardware identity of the device of each node for verification. [10] presented the hash function-based detection mechanism. The messages through the hash function are received by every node. Sybil attackers are identified by confirming the hash received by the side of the message. Node gets hash of correspondent and evaluates it with the earlier hash received in Hello message to validate the identity.

This paper reviews the various Sybil attack detection methods which are based on different behavior characteristics of the Sybil users [11]. It gives the detailed analysis of the Sybil attack detection methods as given in Table 1. Binary comparison between different Sybil attack detection techniques based on various parameters is shown in Table 2.

**Table 1** Comparative analysis of various Sybil attack detection technique

Author	Network	Algorithm name	Detection mechanism	Routing protocol	Simulator/tool used	Detection accuracy	Advantages	Limitations	Countermeasures
Sharmila et al. [12], 2012	WSN	Node ID-based algorithm	Node ID-based detection	AODV	NS2	90%	It consumes less energy and duration of detection time is 75 s	Not specified	Unique physical ID
Alsaedi et al. [13], 2015	Clustered WSN	Not specified	Energy trust system model	Election protocol	OMNet ++	>87%	The system is efficient to detect the Sybil attack	Not specified	Energy trust system
Prabhjotkaur et al. [14], 2016	WSN	Not specified	Misuse detection system	Clustered-based	Not specified	Low	Effective system	It is not able to detect other routing attacks	Centralized IDS scheme
Moradi et al. [15], 2016	WSN	Algorithm based on development phase and keeping phase	Distributed method based on mobile agent	AODV, CSMA CA	C# NET	Good	Removes the adversary nodes from participation in routing while using mobile nodes and increases the security in a network	Not specified	Not specified

(continued)

Table 1 (continued)

Author	Network	Algorithm name	Detection mechanism	Routing protocol	Simulator/tool used	Detection accuracy	Advantages	Limitations	Countermeasures
Lakhanpal et al. [16], 2016	Adhoc Network	MAC and MAP Algorithm	Hybrid approach Of MAC and MAP algorithm	AODV	MATLAB	93.98%	This technique is able to detect the Sybil attacks	Not specified	MAC and MAP algorithm
Evangelista et al. [17], 2016	IoT	Not specified	Lightweight Sybil attack detection framework	802.15.4	NS3	88%	Not specified	Not specified	Not specified
Singh et al. [18], 2017	WSN	TBID technique	Trust-based identity detection	DSDV	NS2	Good	Nodes having average trust value are detected as Sybil nodes which are isolated from a sensor network	Implementation is done in NS2 with 42 nodes only	TBID technique

(continued)

**Table 1** (continued)

Author	Network	Algorithm name	Detection mechanism	Routing protocol	Simulator/tool used	Detection accuracy	Advantages	Limitations	Countermeasures
Kavitha et al. [19], 2014	MANET	NDD algorithm	NDD algorithm-based detection	Geographic routing protocol	NS2	Good	No data damage or Failure, detect both the join and leave and direct Sybil attackers	Not specified	NDD algorithm
Ramesh et al. [21], 2016	MANET	Attracter detection algorithm	Transmission time-based detection	802.11 protocol	C#	Good	Decreases the false positive rate and increases the throughput of the network	Not specified	Not specified
Chang et al. [22], 2013	Mobile cloud computing	CbTMS	Cloud-based trust management system	GPRS	OMNet ++	Good	Not specified	Not specified	Not specified



**Table 2** Binary comparison of various Sybil attack detection techniques

Author	Supports Energy	Delay Time	Throughput	Packet Delivery Rate	Avg. No. of samples	BW utilization	Accuracy
Sharmila et al. [12], 2012	✓	×	✓	✓	✓	×	✓
Alsaedi et al. [13], 2015	✓	×	×	✓	✓	×	✓
Prabhjotkaur et al. [14], 2016	×	×	×	×	×	×	×
Moradi et al. [15], 2016	×	✓	✓	×	✓	×	✓
Lakhanpal et al. [16], 2016	×	×	✓	×	×	×	✓
Evangelista et al. [17], 2016	×	×	×	×	×	×	×
Singh et al. [18], 2017	×	✓	✓	✓	✓	✓	×
Kavitha et al. [19], 2014	×	×	✓	✓	✓	×	✓
Vasudeva et al. [20], 2012	×	×	×	×	×	×	×
Ramesh et al. [21], 2016	×	×	×	×	×	×	×
Chang et al. [22], 2013	✓	×	✓	✓	✓	×	✓

### 3 Analysis and Discussion

Sharmila [12] presented a node ID-based algorithm for detecting the Sybil attack. The node registration is performed in two different phases, and the node ID is allocated dynamically. The detection mechanism is analyzed based on the PDR. It is observed that the PDR of the network has improved after the implementation of the algorithm and the detection of Sybil nodes. This proposed scheme gives the accuracy of 90%, and it consumes less energy. The countermeasure for this scheme is its unique physical ID. Hence, the scheme may be used for the high-security environment.

Alsaedi [13] presented a system based on hierarchical trust detection system to detect the Sybil attack in WSN. This system has three levels to check the ID, location and trust evaluation which is based on the energy of the sensor nodes. Prabhjotkaur [14] a misuse detection system is proposed based on the clustered routing protocol.

The proposed centralized IDS based on misuse detection to detect the malicious cluster head which has the intention of causing the Sybil attack in the wireless sensor network. Any kind of malicious activity in the network will be detected by the base station in the network.

Moradi [15] gives the algorithm which detects and prevents the Sybil attack in the network with two phases. The first phase is network improvement phase, and the second phase is the network maintenance phase. Lakhnpal [16] presented the approach of MAC and MAP algorithm to detect and prevent the Sybil nodes. MAC address is the special identifier address which allocates to the nodes on a physical network segment.

Evangelista [17] strengths and weaknesses of Sybil attack detection are focused when applied in the IoT content dissemination. It is based on the lightweight Sybil attack detection mechanism. Singh [18] proposed a mechanism which is trust-based identity detection. It calculates the trust values of adjacent sensor nodes. The nodes having trust values < the threshold value is considered as Sybil node. The mechanism gives high performance for the factor of throughput, PDR, delay and overhead. Kavitha, [19] proposed neighbor discover distance (NDD) algorithm for Sybil attack detection in MANET. By using this NDD algorithm, the data will be transferred without any damage or loss, from source to destination.

Vasudeva [20] proposed a lowest ID clustering algorithm for detection Sybil attack. In [20], Sybil attack has been illustrated by two different ways: First is lowest ID-based Sybil attack and second is impersonation-based Sybil attack. The paper addresses the challenges of the Sybil attack on the routing protocol. Ramesh [21] proposed transmission time-based detection mechanism for protection against Sybil attack in the network. The algorithm is based on the MAC layer with 802.11 protocol without the burden for any extra hardware. Chan [22] presented a trust management scheme to detect a Sybil attack in the cloud. Cloud-based trust management scheme (CbTMS) performs a characteristic check of Sybil attack and trustworthiness to clarify coverage nodes.

## 4 Conclusion

Numerous personal records and sensitive information are placed openly in today's networks which are vulnerable to Sybil attacks. This paper focuses on security issues, classification of Sybil attack and systematic study of Sybil attack and gives the critical analysis of the various parameters which affect the performance along with the advantages and limitations. It gives the binary comparison between different Sybil attack detection techniques based on various parameters, and there are various Sybil detection techniques discussed in this paper.

## References

1. Sarigiannidis P, Karapistoli E, Economides AA (2016) Analysing indirect sybil attacks in randomly deployed wireless sensor networks. In: 2016 IEEE 27th annual international symposium on personal, indoor, and mobile radio communications (PIMRC), Sept 2016, pp 1–6
2. Al-Qurishi M, Al-Rakhami M, Alamri A, Alrubaian M, Rahman SMM, Hossain MS (2017) Sybil defense techniques in online social networks: a survey. *IEEE Access* 5:1200–1219
3. Sharmila S, Umamaheswari G (2012) Detection of sybil attack in mobile wireless sensor networks. *Int J Eng Sci Adv Technol* 2(2):256–262
4. Valarmathi M, Meenakowshalya A, Bharathi A (2016) Robust sybil attack detection mechanism for social networks-a survey. In: 2016 3rd International conference on advanced computing and communication systems (ICACCS), vol 1, pp 1–5. IEEE
5. Pooja M, Singh DY (2013) Security issues and sybil attack in wireless sensor networks. *Int J P2P Netw Trends Technol* 3(1):7–13
6. Chang W, Wu J (2012) A survey of sybil attacks in networks
7. Sohail Abbas DL-J, Merabti M, Kifayat K (2013) Lightweight sybil attack detection in manets. *IEEE Syst J* 7:236–248
8. Joshi N, Challa M (2014) Secure authentication protocol to detect sybil attacks in manets. *Int J Comput Sci Eng Technol* 5:645–653
9. Hashmi S, Brooke J (2010) Towards sybil resistant authentication in mobile ad hoc networks. In: 2010 Fourth international conference on emerging security information, systems and technologies, pp 17–24
10. Danish Shehzad NUA, Umar AI, Ishaq W (2014) A novel mechanism for detection of sybil attack in manets. In: International conference on computer science and information systems, Oct 2014
11. Alvisi L, Clement A, Epasto A, Lattanzi S, Panconesi A (2013) Sok: the evolution of sybil defense via social networks. In: 2013 IEEE symposium on security and privacy (SP), pp 382–396. IEEE
12. Sharmila S, Umamaheswari G (2013) Node id-based detection of sybil attack in mobile wireless sensor network. *Int J Electron* 100(10):1441–1454
13. Alsaedi N, Hashim F, Sali A (2015) Energy trust system for detecting sybil attack in clustered wireless sensor networks. In: 2015 IEEE 12th Malaysia international conference on communications (MICC), pp 91–95
14. Prabhjotkaur AC, Singh S (2016) Review paper of detection and prevention of sybil attack in wsn using centralized ids. *Int J Eng Sci* 8399
15. Moradi S, Alavi M (2016) A distributed method based on mobile agent to detect sybil attacks in wireless sensor networks. In: 2016 Eighth international conference on information and knowledge technology (IKT), pp 276–280
16. Lakhapal R, Sharma S (2016) Detection prevention of sybil attack in ad hoc network using hybrid map mac technique. In: 2016 International conference on computation of power, energy information and communication (ICCPEIC), pp 283–287
17. Evangelista D, Mezghani F, Nogueira M, Santos A (2016) Evaluation of sybil attack detection approaches in the internet of things content dissemination. In: 2016 Wireless days (WD), pp 1–6
18. Singh R, Singh J, Singh R (2017) A novel sybil attack detection technique for wireless sensor networks. *Adv Comput Sci Technol* 10(2):185–202
19. Kavitha P, Keerthana C, Niroja V, Vivekanandhan V (2014) Mobile-id based sybil attack detection on the mobile adhoc network. *Int J Commun Comput Technol* 2(02)
20. Vasudeva A, Sood M (2012) Sybil attack on lowest id clustering algo-rithm in the mobile ad hoc network. *Int J Netw Secur Appl* 4(5):135

21. Ramesh MM (2016) Sadam: sybil attack detection approach in manets using testbed. *Int J Sci Eng Comput Technol* 6(9):317
22. Chang S-H, Chen Y-S, Cheng S-M (2013) Detection of sybil attacks in participatory sensing using cloud-based trust management system. In: 2013 International symposium on wireless and pervasive computing (ISWPC), pp 1–6. IEEE

# Smart and Autonomous Door Unlocking System Using Face Detection and Recognition



Amrutha Kalturi, Anjali Agarwal, and N. Neelima

**Abstract** Recently, face recognition's popularity has grown beyond imagination and has given a new dimension to the existing security system. From being theoretically proposing the idea on paper to making it a reality, it has gone through a lot of phases. This paper implements ease of access security system to the disabled people. The security is provided in terms of sending a text message to the registered mobile number and buzzing the alarm. The system is trained with images of only known faces of people for whom the door will be unlocked automatically. When the unknown person tries to enter the access to the door is denied. With leading-edge innovations, smart door unlocking system has become more push on it should be well designed to as it is related to home security and impart easy access to the user. This autonomous door unlocking system concerns to keyless door unlock system with the help of face recognition and detection.

**Keywords** Face detection · Face recognition · Security · Door unlock

## 1 Introduction

In the present scenario, relying on electronic devices rather than going for written means or one-on-one communication has been started. The shoot up of the use of electronics has resulted in greater demand for quick and precise authentication and identification to increase reliability. These days PINs are used to get clearances and identification to gain access for buildings or bank accounts and computer systems for security. When our ATM or credit cards are lost, anyone can guess the correct code and get access, but the authenticity of a person is not taken into consideration. So, an unauthorized user can misuse it. Even governments of developed countries

---

A. Kalturi · A. Agarwal (✉) · N. Neelima  
Department of Electronics and Communication Engineering, Amrita School of Engineering,  
Amrita Vishwa Vidyapeetham, Bengaluru, India  
e-mail: [anjaliag1998@gmail.com](mailto:anjaliag1998@gmail.com)

are utilizing facial recognition on a data basis, namely drivers' licenses for a lot more reasons. There is a big difference between facial recognition and facial detection. With facial detection, the search of the computer for a face in an image is by going through the contours of face, and recognition is to confirm the identity of an individual. It can be deployed for door security, personal identification, attendance system, bank accounts operations and so on. Face recognition recognizes a face without any help from human support. This technology can solve this problem as a face is a unique identity of a person. Other techniques require physical interaction with the device, but here, there is no need for that and can be designed with high accuracy and verification rate increases. This can be easily implemented with the present hardware infrastructure, without modifying the cameras and image capture devices as they will work without any problem.

When humans encounter faces, it is easy for them to identify them even if it is positioned in different angles or different lighting conditions and can easily recognize them, but when it comes to training the machine, it becomes difficult and the complexity increases as the number of features is keep on increasing. So, to make the machine understand, various features like the length of the nose, the separation between eyes, the width of eye sockets, lip thickness, the height of cheekbones, the shape of jawline, etc. is used. With numerous elaborated features to be measured, complications could arise.

## 2 Literature Review

The detection and recognition models have been come up from the field of the computer visionary and have proved to be one of the useful and productive applications of image processing and efficient algorithm-based learning. These classifiers provide a successful recognition rate of at least 95%, in which the ratio of false positive is less than 3% of detected faces. OpenCV is open-source library, and it has a multi-platform framework which supports various operating systems such as Windows and Linux and recently started assisting Mac OS also. Its major applications include facial recognition, motion tracking and object identification [1]. OpenCV provides us with Haar classifier which is used for face detection and LBPH can be trained to use for recognition. Both the classifiers process the images in grayscale. Haar classifier inputs its facial features into a series of classifiers to recognize faces in an image. LBP is a texture and visual descriptor which can identify micro visual patterns from our faces. It can test a static image or from a real-time video recorder. The face detector analyzes and segregates into two classes, namely "positive" or "negative" image based on an image's position [2, 3].

To modernize the existing key lock system of doors and to provide more security, the access to the door can be given by using facial recognition [4]. The system is trained by OpenCV classifiers, and output is generated. Based on the result, access to the door is provided [5–10].

The presence of the person at the door is sensed by using two edge sensors which are placed outside the door, and hence, a command is sent through Raspberry Pi to activate the camera [6]. If the captured image and the face match with the trained database, then with the help of a motor revolution, the door is opened [7]. Two edge sensors are placed inside also so that when the individual crosses, the door will get locked automatically [9–12]. If an unauthorized person tries to access the door, then the alert alarm will get started [13–15].

### 3 Proposed Methodology

Facial recognition has become an integral part of upcoming smartphones to enhance security and reliability. This paper implements a security mechanism for the automatic door unlocking system. It is designed to help the disabled people by providing them easy access to the door by informing them through a text message or by ringing an alarm when a person tries to open the door. The working of the model requires three major steps which are face detection by utilizing classifiers which in turn uses the Viola–Jones algorithm, and then recognition is done by using OpenCV LBPH algorithm. The final decision is to open and close the door based on if the person is authenticated. The entire process is shown in Fig. 1.

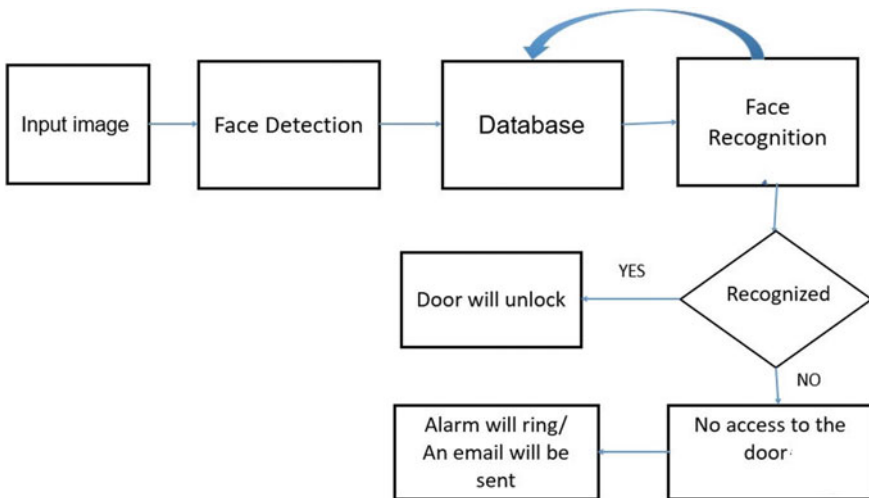


Fig. 1 Block diagram of the proposed smart door unlocking system

### 3.1 Face Detection by Haar Cascade Classifier

Face detection is the subset of “computer vision” technology. Computer vision deals with the objective that how humans extract important information by seeing the objects around us in the same way machines are taught to extract useful information by designing distinct algorithms. In this process, machines are trained by using complex neural networks to find objects and locate faces in the images. The perception can be performed on a real-time video camera or from images. To locate the face, the camera first captures and process to extract the features of a person, and then it detects and identifies it. The detected face is and then compared with the face saved during training by using the dataset provided. If the current face matches and identified that process is termed as face recognition; if they are same, the system will unlock itself. An example of this technology is being used at airport security. The classifier is a program that is used to separate face or non-face from the picture or live stream. The algorithm is trained on lots of positive and negative images so that it can differentiate between a face and non-face image. Two competent classifiers Haar and LBP are provided by OpenCV which are helpful are in detecting faces.

This classifier works on “Haar wavelets” technique that is defined as **“square-shaped” functions which together form a series of rescaled wavelet family or basis.**

Haar classifier consists of four steps.

#### 3.1.1 Haar Feature Selection

Haar wavelets analyze each pixel from the image by dividing into squares. Haar features are extracted from the image by placing the line and edge features as rectangular windows.

#### 3.1.2 Creating Integral Image

Integral image is calculated by adding all the pixels and discards the alternate pixels. This makes the algorithm faster.

$$\text{Sum} = \text{top left} - \text{top right} + \text{bottom right} - \text{bottom left}$$

#### 3.1.3 AdaBoost Training

Every rectangle feature is put on the image to calculate the required feature. The calculated result contains a single value feature which is obtained by the difference between the summation of pixels under white part and black part of the specified



window. To calculate all the features, each window of all possible sizes are placed on all locations of each image. For example, to extract eyes from an image, the edge feature is placed as the region of the eye has more intensity than that of the nose, forehead and cheeks. Now a line feature is placed on the nose as the alternate regions of the nose are darker. Like this, all the rectangular windows are placed and various features are calculated, out of which many are irrelevant. Suppose when a window is placed on the cheek it is not useful because the entire region is the same as there are no light and dark shades. As the maximum region of the image is insignificant so to discard these unnecessary features, the AdaBoost technique is made use. With the help of this method, the accuracy of classification can be enhanced by selecting only relevant features.

### **3.1.4 Cascading Classifiers**

It contains various stages, and in every iteration, a lot of weak learners are identified by decision stumps. Boosting is provided in every stage to improve accuracy by adding the weights to teach weak classifier. Then from all stages, weak learners are summed or cascaded to form a strong classifier.

## **3.2 *Face Recognition by Local Binary Pattern Histogram or LBPH***

Local binary patterns, or LBP, are applied a huge set of images for efficient training. It has a visual and texture descriptor to locate a pattern of micro visuals on the faces of the images. While processing an image, the entire image is divided into blocks of pixels where the block is coded as 3\*3 matrix so each box has 9 pixels which are taken for finding a face and extracting its efficient features. For each box, the central pixel is used to compare with another adjacent if it is larger than the center one it is assigned a value one else zero in the way all the adjacent pixels are updated with binary values and the binary pattern is calculated by reading clockwise direction of the block pixels. Now the calculated pattern is used to plot the histograms for each block which will give a vector for an image. In this way, LBP is used to extract the images. It has fast computational speed when compared with Haar.

## **3.3 *Access Provision to the Door***

The main target of the home security systems is to detect and realize the kind of people entering or leaving the place. Consequent to the monitoring through passwords and pins, people faces can be used as the face is one of the biometric traits. This will have

a profound impact on the present security system. The proposed autonomous door unlock system helps the old age people and also used to prevent robberies in high-security areas like a home environment. This system is powered by the Raspberry Pi circuit.

### **3.3.1 Raspberry Pi**

Raspberry Pi a miniature computer which was developed based on ARM and single board technologies. It serves various educational utilities and can be handled by anyone with ease. It is like a basic computer which enables all the operations that a PC can solve. It consists of an SD card reader slot available on the board as hard drive for Raspberry Pi and the power supply is given by the USB slot and to view the video output it can be connected to an RCA TV set, a modern desktop, or a smart TV using an HDMI port. The power consumption is low, about three watts, which is a major advantage of this miniature computer.

### **3.3.2 Relay**

Relays are electromechanically operated switches which are made of electromagnets, they are usually known as standard relays. The signals that are received from one side of the device controls the switching operations on the other end of the device. The prior objective of the device is to see the signal and connect or disconnect a contact, without human's interaction to switch ON or OFF. By using a low power signal, a high-powered circuit is controlled. Like the DC voltage is used to control the circuit which is activated by AC or high voltage in operating home appliances with DC voltage from the microcontrollers.

### **3.3.3 Solenoid Lock**

It is a type of lock that operates by using the concept of electro-mechanical theory; DC creates a magnetic field that helps in moving the slug. It operates in two modes: One mode deals with power-on which helps the door to unlock by using electromagnet present in the lock by pulling plunger. It is used in various applications like cabinet locking, home automation and electronic door control attendance systems. It has a shockproof design so it is used in emergency purposes like crime threat and accidental fires.

### 3.4 Working Process

In the proposed system, the classifier is trained with the given dataset. Face detection is taking place by the use of Haar classifier and face recognition using LBPH algorithm. The frames are taken through PI camera, and if the person is authenticated, then the access to the door is given.

A virtual environment for Raspberry Pi is created with the help of Virtual Box using a Linux platform. Configuration of the camera is done and the code for real-time face recognition is successfully implemented in the virtual machine. The code is then directly dumped on the hard disk, i.e., on SD card. Solenoid lock is used to unlock the door, and it needs 7–12 V to operate, but Raspberry Pi provides 3.3 V output, so a relay is used to increase the voltage level. Relay utilizes the external power and helps in operating the lock. It has three pins, ground, Vcc and not-connected. The Vcc of the relay is connected to 5 V of Raspberry Pi and GND of the relay to GND of Raspberry Pi. The signal pin of the relay module is connected to the programmed pin of Raspberry Pi. The negative is connected from DC power source to the negative of the solenoid door lock to another end of the relay. DC power source positive pin is connected to the common of the relay module and then connect normally open from the relay module to positive of the solenoid door lock.




## 4 Results

Raspberry Pi was programmed successfully using Virtual Box software by using python with open CV to achieve face recognition door unlocking system. The people who are granted access the door, when they come in front of the door, their image will be captured, pre-processed and recognized based on the available dataset that is used to train the system. The dataset contains 200 images of which 100 each individual of two persons are considered. If a match is found, then the door will be unlocked. If an unauthorized person comes, then the door will not be opened and access is denied. An alarm will rise immediately, and an email will be sent to the registered email ID. Figure 2 displays the results for both cases (Fig. 3).

Confidence values are used to differentiate unknown face and known face from the trained database (Table 1; Fig. 4).

## 5 Conclusion

A method for automatic door unlocking system is proposed that utilizes facial recognition, and a database of the authorized person is created and stored. The system is trained on the provided database. Whenever a person is detected at the door, the commands are sent using Raspberry Pi. The dataset from a real-time scenario was

<p><b>Training dataset</b></p> 	<p><b>Test data</b></p>  <p>Access granted</p>
<p>“ Unauthorised person ” Access is not granted.</p>	

**Fig. 2** Results for authorized and unauthorized access of the proposed system

taken and tested. If the person is authorized, he/she is given access to enter through the door. A text message and an email are sent to registered email ID to know the presence of an unknown person trying to enter the door. This system is mainly focused to enhance the security and safety of the disabled people and provide them ease of access to the door.

```
In [4]: runfile('D:/PP/opencv-face-recognition/opencv-face-recognition/
videoTester.py', wdir='D:/PP/opencv-face-recognition/opencv-face-recognition')
confidence: 44.314426511908366
label: 1
confidence: 40.017753608743135
label: 1
confidence: 39.906339555886206
label: 1
confidence: 43.174233577469316
label: 1
confidence: 42.93912160491611
label: 1
confidence: 41.14997294210464
label: 1
confidence: 40.715647857538436
label: 1
confidence: 40.522126120705764
label: 1
confidence: 42.39088597574739
label: 1
confidence: 39.43925021330714
label: 1
confidence: 40.93524305066209
label: 1
confidence: 43.137045596126015
label: 1
confidence: 41.19017171676194
label: 1
confidence: 43.210458846967136
label: 1
confidence: 42.85423895362213
label: 1
confidence: 167.96883639131445
label: 0
email sent
```

Fig. 3 Confidence values

Table 1 Confidence values

Confidence values	Face recognized
<50	Known face
>=50	Unknown face

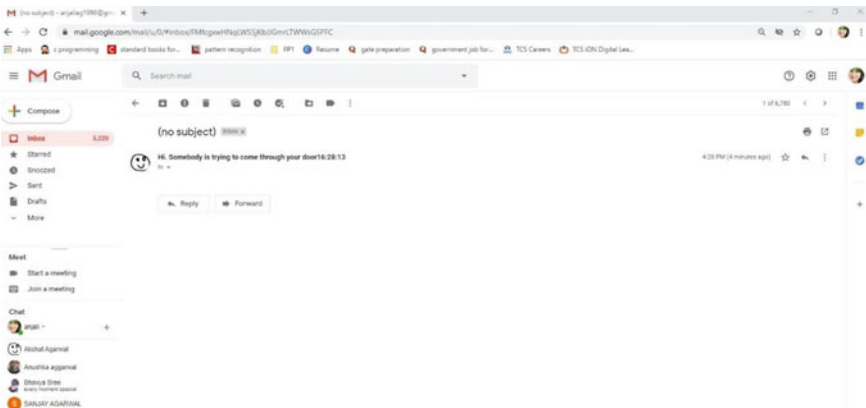


Fig. 4 Mail is sent to the registered mail ID

## References

1. Emami S, Suci VP (2012) Facial recognition using OpenCV senior systems software engineer mobile computer vision team. NVIDIA Australia 2IT&C security master Romania Mar 2012
2. Madhuram M, Kumar BP, Sridhar L, Prem N, Prasad V (2018) Face detection and recognition using OpenCV. In: 2012 24th Chinese Control and Decision Conference (CCDC)- Oct 2018
3. Ahmed A, Ali F (2018) LBPH based improved face recognition, at low resolution Awais Ahmed School of information & software engineering. University of Electronic Science & Technology of China Chengdu, China 2018
4. Roy S, Uddin MN; Haque MZ, Kabir MJ, Design and implementation of the smart door lock system with face recognition method using the linux platform raspberry Pi. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
5. Azeem A, Rao SM, Rao KR, Basha SA, Pedarla H, Gopi M, Door unlock using face recognition. Int J Adv Res Electron Commun Eng (IJARECE) 6(4)
6. Vamsi TK, Sai KC, Vijayalakshmi M (2019) Face recognition-based door unlocking system using raspberry Pi. Int J Adv Res Ideas Innov Technol
7. Maheshwari K, Nalini N (2017) VIT University, facial recognition enabled smart door using microsoft face API. Int J Eng Trends Appl (IJETA) 4(3)
8. Gupte NN, Shelar MR (2013) Smart door locking system. Int J Eng Res Technol (IJERT) 2(11)
9. Saraf T, Shukla K, Balkhande H, Deshmukh A (2018) Automated door access control system using face recognition. Int Res J Eng Technol (IRJET)
10. Yan K, Huang S, Song Y, Liu W, Fan N (2017) Face recognition based on convolution neural network. In: Proceedings of the 36th Chinese control conference, July 26–28, 2017
11. Zhang S, Qiao H (2003) Face recognition with support vector machine. In: International conference on robotics, intelligent systems and signal processing
12. Paul LC, Al Sumam A (2012) Face recognition using principal component analysis method. Int J Adv Res Comput Eng Technol (IJARCET) 1(9)
13. Manoharan S (2019) Smart image processing algorithm for text recognition, information extraction and vocalization for the visually challenged. J Innov Image Process (JIIP) 1(01):31–38
14. Koresh MHJD (2019) Computer vision-based traffic sign sensing for smart transport. J Innov Image Process (JIIP) 1(01):11–19
15. Hassan H, Bakar RA, Mokhtar ATF (2012) Face recognition based on auto-switching magnetic door lock system using microcontroller. In: 2012 International conference on system engineering and technology, Bandung, Indonesia, Sept, pp 11–12

# Frequency Reconfigurable Circular Patch Antenna



R. S. Janisha, D. Vishnu, and O. Sheeba

**Abstract** The design of a frequency reconfigurable antenna placing PIN diode at the ground plane is proposed here. The geometry consists of a circular patch is mounted on an FR4 substrate. The PIN diode is used to disturb the current flow by changing the length of the conductor thereby changing the area. The biasing process will be easier when the diode is placed at the ground plane. So the fabrication of the antenna becomes also easy. The proposed antenna is useful in WLAN 2.45 GHz and UWB applications.

**Keywords** Frequency reconfiguration · Micro-electro-mechanical switch · A microstrip patch antenna · PIN diodes · Reconfiguration · Resonant frequency

## 1 Introduction

In a wireless communication system, an antenna is responsible for navigation, communication and surveillance. The antenna is responsible for the generation and degradation of electromagnetic signals. As the number of the antenna increases it will increase the system performance in terms of better reception and transmission [1] The different types of antennas include wire antenna, aperture antenna, reflector antenna, lens antenna, etc. Proper calculation of antenna dimensions, selection of feeding techniques and polarization are the important aspects in the patch antenna design. Failure in one of the above will leads to performance degradation. In the early stages, antennas are designed to meet a single application, so it is designed to resonate in a

---

R. S. Janisha (✉) · D. Vishnu · O. Sheeba  
Department of ECE, TKM College of Engineering, Kollam, Kerala, India  
e-mail: [rsjanisha@gmail.com](mailto:rsjanisha@gmail.com)

D. Vishnu  
e-mail: [vishnud2118@gmail.com](mailto:vishnud2118@gmail.com)

O. Sheeba  
e-mail: [shb.odattil@gmail.com](mailto:shb.odattil@gmail.com)

single operating frequency [2]. But due to increasing requirements in communication, more advancements and researches are done in antenna design and fabrication. One of the major such advancement is the reconfigurable antenna. This allows us to use a single antenna for multiple applications. Reconfiguration can be frequency reconfiguration, radiation pattern reconfiguration, polarization reconfiguration, and hybrid reconfiguration.

1. Radiation Pattern Reconfigurable Antenna: Using pattern reconfiguration the signal with the desired direction can be obtained. The radiation pattern is changed in terms of shape, direction, gain [3]
2. Polarization Reconfigurable Antenna: Polarisation reconfigurable antenna is used for switching different polarization modes. It helps avoid polarisation mismatch in the portable device [4]
3. Frequency Reconfigurable Antenna: These antennas are capable of changing the frequency of operation by using electronic switches (varactor, PIN diodes), RF MEMS, tunable materials, etc. [5]
- iv. Compound/hybrid Reconfiguration: They have the capability of tuning several antenna parameters simultaneously [3, 6, 7].

## 2 Related Work in Frequency Reconfigurable Antennas

Frequency Reconfiguration is achieved by either by Integrated active switches such as varactor diodes [8, 9] or PIN diodes [10] or Using RF MEMS [11] or Tunable materials or combination of lumped elements [12]. RF MEMS has gained significant attention in reconfigurability than electronic switches due to its small size and superior performances [13]. They provide good performance in high power applications and also provide high isolation. Reconfigurability can be achieved by the use of tuning elements such as materials like liquid crystals and ferrites. Due to its nonlinearity (that is its dielectric constant changes under different voltage levels) it can be used as the substrate for reconfigurable antennas. Reconfigurable antenna for compact wireless devices using PIN diode is designed in [14]. Leonardo Araque Quijano et al. [15] proposed a compact frequency reconfigurable antenna by using optimization. Reconfigurability is achieved by MEMS switches which are connected at the edge sectors of the radiating element. With the help of varactor diodes, the frequency range obtained is 2–4 GHz. An optically pumped frequency reconfigurable antenna [16], besides the conventional frequency reconfigurable switches here, used photoconductive silicon elements as optical switches. High-performance PIN diode [17] is responsible for microwave switching and is used in telemetry applications. Using a cognitive radio approach an efficient device to device communication is explained in [18].



### 3 Proposed Antenna Design

Here a frequency reconfigurable antenna placing PIN diodes at the ground plane is proposed the results are simulated in HFSS. The number of switches needed for pattern reconfiguration is more compared to frequency reconfiguration. Also, the placement of switches in the pattern is quite complex compared to frequency reconfiguration. The proposed antenna consists of a circular patch mounted above an FR4 substrate of dielectric constant 4.4. By switching ON and OFF values of the PIN diode, the resonating length and thereby excitation can be controlled. Hence achieves two different operating frequency.

#### 3.1 Design of Patch Antenna

Antenna design equations are the basic steps in the proposed one. A clear idea must have about the values of dielectric constant, height of the substrate and resonating frequency. That is, first have to set the  $\epsilon_r$ ,  $h$ ,  $f_r$ . Using these values the parameters like patch width (1), effective dielectric constant (2), effective length (3), actual length (4) can be calculated.

1. For calculation of width,

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

where  $c$ —free space velocity of light

$f_r$ —operating frequency

$\epsilon_r$ —dielectric constant of the substrate

2. Effective dielectric constant,

$$\epsilon_{\text{eff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + \frac{12h}{W}\right)^{-1/2} \quad (2)$$

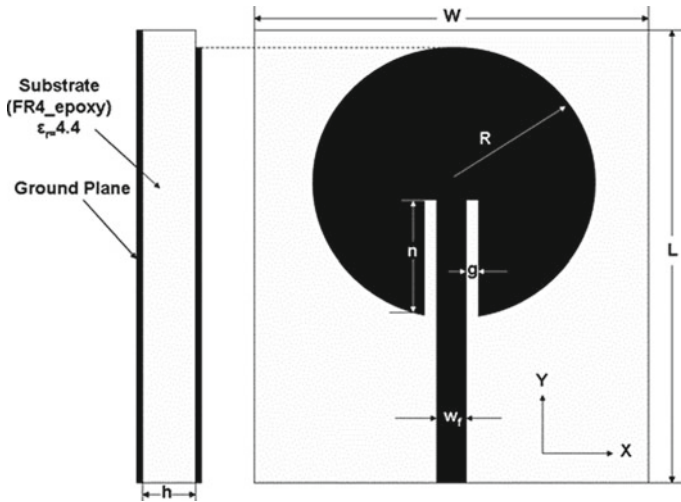
where  $h$ —antenna thickness

3. Calculation of length

$$\Delta L = 0.412 h \frac{(\epsilon_{\text{eff}} + 0.3) \left(\frac{W}{h} + 0.264\right)}{(\epsilon_{\text{eff}} - 0.258) \left(\frac{W}{h} + 0.8\right)} \quad (3)$$

- iv. Calculation of actual length,

$$L = \frac{C}{2f_r \sqrt{\epsilon_{\text{eff}}}} - 2\Delta \frac{1}{2}L \quad (4)$$



**Fig. 1** Circular patch with inset feed

Ground plane parameters such as its length and width are taken as,  $L_g = 2L$ ;  $W_g = 2W$

Figure 1. shows the proposed circular patch antenna geometry having inset feed. The parameters of FR4 substrate include thickness = 1.6, relative permittivity = 4.4, and loss tangent = 0.02. The patch radius is denoted as R. Radiating circular patch and ground is present at either side of the substrate. Ground plane parameters are denoted as length, L and width, W. The inset feed length ( $L_f$ ) and width ( $W_f$ ) are designed with 50  $\Omega$  input impedance to provide a better impedance match

The radius of the patch can be determined by using inset feed design equation as follows;

$$R = \frac{F}{\sqrt{\left\{ 1 + \frac{2h}{\pi \epsilon_r F} \left( \ln \frac{\pi F}{2h} + 1.7726 \right) \right\}}} \tag{5}$$

$$F = \frac{8.791 * 10^9}{f_r \sqrt{\epsilon_r}} \tag{6}$$

where  $f_r$  is the operating frequency

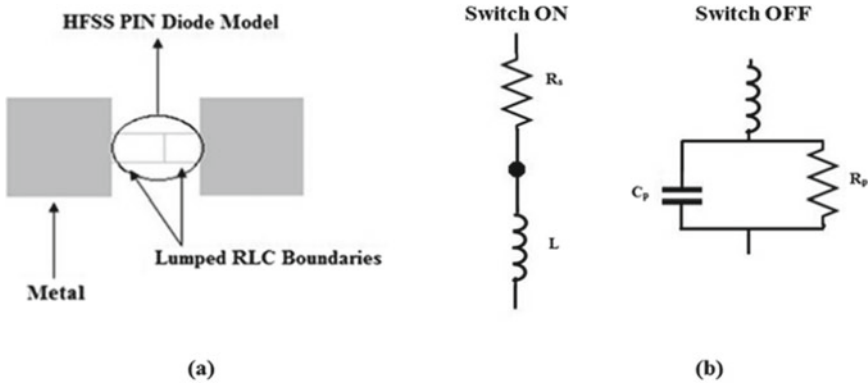
Circular patch antenna using inset feed will help to increase the bandwidth and return loss ( $S_{11}$ ). Here choosing FR4 epoxy as the antenna substrate with a dielectric constant of 4.4. By choosing FR4 epoxy as the substrate the dielectric losses can be minimized. Also, inset feed will provide better impedance matching and return loss [20]. The proposed antenna is useful for 2.4 GHz applications which are come under ISM frequency band (Table 1).

**Table 1** Circular patch antenna design dimensions

Operating frequency	2.4 GHz
Dielectric constant	4.4
Radius of patch	17 mm
Thickness of substrate	1.6 mm
Feed type	Inset feed
Feedline length	13.84 mm
Feedline width	3 mm
The gap in inset feed	0.3 mm
Substrate length	40 mm
Substrate width	30 mm

### 4 Placement of PIN Diodes and Reconfigurability

To obtain frequency reconfigurability the PIN diode placement is done in the ground plane. By changing ON/OFF values of PIN diode the antenna resonates at two discrete frequencies. Placement of diode in ground plane needs simple biasing circuit thereby connection of biasing circuit becomes easy. It gives the best simulation results compared to the placement of diode on patch. Other advantages include it gives a constant bandwidth for two switching frequencies, negligible decrease in gain value and antenna efficiency, reduced size, and low diode loading effect (Fig. 2; Table 2).



**Fig. 2** Lumped RLC equivalent model of pin diode in HFSS

**Table 2** ON and OFF state values of lumped RLC

Diode ON	Diode OFF
Resistance $R_s = 5 \Omega$	Resistance $R_p = 5 K\Omega$
Inductance $L = 0.4 \text{ nH}$	Inductance $L = 0.4 \text{ nH}$
	Capacitance $C_p = 0.5 \text{ pF}$

## 5 Effect of Diodes in the Resonating Frequency

The concept of radiation can be easily analyzed the theory of accelerating charged particles or by the time-varying current. The current distribution greatly depends on the length and thereby the effective area. The current distribution is different for physical and effective length. In physical length (conductor length) the current distribution is asymptotic whereas in effective length the current distribution is uniform that is average current is distributed all over the conductor.

The effective length;

$$L_{\text{eff}} = \frac{2L_{\text{phy}}}{\pi} \quad (7)$$

The effective area;

$$A_e = \frac{\lambda^2 G}{4\pi} \quad (8)$$

where  $G$  is gain

The equation for current distribution for an ideal case is,

$$\text{Current, } I = I_0 \sin\left(\frac{2\pi x}{\lambda}\right) \sin \omega t \quad (9)$$

$I_0$  the maximum current and  $x/\lambda$  is the distance.

When slots are introduced in the antenna geometry either at the ground plane or patch will disturb the flow of current. Switches can be easily placed in these slots. Here using a PIN diode to disturb the current flow. When the diode is ON, the resonating length of the antenna became larger and thereby resonates at a lower frequency. On the other hand, when the diode is OFF, the length became shorter and antenna resonate in high frequency.

## 6 Results and Discussion

### 6.1 Circular Patch Antenna Without Diodes

Figure 3 depicts the simulated return loss( $S_{11}$ ) of the designed antenna. The antenna operates at 2.4 GHz with reflection coefficient value  $-10$  dB. This frequency can be used in the various wireless application, Bluetooth and radio frequency identification.

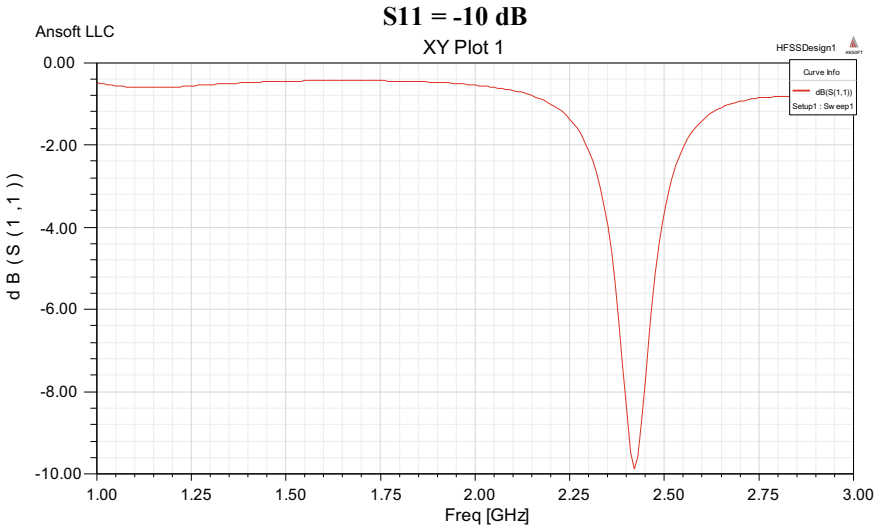


Fig. 3 Simulated result of  $S_{11}$  parameter without diodes

## 6.2 Circular Patch with Diodes

### 6.2.1 When the Diode is ON

Figure 4 illustrates the simulated return loss of the designed antenna with a diode is in ON state. The antenna resonating frequency obtained as 5.4 GHz with reflection coefficient  $-25.52$  dB. The frequency is used in ELAN and the range of frequencies 5.25–5.85 GHz is widely accepted for microwave access.

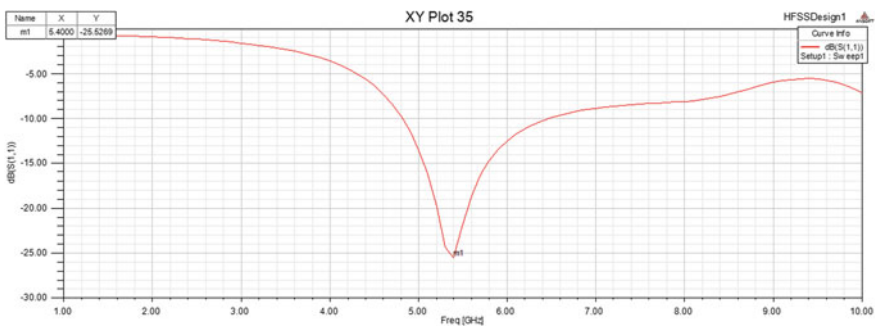
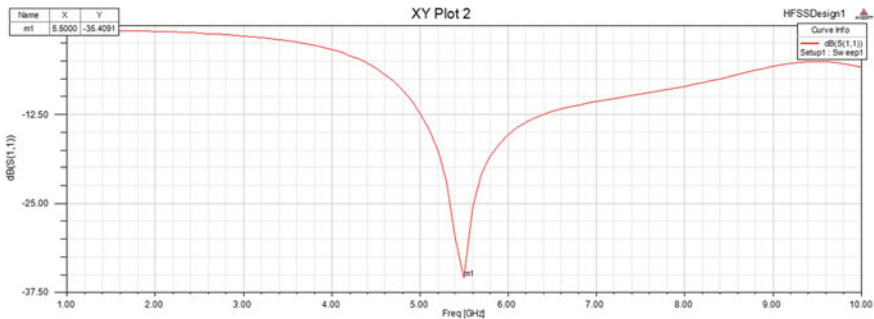


Fig. 4 Simulated  $S_{11}$  parameter of Antenna in ON State



**Fig. 5** Simulated  $S_{11}$  parameter of Antenna in OFF State

### 6.2.2 When the Diode is OFF

Figure 5. depicts the simulated  $S_{11}$  parameter of the designed antenna with a diode is OFF. The resonating frequency of the designed antenna is obtained as 5.5 GHz with reflection coefficient  $-35.4091$  dB. 5.5 GHz is used for Wi-Fi applications.

## 7 Conclusion and Future Scope

Frequency reconfigurable circular patch antenna for various wireless applications is designed and simulated in HFSS. The antenna has an overall size of  $30 \times 40 \times 1.6$ . The antenna structure consists of a circular radiating patch with a diode in its ground plane. Placement of PIN diode in ground plane needs simple biasing circuit thereby connection of biasing circuit becomes easy when compared to PIN diode on the patch. When the diode is ON the resonating frequency is 5.4 GHz with a return loss of  $-25.52$  dB and when the diode is OFF it resonates at 5.5 GHz with a return loss of  $-35.40$ . The operating frequency depends on the current path length. When diode ON, the path length increases and when the diode is OFF path length decreases. The antenna is useful in Bluetooth, Wi-Fi applications, radio frequency identification, etc. The range of frequencies 5.25–5.85 GHz is widely accepted for microwave access. One of the limitations associated with the proposed antenna is the implementation of the biasing circuit, it will increase the complexity of the structure. If the number PIN diodes are increased antenna can be made to resonate at multiple frequencies. Also, Diodes can be placed alternately on patch and substrate to get better results.

## References

1. Noha Hassan, Xavier Fernando (2017) Massive MIMO wireless networks: an overview. *Electronics* 6(63). <https://doi.org/10.3390/electronics6030063>
2. Shimu NJ, Ahmed A (2016) Design and performance analysis of rectangular microstrip patch antenna at 2.45 GHz. In: 2016 5th International conference on informatics, electronics and vision (ICIEV), Dhaka, pp 1062–1066. <https://doi.org/10.1109/iciev.2016.7760161>
3. Selvaraj M, Sreeja BS (2016) Pattern reconfigurable antenna for wireless applications using a parasitic ring. In: 2016 International conference on wireless communications, signal processing and networking (WiSPNET), Chennai, pp 977–980. <https://doi.org/10.1109/wispnet.2016.7566280>
4. Cheung SW, Zhou CF, Li QL, Yuk TI (2016) A simple polarization-reconfigurable antenna. In: 2016 10th European conference on antennas and propagation (EuCAP), Davos, pp 1–4. <https://doi.org/10.1109/eucap.2016.7481801>
5. Rahim MKA, Hamid MR, Samsuri NA, Murad NA, Yusoff MFM, Majid HA (2016) Frequency reconfigurable antenna for future wireless communication system. In: 2016 46th European microwave conference (EuMC), London, pp 965–970. <https://doi.org/10.1109/eumc.2016.7824506>
6. Nguyen-Trong N, Hall L, Fumeaux C (2016) A frequency- and pattern-reconfigurable center-shorted microstrip antenna. *IEEE Antennas Wirel Propag Lett* 15:1955–1958. <https://doi.org/10.1109/lawp.2016.2544943>
7. Huff GH, Feng J, Zhang S, Bernhard JT (2003) A novel radiation pattern and frequency reconfigurable single turn square spiral microstrip antenna. *IEEE Microwave Wirel Compon Lett* 13(2):57–59. <https://doi.org/10.1109/lmwc.2003.808714>
8. Hum SV, Xiong HY (2010) Analysis and Design of a differentially-fed frequency agile microstrip patch antenna. *IEEE Trans Antennas Propag* 58(10):3122–3130. <https://doi.org/10.1109/tap.2010.2055805>
9. Ferrero F, Chevalier A, Ribero JM, Staraj R, Mattei JL, Queffelec Y (2011) A new magneto-dielectric material loaded, tunable UHF antenna for handheld devices. *IEEE Antennas Wirel Propag Lett* 10:951–954. <https://doi.org/10.1109/lawp.2011.2167118>
10. Shayne SV, Augustin G, Aanandan CK, Mohanan P, Vasudevan K (2005) A reconfigurable dual-frequency slot-loaded microstrip antenna controlled by pin diodes. *Microwave Opt Technol Lett* 44(4)
11. Deng Z, Yao Y, Ka-Band frequency reconfigurable microstrip antenna based on MEMS technology
12. Yang SS, Kishk AA, Lee K (2008) Frequency reconfigurable U-Slot microstrip patch antenna. *IEEE Antennas Wirel Propag Lett* 7:127–129. <https://doi.org/10.1109/lawp.2008.921330>
13. Cetiner BA, Roqueta Crusats G, Jofre L, Biyikli N (2010) RF MEMS integrated frequency reconfigurable annular slot antenna. *IEEE Trans Antennas Propag* 58(3):626–632. <https://doi.org/10.1109/tap.2009.2039300>
14. Leonardo Araque Quijano and G. Vecchi (2009) Optimization of an innovative type of compact frequency-reconfigurable antenna. *IEEE Trans Antennas Propag* 57(1):9–18. <https://doi.org/10.1109/tap.2008.2009649>
15. Lai MI, Wu TY, Hsieh JC, Wang CH, Jeng SK, Design of reconfigurable antennas based on an L-shaped slot and PIN diodes for compact wireless devices. <https://doi.org/10.1049/iet-map:20080049>
16. Tawk Y, Albrecht AR, Hemmady S, Balakrishnan G, Christodoulou CG (2010) Optically pumped frequency reconfigurable antenna design. *IEEE Antennas Wirel Propag Lett* 9:280–283. <https://doi.org/10.1109/lawp.2010.2047373>

17. Zaidi A, Baghdad A, Ballouk A, Badri A (2016) Design and optimization of inset fed circular microstrip patch antenna using DGS structure for applications in the millimetre wave band. In: 2016 International conference on wireless networks and mobile communications (WINCOM), Fez, pp 99–103. <https://doi.org/10.1109/wincom.2016.7777198>
18. Sultana A, Zhao L, Fernando X, Efficient resource allocation in device-to-device communication using cognitive radio technology. *IEEE Trans Veh Technol.* <https://doi.org/10.1109/tvt.2017.2743058>



# A New Avenue to the Reciprocity Axioms of Multidimensional DHT Through Those of the Multidimensional DFT



B. N. Madhukar and S. H. Bharathi

**Abstract** The main objective of this paper is to acquaint the reader with a trail-blazing pike into the reciprocity axioms of the multidimensional discrete Hartley transform (DHT) through the multidimensional discrete Fourier transform (DFT). Spin-offs for the reciprocity axioms of multidimensional DHT are tendered through the reciprocity axioms of the multidimensional DFT. This is due to the mathematical hookup betwixt the DFT and DFT. DHT is a real transform, whilst the DFT is a complex transform. Real transforms are posthaste in implementation in contrast to the complex transforms. Ergo, the axioms deduced for the reciprocity of multidimensional DHT in this paper, can be used in real time for the implementation of reciprocity axioms of multidimensional DFT at a brisk pace, by half of the cent percentage than their actual use. Thence, the axioms for multidimensional DHT can be used in real-time applications that use multidimensional DFT such as digital video processing, colour image processing.

**Keywords** Reciprocity · Discrete · Temporal · Frequency · Signal processing

## 1 Introduction

DFT is one of the most rhapsodized discrete transforms in the bailiwick of digital signal processing. But, it also has its fors and againsts, and due to this, many other mathematical transforms have been new fangled in the belles-lettres [1]. One such reason being is that its complex in nature and the mensuration time taken is humongous howbeit FFT algorithms are on tap. DHT flowered by Ronald N. Bracewell in 1983 is one such transform. DHT is a real transform which has its seedbeds in the

---

B. N. Madhukar (✉) · S. H. Bharathi  
Reva University, Bengaluru, India  
e-mail: [madhukarbn890@gmail.com](mailto:madhukarbn890@gmail.com)

S. H. Bharathi  
e-mail: [bharathish@reva.edu.in](mailto:bharathish@reva.edu.in)

infinite Hartley transform flowered by R V L Hartley in 1942 [2–4]. DHT has also Junoesque axioms that are as à la mode as that of the DFT. DHT is also rangy to  $\mathbb{W}$ -Dimensions ( $\mathbb{W}$ D). This paper acquaints the spin-offs of the reciprocity axioms of multidimensional DHT athwart the reciprocity axioms of the DFT.

The constitution of the paper is thence: Ghetto-II and III dish out the vignettes of the multidimensional DFT and DHT, Ghetto-III acquaints the hookup betwixt DFT and DHT, Ghetto-IV regales the interrelation betwixt DFT and DHT, and Ghetto-V doles out the spin-offs of the reciprocity axioms for DHT by capitalizing on the reciprocity axioms for DFT supplanted by Ghetto-VI and VII that acquaints the discussion and conclusion reaped from this research work.

## 2 Multidimensional DFT

This section gives the delineations of the onward and backward transforms of DFT in one, two, three, and  $\mathbb{W}$ -dimensions.

### 2.1 1D-DFT

If  $\mathfrak{c}(\eta)$  is a unidimensional discrete-temporal signal,  $\forall 0 \leq \eta \leq \mathbb{W} - 1$ , then its 1D-DFT is delineated as [5–8]

$$\begin{aligned} \mathbb{C}(\gamma) &= \mathfrak{F}[\mathfrak{c}(\eta)] \\ \mathbb{C}(\gamma) &\triangleq \sum_{\eta=0}^{\mathbb{W}-1} \mathfrak{c}(\eta) \cdot e^{-\frac{j2\pi\gamma\eta}{\mathbb{W}}} \end{aligned} \quad (1)$$

$\forall 0 \leq \gamma \leq \mathbb{W} - 1$ . Here,  $\eta$  and  $\gamma$  betoken discrete temporal-frequency and discrete-frequency variables apiece. The constant  $\mathbb{W}$  coheres to the number of samples in  $\mathfrak{c}(\eta)$  and  $\mathbb{C}(\gamma)$  apiece. Additionally,  $\mathfrak{c}(\eta)$  can be contrived from  $\mathbb{C}(\gamma)$  by the exertion of 1D-inverse discrete Fourier transform (1D-IDFT) which is delineated as [1, 5, 7, 8],

$$\begin{aligned} \mathfrak{c}(\eta) &= \mathfrak{F}^{-1}[\mathbb{C}(\gamma)] \\ \mathfrak{c}(\eta) &\triangleq \frac{1}{\mathbb{W}} \sum_{\gamma=0}^{\mathbb{W}-1} \mathbb{C}(\gamma) \cdot e^{\frac{j2\pi\gamma\eta}{\mathbb{W}}} \end{aligned} \quad (2)$$

$\forall 0 \leq \eta \leq \mathbb{W} - 1$ . If  $\mathfrak{c}(\eta)$  and  $\mathbb{C}(\gamma)$  perdure, they set up a DFT dyad.

$$\mathfrak{c}(\eta) \stackrel{\text{DFT}}{\Leftrightarrow} \mathbb{C}(\gamma) \quad (3)$$

## 2.2 2D-DFT

If  $\mathfrak{c}(\eta_1, \eta_2)$  is a 2D discrete-temporal signal with  $0 \leq \eta_1, \eta_2 \leq \mathbb{W} - 1$ , then its 2D-DFT is delineated as [5–9],

$$\begin{aligned} \mathbb{C}(\gamma_1, \gamma_2) &= \mathfrak{F}[\mathfrak{c}(\eta_1, \eta_2)] \\ \mathbb{C}(\gamma_1, \gamma_2) &\triangleq \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathfrak{c}(\eta_1, \eta_2) \cdot e^{-\frac{j2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2)} \end{aligned} \quad (4)$$

$\forall 0 \leq \gamma_1, \gamma_2 \leq \mathbb{W} - 1$ . Here,  $\eta_1, \eta_2$  and  $\gamma_1, \gamma_2$  betoken discrete temporal-variables and frequency-variables apiece. Additionally,  $\mathfrak{c}(\eta_1, \eta_2)$  can be contrived from  $\mathbb{C}(\gamma_1, \gamma_2)$  by the exertion of 2D-IDFT which is delineated as [3, 5, 7, 8],

$$\begin{aligned} \mathfrak{c}(\eta_1, \eta_2) &= \mathfrak{F}^{-1}[\mathbb{C}(\gamma_1, \gamma_2)] \\ \mathfrak{c}(\eta_1, \eta_2) &\triangleq \frac{1}{\mathbb{W}^2} \sum_{\gamma_1=0}^{\mathbb{W}-1} \sum_{\gamma_2=0}^{\mathbb{W}-1} \mathbb{C}(\gamma_1, \gamma_2) \cdot e^{\frac{j2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2)} \end{aligned} \quad (5)$$

$\forall 0 \leq \eta_1, \eta_2 \leq \mathbb{W} - 1$ . If  $\mathfrak{c}(\eta_1, \eta_2)$  and  $\mathbb{C}(\gamma_1, \gamma_2)$  perdure, they setup a 2D-DFT dyad [9–11].

$$\mathfrak{c}(\eta_1, \eta_2) \stackrel{\text{DFT}}{\Leftrightarrow} \mathbb{C}(\gamma_1, \gamma_2) \quad (6)$$

## 2.3 3D-DFT

If  $\mathfrak{c}(\eta_1, \eta_2, \eta_3)$  is a 3D discrete-temporal signal with  $0 \leq \eta_1, \eta_2, \eta_3 \leq \mathbb{W} - 1$ , then its 3D-DFT is delineated as [5–8, 10, 12],

$$\begin{aligned} \mathbb{C}(\gamma_1, \gamma_2, \gamma_3) &= \mathfrak{F}[\mathfrak{c}(\eta_1, \eta_2, \eta_3)] \\ \mathbb{C}(\gamma_1, \gamma_2, \gamma_3) &\triangleq \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathfrak{c}(\eta_1, \eta_2, \eta_3) \cdot e^{-\frac{j2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2 + \eta_3\gamma_3)} \end{aligned} \quad (7)$$

$\forall 0 \leq \gamma_1, \gamma_2, \gamma_3 \leq \mathbb{W} - 1$ . Here,  $\eta_1, \eta_2, \eta_3$  and  $\gamma_1, \gamma_2, \gamma_3$  betoken temporal-variables and frequency-variables apiece. Additionally,  $\mathfrak{c}(\eta_1, \eta_2, \eta_3)$  can be contrived from  $\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)$  by the exertion of 3D-IDFT which is delineated as [5, 7, 8, 13, 14],

$$\mathfrak{c}(\eta_1, \eta_2, \eta_3) = \mathfrak{F}^{-1}[\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)]$$

$$\mathfrak{C}(\eta_1, \eta_2, \eta_3) = \frac{1}{\mathbb{W}^3} \sum_{\gamma_1=0}^{\mathbb{W}-1} \sum_{\gamma_2=0}^{\mathbb{W}-1} \sum_{\gamma_3=0}^{\mathbb{W}-1} \mathbb{C}(\gamma_1, \gamma_2, \gamma_3) \cdot e^{\frac{j2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2 + \eta_3\gamma_3)} \quad (8)$$

$\forall 0 \leq \eta_1, \eta_2, \eta_3 \leq \mathbb{W} - 1$ . If  $\mathfrak{C}(\eta_1, \eta_2, \eta_3)$  and  $\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)$  perdure, they setup a 3D-DFT dyad [11].

$$\mathfrak{C}(\eta_1, \eta_2, \eta_3) \stackrel{\text{DFT}}{\Leftrightarrow} \mathbb{C}(\gamma_1, \gamma_2, \gamma_3) \quad (9)$$

## 2.4 WD-DFT

If  $\mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n)$  is an ND discrete-temporal signal with  $0 \leq \eta_1, \eta_2, \dots, \eta_n \leq \mathbb{W} - 1$ , then its WD-DFT is delineated as [5–8, 12, 15–17],

$$\begin{aligned} \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n) &= \mathfrak{F}[\mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n)] \\ \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n) &= \sum_{\gamma_1=0}^{\mathbb{W}-1} \sum_{\gamma_2=0}^{\mathbb{W}-1} \dots \sum_{\gamma_n=0}^{\mathbb{W}-1} \mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n) \\ &\quad \cdot e^{-\frac{j2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2 + \dots + \eta_n\gamma_n)} \end{aligned} \quad (10)$$

$\forall 0 \leq \gamma_1, \gamma_2, \dots, \gamma_n \leq \mathbb{W} - 1$ . Here,  $\eta_1, \eta_2, \dots, \eta_n$  and  $\gamma_1, \gamma_2, \dots, \gamma_n$  betoken discrete temporal-variables and frequency-variables apiece. Additionally,  $\mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n)$  can be contrived from  $\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$  by the exertion of WD-IDFT which is delineated as [5–8, 18].

$$\begin{aligned} \mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n) &= \mathfrak{F}^{-1}[\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)] \\ \mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n) &= \frac{1}{\mathbb{W}^n} \sum_{\gamma_1=0}^{\mathbb{W}-1} \sum_{\gamma_2=0}^{\mathbb{W}-1} \dots \sum_{\gamma_n=0}^{\mathbb{W}-1} \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n) \\ &\quad \cdot e^{\frac{j2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2 + \dots + \eta_n\gamma_n)} \end{aligned} \quad (11)$$

$0 \leq \eta_1, \eta_2, \dots, \eta_n \leq \mathbb{W} - 1$ . If  $\mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n)$  and  $\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$  perdure, they setup a WD-DFT dyad [16–18].

$$\mathfrak{C}(\eta_1, \eta_2, \dots, \eta_n) \stackrel{\text{DFT}}{\Leftrightarrow} \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n) \quad (12)$$

### 3 Multidimensional DHT

This section gives the delineations of the onward and backward transforms of DHT in one dimensions, two dimensions, three dimensions, and  $\mathbb{W}$ -dimensions.

#### 3.1 1D-DHT

If  $\mathfrak{c}(\eta)$  is a unidimensional discrete-temporal signal with  $0 \leq \eta \leq N - 1$ , then its DHT is delineated as [5, 14, 19–22],

$$\begin{aligned} \mathbb{C}_H(\gamma) &= \mathfrak{H}[\mathfrak{c}(\eta)] \\ \mathbb{C}_H(\gamma) &\triangleq \sum_{\eta=0}^{\mathbb{W}-1} \mathfrak{c}(\eta) \cdot \text{cas}\left(\frac{2\pi\gamma\eta}{\mathbb{W}}\right) \end{aligned} \tag{13}$$

$\forall 0 \leq \gamma \leq \mathbb{W} - 1$ . The  $\text{cas}(\cdot)$  is vociferated the cosine and sine kernel. Also,  $\text{cas}(\mu) = \cos(\mu) + \sin \mu$ , where  $\mu$  is any real number [6, 9]. Here,  $\eta$  and  $\gamma$  betoken discrete temporal-variables and frequency-variables apiece. Additionally,  $\mathfrak{c}(\eta)$  can be contrived from  $\mathbb{C}(\gamma)$  by the exertion of 1D-inverse discrete Hartley transform (1D-IDHT) which is delineated as [3, 4, 23],

$$\begin{aligned} \mathfrak{c}(\eta) &= \mathfrak{H}^{-1}[\mathbb{C}_H(\gamma)] \\ \mathfrak{c}(\eta) &\triangleq \frac{1}{\mathbb{W}} \sum_{\gamma=0}^{\mathbb{W}-1} \mathbb{C}_H(\gamma) \cdot \text{cas}\left(\frac{2\pi\gamma\eta}{\mathbb{W}}\right) \end{aligned} \tag{14}$$

$\forall 0 \leq \eta \leq \mathbb{W} - 1$ . It is tectonic to weigh in that the DHT harnesses the coequal kernel in its delineations of the onward and rearward transforms. Ergo, DHT is an involutory discrete transform [18, 22, 24]. If  $\mathfrak{c}(\eta)$  and  $\mathbb{C}_H(\gamma)$  perdure, they setup a DHT dyad.

$$\mathfrak{c}(\eta) \stackrel{\text{DHT}}{\Leftrightarrow} \mathbb{C}_H(\gamma) \tag{15}$$

### 3.2 2D-DHT

If  $\mathfrak{c}(\eta_1, \eta_2)$  is a 2D discrete-temporal signal with  $0 \leq \eta_1, \eta_2 \leq \mathbb{W} - 1$ , then its 2D-DHT is delineated as [5, 14, 19–21, 25],

$$\begin{aligned} \mathbb{C}_H(\gamma_1, \gamma_2) &= \mathfrak{H}[\mathfrak{c}(\eta_1, \eta_2)] \\ \mathbb{C}_H(\gamma_1, \gamma_2) &\triangleq \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathfrak{c}(\eta_1, \eta_2) \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2) \right] \end{aligned} \quad (16)$$

$\forall 0 \leq \gamma_1, \gamma_2 \leq \mathbb{W} - 1$ . Here,  $\eta_1, \eta_2$  and  $\gamma_1, \gamma_2$  betoken discrete temporal-variables and frequency-variables apiece. Additionally,  $\mathfrak{c}(\eta_1, \eta_2)$  can be contrived from  $\mathbb{C}(\gamma_1, \gamma_2)$  by the exertion of 2D-IDHT which is delineated as [3, 4, 23, 26],

$$\begin{aligned} \mathfrak{c}(\eta_1, \eta_2) &= \mathfrak{H}^{-1}[\mathbb{C}_H(\gamma_1, \gamma_2)] \\ \mathfrak{c}(\eta_1, \eta_2) &\triangleq \frac{1}{\mathbb{W}^2} \sum_{\gamma_1=0}^{\mathbb{W}-1} \sum_{\gamma_2=0}^{\mathbb{W}-1} \mathbb{C}(\gamma_1, \gamma_2) \\ &\quad \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2) \right] \end{aligned} \quad (17)$$

$\forall 0 \leq \eta_1, \eta_2 \leq \mathbb{W} - 1$ . If  $\mathfrak{c}(\eta_1, \eta_2)$  and  $\mathbb{C}_H(\gamma_1, \gamma_2)$  perdure, they setup a 2D-DHT dyad.

$$\mathfrak{c}(\eta_1, \eta_2) \stackrel{\text{DHT}}{\Leftrightarrow} \mathbb{C}_H(\gamma_1, \gamma_2) \quad (18)$$

### 3.3 3D-DHT

If  $\mathfrak{c}(\eta_1, \eta_2, \eta_3)$  is a 3D discrete-temporal signal with  $0 \leq \eta_1, \eta_2, \eta_3 \leq \mathbb{W} - 1$ , then its 3D-DHT is delineated as [5, 14, 19–21, 26],

$$\begin{aligned} \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) &= \mathfrak{H}[\mathfrak{c}(\eta_1, \eta_2, \eta_3)] \\ \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) &\triangleq \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathfrak{c}(\eta_1, \eta_2, \eta_3) \\ &\quad \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3) \right] \end{aligned} \quad (19)$$

$\forall 0 \leq \gamma_1, \gamma_2, \gamma_3 \leq \mathbb{W} - 1$ . Here,  $\eta_1, \eta_2, \eta_3$  and  $\gamma_1, \gamma_2, \gamma_3$  betoken discrete temporal-variables and frequency-variables apiece. Additionally,  $\mathbb{C}(\eta_1, \eta_2, \eta_3)$  can be contrived from  $\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)$  by the exertion of 3D-IDHT which is delineated as [3, 4, 23, 27],

$$\begin{aligned} \mathbb{C}(\eta_1, \eta_2, \eta_3) &= \mathfrak{H}^{-1}[\mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3)] \\ \mathbb{C}(\eta_1, \eta_2, \eta_3) &\triangleq \frac{1}{\mathbb{W}^3} \sum_{\gamma_1=0}^{\mathbb{W}-1} \sum_{\gamma_2=0}^{\mathbb{W}-1} \sum_{\gamma_3=0}^{\mathbb{W}-1} \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) \\ &\quad \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3) \right] \end{aligned} \tag{20}$$

$\forall 0 \leq \eta_1, \eta_2, \eta_3 \leq \mathbb{W} - 1$ . If  $\mathbb{C}(\eta_1, \eta_2, \eta_3)$  and  $\mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3)$  perdure, they setup a 3D-DHT dyad.

$$\mathbb{C}(\eta_1, \eta_2, \eta_3) \stackrel{\text{DHT}}{\Leftrightarrow} \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) \tag{21}$$

### 3.4 $\mathbb{W}\text{D-DHT}$

If  $\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)$  is a  $\mathbb{W}\text{D}$  discrete-temporal signal with  $0 \leq \eta_1, \eta_2, \dots, \eta_n \leq \mathbb{W} - 1$ , then its  $\mathbb{W}\text{D-DHT}$  is delineated as [5, 14, 19–21, 26],

$$\begin{aligned} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) &= \mathfrak{H}[\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)] \\ \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) &\triangleq \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \\ &\quad \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n \} \right] \end{aligned} \tag{22}$$

$0 \leq \gamma_1, \gamma_2, \dots, \gamma_n \leq \mathbb{W} - 1$ . Here,  $\eta_1, \eta_2, \dots, \eta_n$  and  $\gamma_1, \gamma_2, \dots, \gamma_n$  betoken discrete temporal-variables and frequency-variables apiece. Additionally,  $\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)$  can be contrived from  $\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$  by the exertion of  $\mathbb{W}\text{D-IDHT}$  which is delineated as [3, 4, 23, 28],

$$\begin{aligned} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) &= \mathfrak{H}^{-1}[\mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n)] \\ \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) &\triangleq \frac{1}{\mathbb{W}^n} \sum_{\gamma_1=0}^{\mathbb{W}-1} \sum_{\gamma_2=0}^{\mathbb{W}-1} \dots \sum_{\gamma_n=0}^{\mathbb{W}-1} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) \\ &\quad \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n \} \right] \end{aligned} \tag{23}$$

$\forall 0 \leq \eta_1, \eta_2, \dots, \eta_n \leq \mathbb{W} - 1$ . If  $\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)$  and  $\mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n)$  perdure, they setup a  $\mathbb{W}$  D-DHT dyad [27].

$$\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \stackrel{\text{DHT}}{\Leftrightarrow} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) \quad (24)$$

## 4 Interrelation Betwixt DFT and DHT

DFT and DHT are communed by flat-out guileless mathematical articulations. Suppose  $\mathbb{C}_R(\gamma) = \text{Re}[\mathbb{C}(\gamma)]$  and  $\mathbb{C}_I(\gamma) = \text{Im}[\mathbb{C}(\gamma)]$  cohere to the real and imaginary parts of the DFT,  $\mathbb{C}(\gamma) = \mathfrak{F}[\mathbb{C}(\eta)]$ , then the DHT of  $\mathbb{C}(\eta)$ , viz.  $\mathbb{C}_H(\gamma)$  is habituated by the articulation [13, 20, 23, 29],

$$\mathbb{C}_H(\gamma) = \mathbb{C}_R(\gamma) - \mathbb{C}_I(\gamma) = \text{Re}[\mathbb{C}(\gamma)] - \text{Im}[\mathbb{C}(\gamma)] \quad (25)$$

This bestows the DHT athwart the DFT. The DFT can be notched up from the DHT by exploiting the kindred,

$$\mathbb{C}(\gamma) = \frac{1}{2}[\mathbb{C}_H(\gamma) + \mathbb{C}_H(\mathbb{W} - \gamma)] - j\frac{1}{2}[\mathbb{C}_H(\gamma) - \mathbb{C}_H(\mathbb{W} - \gamma)] \quad (26)$$

Equations (25) and (26) can be long drawn-out to multidimensions [14, 15, 23, 26–28].

## 5 Spin-Offs of the Reciprocity Axioms for Multidimensional DHT from Those of the Multidimensional DFT

### 5.1 Spin-off of the Reciprocity Axiom for 1D-DHT from that of 1D-DFT

**Axiom 1** If  $\mathbb{C}(\eta)$  and  $\mathbb{C}_H(\gamma) = \mathbb{C}(\gamma)$  setup a 1D-DHT dyad, then  $\mathbb{W}\mathbb{C}_H(\gamma) = \mathbb{W}\mathbb{C}(\gamma)$  and  $\mathbb{C}(\eta)$  setup a 1D-DHT dyad.

*Proof* This is proved in [29].



### 5.2 Spin-off of the Reciprocity Axiom for 2D-DHT from that of 2D-DFT

**Axiom 2** If  $\mathbb{C}(\eta_1, \eta_2)$  and  $\mathbb{C}_H(\gamma_1, \gamma_2) = \mathbb{C}(\gamma_1, \gamma_2)$  setup a 2D-DHT dyad, then,  $\mathbb{W}^2\mathbb{C}_H(\gamma_1, \gamma_2) = \mathbb{W}^2\mathbb{C}(\gamma_1, \gamma_2)$  and  $\mathbb{C}(\eta_1, \eta_2)$  setup a 2D-DHT dyad.

**Proof** The reciprocity axiom for 2D-DFT is habituated by

$$\begin{aligned} \mathbb{W}^2\mathbb{C}[((- \gamma_1))_{\mathbb{W}}, ((- \gamma_2))_{\mathbb{W}}] &= \mathbb{W}^2\mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) = \mathfrak{F}[\mathbb{C}(\eta_1, \eta_2)] \\ \mathbb{W}^2\mathbb{C}[((- \gamma_1))_{\mathbb{W}}, ((- \gamma_2))_{\mathbb{W}}] &= \mathbb{W}^2\mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) \\ &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot e^{-\frac{2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2)} \end{aligned} \tag{27}$$

Inasmuch as  $\mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2)$  is complex, can be transcribed,

$$\begin{aligned} &\mathbb{W}^2\mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) + j\mathbb{W}^2\mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) \\ &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot \cos\left\{\frac{2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2)\right\} \\ &\quad - j \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot \sin\left\{\frac{2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2)\right\} \end{aligned} \tag{28}$$

It acquiesces to

$$\mathbb{W}^2\mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) = \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot \cos\left\{\frac{2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2)\right\} \tag{29}$$

and

$$\mathbb{W}^2\mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) = - \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot \sin\left\{\frac{2\pi}{\mathbb{W}}(\eta_1\gamma_1 + \eta_2\gamma_2)\right\} \tag{30}$$

The hookup betwixt 2D-DHT and 2D-DFT is habituated by

$$\mathbb{C}_H(\gamma_1, \gamma_2) = \mathbb{C}_R(\gamma_1, \gamma_2) - \mathbb{C}_I(\gamma_1, \gamma_2) = \text{Re}[\mathbb{C}(\gamma_1, \gamma_2)] - \text{Im}[\mathbb{C}(\gamma_1, \gamma_2)] \tag{31}$$

where  $\mathbb{C}_R(\gamma_1, \gamma_2) = \text{Re}[\mathbb{C}(\gamma_1, \gamma_2)]$  and  $\mathbb{C}_I(\gamma_1, \gamma_2) = \text{Im}[\mathbb{C}(\gamma_1, \gamma_2)]$  apiece. Putting out the barometers for reciprocity and chronicling that there is an about-face in the sign of frequency solely for the real and imaginary parts of the 2D-DFT, we get

$$\begin{aligned}\mathbb{W}^2 \mathbb{C}_H(\gamma_1, \gamma_2) &= \mathbb{W}^2 \mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) - \mathbb{W}^2 \mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2) \\ \mathbb{W}^2 \mathbb{C}_H(\gamma_1, \gamma_2) &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot \left[ \cos \left\{ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2) \right\} \right. \\ &\quad \left. + \sin \left\{ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2) \right\} \right]\end{aligned}\quad (32)$$

$$\Rightarrow \mathbb{W}^2 \mathbb{C}_H(\gamma_1, \gamma_2) = \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot \text{cas} \left\{ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2) \right\} \quad (33)$$

Penciling in  $\mathbb{C}_H(\gamma_1, \gamma_2) = \mathbb{C}(\gamma_1, \gamma_2)$  in Eq. (33), got

$$\Rightarrow \mathbb{W}^2 \mathbb{C}(\gamma_1, \gamma_2) = \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2) \cdot \text{cas} \left\{ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2) \right\} = \mathfrak{H}[\mathbb{C}(\eta_1, \eta_2)] \quad (34)$$

Thus,  $\mathbb{W}^2 \mathbb{C}_H(\gamma_1, \gamma_2) = \mathbb{W}^2 \mathbb{C}(\gamma_1, \gamma_2)$  and  $\mathbb{C}(\eta_1, \eta_2)$  setup a 2D-DHT dyad. The interrelation between the DFT and DHT is the key point in establishing the reciprocity axiom for 2D-DHT from 2D-DFT.

### 5.3 Spin-off of the Reciprocity Axiom for 3D-DHT from that of 3D-DFT

**Axiom 3** If  $\mathbb{C}(\eta_1, \eta_2, \eta_3)$  and  $\mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) = \mathbb{C}(\gamma_1, \gamma_2, \gamma_3)$  setup a 3D-DHT dyad, then  $\mathbb{W}^3 \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) = \mathbb{W}^3 \mathbb{C}(\gamma_1, \gamma_2, \gamma_3)$  and  $\mathbb{C}(\eta_1, \eta_2, \eta_3)$  setup a 3D-DHT dyad.

**Proof** The reciprocity axiom for 3D-DFT is habituated by

$$\begin{aligned}\mathbb{W}^3 \mathbb{C}[\{(-\gamma_1)_{\mathbb{W}}, \{(-\gamma_2)_{\mathbb{W}}, \{(-\gamma_3)_{\mathbb{W}}\}\}] &= \mathbb{W}^3 \mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3) \\ &= \mathfrak{F}[\mathbb{C}(\eta_1, \eta_2, \eta_3)]\end{aligned}\quad (35)$$

$$\mathbb{W}^3 \mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3) = \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot e^{-\frac{j2\pi}{\mathbb{W}} \{\eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3\}} \quad (36)$$

Inasmuch as  $\mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3)$  is complex, can be transcribed,

$$\mathbb{W}^3 \mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3) + j \mathbb{W}^3 \mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3)$$

$$\begin{aligned}
 &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot \cos \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right] \\
 &- \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot \sin \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right]
 \end{aligned} \tag{37}$$

Here,

$$\begin{aligned}
 &\mathbb{W}^3 \mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3) \\
 &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot \cos \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right]
 \end{aligned} \tag{38}$$

and

$$\begin{aligned}
 &N^3 \mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3) = \\
 &- \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot \sin \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right]
 \end{aligned} \tag{39}$$

The hookup betwixt 3D-DHT and 3D-DFT is habituated by

$$\begin{aligned}
 \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) &= \mathbb{C}_R(\gamma_1, \gamma_2, \gamma_3) - \mathbb{C}_I(\gamma_1, \gamma_2, \gamma_3) \\
 &= \text{Re}[\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)] - \text{Im}[\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)]
 \end{aligned} \tag{40}$$

whence,  $\mathbb{C}_R(\gamma_1, \gamma_2, \gamma_3) = \text{Re}[\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)]$  and  $\mathbb{C}_I(\gamma_1, \gamma_2, \gamma_3) = \text{Im}[\mathbb{C}(\gamma_1, \gamma_2, \gamma_3)]$  apiece. Putting out the barometers for reciprocity and chronicling that there is an about-face in the sign of frequency solely for the real and imaginary parts of the 3D-DFT, we get

$$\begin{aligned}
 \mathbb{W}^3 \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) &= \mathbb{W}^3 \mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3) \\
 &- \mathbb{W}^3 \mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \mathbb{W} - \gamma_3)
 \end{aligned} \tag{41}$$

$$\begin{aligned}
 \Rightarrow \mathbb{W}^3 \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot \cos \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right] \\
 &+ \sin \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right] \\
 \Rightarrow \mathbb{W}^3 \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right]
 \end{aligned} \tag{42}$$

Pencilising in  $\mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) = \mathbb{C}(\gamma_1, \gamma_2, \gamma_3)$  in Eq. (42), got

$$\begin{aligned} \Rightarrow \mathbb{W}^3 g(\gamma_1, \gamma_2, \gamma_3) &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \sum_{\eta_3=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \eta_3) \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} \{ \eta_1 \gamma_1 + \eta_2 \gamma_2 + \eta_3 \gamma_3 \} \right] \\ &= \mathfrak{H}[\mathbb{C}(\eta_1, \eta_2, \eta_3)] \end{aligned} \quad (43)$$

Thus,  $\mathbb{W}^3 \mathbb{C}_H(\gamma_1, \gamma_2, \gamma_3) = \mathbb{W}^3 \mathbb{C}(\gamma_1, \gamma_2, \gamma_3)$  and  $\mathbb{C}(\eta_1, \eta_2, \eta_3)$  setup a 3D-DHT dyad. The interrelation between the DFT and DHT is the key point in establishing the reciprocity axiom for 3D-DHT from 3D-DFT.

#### 5.4 Spin-off of the reciprocity axiom for $\mathbb{W}D$ -DHT from that of $\mathbb{W}D$ -DFT

**Axiom 4** If  $\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)$  and  $\mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) = \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$  setup a  $\mathbb{W}D$ -DHT dyad, then  $\mathbb{W}^{\mathbb{W}} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) = \mathbb{W}^{\mathbb{W}} \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$  and  $\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)$  setup a  $\mathbb{W}D$ -DHT dyad.

**Proof** The reciprocity axiom for  $\mathbb{W}D$ -DFT is habituated by

$$\begin{aligned} \mathbb{W}^{\mathbb{W}} \mathbb{C}[((- \gamma_1)_{\mathbb{W}}, ((- \gamma_2)_{\mathbb{W}}), \dots, ((- \gamma_n)_{\mathbb{W}}))] &= \mathbb{W}^{\mathbb{W}} \mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n) \\ &= \mathfrak{F}[\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)] \end{aligned} \quad (44)$$

$$\begin{aligned} \Rightarrow \mathbb{W}^{\mathbb{W}} \mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n) \\ = \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot e^{-j \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n)} \end{aligned} \quad (45)$$

Inasmuch as  $\mathbb{C}(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n)$  is complex, we have

$$\begin{aligned} \mathbb{W}^{\mathbb{W}} \mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n) &+ j \mathbb{W}^{\mathbb{W}} \mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n) \\ = \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot \cos \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \\ - j \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot \sin \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \end{aligned} \quad (46)$$

Here,

$$\mathbb{W}^{\mathbb{W}} \mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n)$$

$$= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot \cos \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \quad (47)$$

and

$$\begin{aligned} & \mathbb{W}^{\mathbb{W}} \mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n) \\ &= - \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot \sin \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \end{aligned} \quad (48)$$

The hookup betwixt  $\mathbb{W}\text{D-DHT}$  and  $\mathbb{W}\text{D-DFT}$  is habituated by

$$\begin{aligned} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) &= \mathbb{C}_R(\gamma_1, \gamma_2, \dots, \gamma_n) - \mathbb{C}_I(\gamma_1, \gamma_2, \dots, \gamma_n) \\ &= \text{Re}[\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)] - \text{Im}[\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)] \end{aligned} \quad (49)$$

where  $\mathbb{C}_R(\gamma_1, \gamma_2, \dots, \gamma_n) = \text{Re}[\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)]$  and  $\mathbb{C}_I(\gamma_1, \gamma_2, \dots, \gamma_n) = \text{Im}[\mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)]$  apiece. Putting out the barometers for reciprocity and chronicling that there is an about-face in the sign of frequency solely for the real and imaginary parts of the  $\mathbb{W}$  D-DFT, got

$$\begin{aligned} \Rightarrow \mathbb{W}^{\mathbb{W}} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) &= \mathbb{W}^{\mathbb{W}} \mathbb{C}_R(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n) \\ &\quad - \mathbb{W}^{\mathbb{W}} \mathbb{C}_I(\mathbb{W} - \gamma_1, \mathbb{W} - \gamma_2, \dots, \mathbb{W} - \gamma_n) \end{aligned} \quad (50)$$

$$\begin{aligned} \Rightarrow \mathbb{W}^{\mathbb{W}} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot \left\{ \cos \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \right. \\ &\quad \left. + \sin \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \right\} \\ \Rightarrow \mathbb{W}^{\mathbb{W}} \mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) &= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \end{aligned} \quad (51)$$

Pencilling in  $\mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) = \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$  in Eq. (51), we get

$$\Rightarrow \mathbb{W}^{\mathbb{W}} \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$$

$$\begin{aligned}
&= \sum_{\eta_1=0}^{\mathbb{W}-1} \sum_{\eta_2=0}^{\mathbb{W}-1} \dots \sum_{\eta_n=0}^{\mathbb{W}-1} \mathbb{C}(\eta_1, \eta_2, \dots, \eta_n) \cdot \text{cas} \left[ \frac{2\pi}{\mathbb{W}} (\eta_1 \gamma_1 + \eta_2 \gamma_2 + \dots + \eta_n \gamma_n) \right] \\
&= \mathfrak{H}[\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)]
\end{aligned} \tag{52}$$

Thus,  $\mathbb{C}_H(\gamma_1, \gamma_2, \dots, \gamma_n) = \mathbb{C}(\gamma_1, \gamma_2, \dots, \gamma_n)$  and  $\mathbb{C}(\eta_1, \eta_2, \dots, \eta_n)$  setup a  $\mathbb{W}\text{D-DHT}$  dyad. The interrelation between the DFT and DHT is the key point in establishing the reciprocity axiom for  $\mathbb{W}\text{D-DHT}$  from  $\mathbb{W}\text{D-DFT}$ .

## 6 Discussion

The ontogenesis of reciprocity axioms for multidimensional DHT from those of the multidimensional DFT is conferred. Reckoning time is used in the technical belles-lettres as one of the performance cadent in evaluating the fruition of these two transforms, and the DHT takes less time than that of DFT. Besides, this is bonafide in the multidimensional case. It is serendipitous to write the outright number of real multiplications and real additions for the exertion of the multidimensional DHT. If an archetype is taken, the all-out number of real multiplications for reckoning the 3D-DHT using the row-column approach is  $3\mathbb{W}^3 \log_2 \mathbb{W}$  and the all-out number of real additions is  $3\mathbb{W}^3 \left[ 1 + \frac{3}{2} \log_2 \mathbb{W} \right]$  apiece. Multidimensional DHT is predominantly serviceable in applications where the involution of the DHT comes into be dextrous, specifically in cyclic convolution and memory critical fiefdoms and vector set up times are crescentic. Ergo, the reciprocity axioms can be middingly used for such purposes in conjunction with fast DHT algorithms as the FFT algorithms take a great deal of time for pursuance in such applications. Thence, it is proposed in this paper that these axioms developed for multidimensional DHT can be used for signal processing applications for which the multidimensional DFT reciprocity axioms are used as real transforms get executed swiftly than complex transforms and as such DHT is a real transform. A typical illustration is the finding of the onward and rearward transforms of multidimensional DFT in baronies such as 3D imaging using multidimensional DHT reciprocity axioms instead of using the former's delineations of them. This is a clear advantage of the reciprocity axioms developed in this paper over the existing ones and can be chronicled seriously while implementing signal processing algorithms on pronto VLSI architectures where throughput with precision is numero uno. As well, the algorithms developed in this paper can be utilitarian in Power Spectrum Estimation and other application fiefdoms. The axioms can also be used in the implementation of multidimensional signal processing algorithms on FPGAs, ASICs, etc.

## 7 Conclusions

This paper prates the reciprocity axioms for multidimensional DHT by finagling a distinctive pike, i.e., by drawing upon the reciprocity axioms for multidimensional DFT. The reciprocity axioms abet in the mensuration of the onward and the sternforemost transforms of signals bearing the suchlike envelope either in the discrete-temporal or discrete-frequency domains. In discrete-time signal processing, one bumps into signals bearing the suchlike envelope in the temporal and frequency fiefs. The reciprocity axioms of DHT can be drawn upon for guesstimating the onward and sternforemost transforms predicated on the extant DHT dyad. Inasmuch as DHT is a real transformation, the mensuration deadweight is sunken by partway of the cent per cent which soliloquizes of this probity in realistic pursuance. Also, it has been mentioned in the literature that the computation time taken for finding multidimensional DFT of signals is more if its delineation is used, but is less if the multidimensional DHT is used for finding the same and computation time is used as the performance metric. In consequence, whilst using the reciprocity axioms of multidimensional DFT, we can use the reciprocity axioms of multidimensional DHT developed in this research paper can be used as there are fast algorithms available for the latter which can give accurate results than those of the former. These axioms can be adroitly drawn upon in unriddling of random difference equations, difference equations, in the barony of multidimensional signal processing such as remote sensing image processing, etc. Depending upon the architecture of the processor and the compiler of the software used, the hustle of execution of the multidimensional fast DHT algorithms can be jetted up in cartel with the axioms evolved in this research paper.

**Acknowledgements** The authors are chuffed to their household constituents and the REVA University Governance, Bengaluru, for inspiring us for hauling this disquisition work.

## References

1. Shi Y (2019) Multi-dimensional processing for big data with Noise. In: IEEE international conference on power, intelligent computing and systems, vol 14, pp 686–690
2. Hartley RVL (1942) A more symmetrical fourier analysis applied to transmission Problems. In: IRE, pp 144–150
3. Madhukar BN Bharathi SH (2020) Similarity properties for two versions of 1D-, 2D-, 3D-, and ND-Infinite Hartley Transform (IHT). *Int J Adv Sci Technol* 29:11267–11274
4. Jones K (2010) *The regularized fast Hartley transform*. Springer, New York
5. Madhukar BN, Bharathi SH (2020) An artery to the reciprocity properties of N-dimensional Infinite Hartley Transform through N-dimensional Infinite Fourier Transform. *Int J Adv Sci Technol* 29:8591–8599
6. Dudgeon D, Mersereau MR (1984) *Multidimensional Digital Signal processing*. Prentice-Hall, New Jersey
7. Gonzalez RC, Woods RE (2020) *Digital image processing*. Pearson Education Inc, New Delhi

8. Oppenheim AV, Schaffer RW (2020) Discrete-time signal processing. Pearson Education Inc, New Delhi
9. Palanisamy R (2014) The impact of privacy concerns on trust, attitude and intention of using a search engine: an empirical analysis. *Int J Electron Bus (IJEB)* 11:280–302
10. Bracewell RN (1983) Discrete Hartley Transform. *J Opt Soc Am A* 73:1832–1835
11. Bracewell RN (1990) Assessing the Hartley transform. *IEEE Trans Acoust Speech Sig Process ASSP* 38:2174–2176
12. Bracewell RN (1984) The fast Hartley transform. *Proc IEEE* 72:1010–1018
13. Madhukar BN, Jain S (2016) A duality theorem for Infinite Hartley Transform. In: *IEEE 5th international conference on communication and signal processing 2016 (ICCSP'16)*. Melmaruvathur, pp 109–113
14. Bracewell RN (1986) The Hartley transform. Oxford University Press, London
15. Madhukar BN, Jain S (2016) A duality theorem for the Discrete Sine Transform–IV (DST–IV). In: *IEEE 3rd international conference on advanced computing and communication systems 2016 (ICACCS–2016)*, pp 1–6
16. Olejniczak KJ, Heydt GT (1994) Scanning the special section on the Hartley transform. *Proc IEEE* 82:372–380
17. Bracewell RN (1987) Alternative to split-radix Hartley transform. *Electron Letters* 23:1148–1149
18. Bracewell RN, Buneman O, Hao H, Villasenor J (1986) Fast two-dimensional Hartley transforms. *Proc IEEE* 74:1282–1283
19. Madhukar BN, Bharathi SH (2018) A new property of the Discrete Cosine Transform–IV (DCT–IV). In: *IEEE international conference on communication and signal processing 2018 (ICCSP'18)*. Melmaruvathur, pp 0144–0147
20. Madhukar BN, Jain S, Satyanarayana PS (2015) Duality theorem for Discrete Hartley Transform. *Int J Appl Eng Res (IJAER)* 10:34699–34703
21. Madhukar BN, Bharathi SH (2019) Duality properties for the variants of the Discrete W Transform (DWT). In: *2019 IEEE international conference on distributed computing, VLSI, electrical circuits and robotics (DISCOVER)*, pp 318–322
22. Hao H, Bracewell RN (1987) A three-dimensional DFT algorithm using the fast Hartley transform. *Proc IEEE* 75:264–266
23. Madhukar BN, Bharathi SH (2019) Similarity properties for the offshoots of the Generalized Discrete Hartley Transform (GDHT). In: *2019 IEEE international conference on distributed computing, VLSI, electrical circuits and robotics (DISCOVER)*, pp 205–209
24. Buneman O (1987) Multidimensional Hartley transforms. *Proc IEEE* 75:267
25. Popovic M, Sevic D (1994) A new look at the comparison of the Fast Hartley and Fourier Transforms. *IEEE Trans Sig Process* 42:2178–2182
26. Buneman O (1986) Conversion of FFT's to Fast Hartley Transforms. *SIAM J Sci Comput (SISC)* 7:624–638
27. Narasimhan SV, Veena S (2020) Signal Processing. Narosa Publishing House, New Delhi
28. Sorensen H, Jones D, Burrus C, Heideman M (1985) On computing the discrete Hartley Transform. *IEEE Trans Acoust Speech Sig Process* 33:1231–1238
29. Madhukar BN, Jain S (2015) A duality theorem for the Discrete Sine Transform (DST). In: *IEEE international conference on applied and theoretical computing and communication technology (iCATccT–2015)*. Davanagere, pp 156–160



# A Deep Learning Approach for Detecting and Classifying Cancer Types



G. Murugesan, G. Preethi, and S. Yamini

**Abstract** Cancer is one of the deadly diseases worldwide. It takes a lot of time and effort to detect it at the early stages. Various pathological and imaging techniques are widely used by doctors/physicians to ensure the presence of cancerous cells in the body. The techniques currently being used are highly time-consuming and lack accuracy. This paper deals with applying deep learning algorithm to detect the presence of different types of cancer. Brain, lung, and skin cancer are determined using Convolution Neural Networks (CNN). By utilizing pre-trained weights of VGG16, the model is trained and fine-tuned with cancerous and normal CT scan images. Experimental evidence shows that the adopted transfer learning (such as using pre-trained weights of VGG16) method provides better accuracy than traditional methods. The accuracy obtained is greater than 95%.

**Keywords** Deep learning · Cancer detection · Types of cancer · Deep learning and cancer · VGG16 · Detection with image · Skin cancer · Brain cancer · Lung cancer

## 1 Introduction

Cancer is a disease which includes uncontrolled growth of cells. They propagate to different parts of the body and invade cells and tissues. They are classified based on the cells in which they grow or originate. They acquire the size and shape of the body cell in which they are impregnated. There are different types and stages of cancer. Early detection always helps in curing the disease. According to the American cancer

---

G. Murugesan (✉) · G. Preethi · S. Yamini  
Department of Computer Science and Engineering, St. Joseph's College of Engineering, Chennai  
-119, India

e-mail: [murugesang@stjosephs.ac.in](mailto:murugesang@stjosephs.ac.in)

G. Preethi

e-mail: [preethigopinath199911@gmail.com](mailto:preethigopinath199911@gmail.com)

S. Yamini

e-mail: [yaminisubramani99@gmail.com](mailto:yaminisubramani99@gmail.com)

society, skin cancer accounts for more than 96,000 deaths, 142,670 deaths from lung cancer, breast cancer for about 42,000, and about 17,800 deaths from brain tumor in the year of 2019. Earlier, microwave scanning techniques and Computer Aided Designs (CAD) were widely adopted for cancer diagnosis. Later, researchers started using machine learning and deep learning approaches for diagnosis and prediction of the disease.

Deep learning is a part of machine learning methodology that are based on neural networks. There are different classes of neural networks like Artificial Neural Network (ANN), Convolution Neural Network (CNN) [1], Deep belief Networks, Recurrent Neural Network (RNN), etc. They are widely applied in different fields of study. They are also used for automating several manual tasks. Cancer research with deep learning [2] is bringing new discoveries every day. With the aid of trained model, various predictions and detections are made.

This paper experiments the usage deep learning model for detecting three different types of cancer. Convolution neural network is utilized for this purpose. The model uses pre-trained weights used in VGG16 architecture (ImageNet) [3] and fine-tuned. This model is then exploited for detecting the presence of cancer in brain, lungs, and skin.

## 2 Literature Survey

A method based on segmentation of sputum cells for detecting lung cancer at the earlier stage was proposed in [4]. It used Bayesian classification for detecting the disease. This method utilized the threshold parameter for prediction. The accuracy of this method was comparatively higher than the earlier rule-based segmentation methods (above 85%). Still, this is not enough for early detection. The first clinical 3D microwave tomography technique for cancer detection was demonstrated in [5]. The result obtained by this microwave imaging system is fairly good. As the power used is 1/1000th of that utilized by cellphones, it is safer; besides, it is also faster. It costs less and is not penetrative, but it requires examination of the images by physicians. A method to efficiently combine different neural network in one architecture was explained in [6]. The accuracy of this method proved to be higher than the individual deep learning models. Yet, it was based only on skin cancer and also took long hours to train the model.

An automated melanoma detection device using image processing techniques was proposed in [7]. The skin lesion images were studied, and during testing, various image processing techniques were applied to classify if the lesions are of melanoma type. Features such as color, shape, diameter are compared, and finally, the result is obtained. This is comparatively faster than traditional methods which involved biopsy and analysis of histopathological images in the laboratory. A deep learning method in [8] uses thoracic MR images for cancer detection. A faster Region-based Convolution Neural Network (R-CNN) was developed with spatial three channeled input, optimized parameters, and transfer learning. This was developed for lung

nodule detection. Experimental results showed that the above devised method could detect most of the nodules. Perhaps, the false positive areas are greatly reduced by this method. In [9], the authors studied how to utilize the deep learning models and transfer learning for brain tumor detection. In this paper, two different methods were estimated. Two pre-trained networks, namely Inception-v3 and Densnet 201 were taken into consideration. Features were extracted from inception modules and densnet blocks, concatenated and then passed to softmax classifier for classification. They are evaluated individually. Both these methods provided reasonable accuracy in detecting brain tumors.

### 3 Materials and Methods

#### 3.1 Dataset

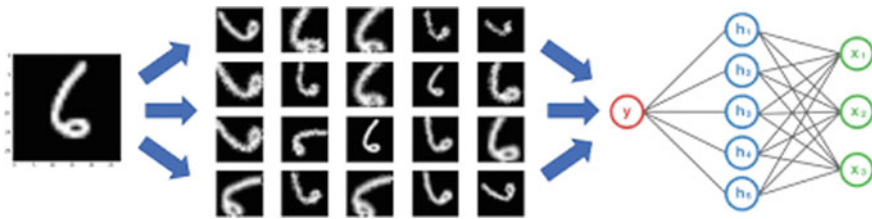
In this work, images collected from Google are used. CT scan images and dermoscopic images in .jpg and .jpeg formats are collected. In total, 1595 images were used; of which 260 brain tumor, 268 normal brain images, 276 lung cancer images, 263 normal lung images, 262 normal skin, and 266 skincancer images are collected and organized in six separate folders. The folders are properly named such that they act as labels during training. Of these images 75% are used for training and 25% for testing and validation. Python libraries such as Tensorflow, keras are used for training the model.

#### 3.2 Pre-processing

In this stage, two important pre-processing steps are performed. Initially, all the images collected are resized to standard size 224 \* 224 px. Then, the pixels values are transformed into numpy arrays, so that they could be processed in the neural network.

#### 3.3 Data Augmentation

Data augmentation [10] as shown in Fig. 1 is a strategy that enables practitioners to significantly increase the data diversity. Augmentation is dynamically performed while training the network. Despite the availability of large datasets, extracting the proper data for the problem domain is highly important. Thus, by improving the diversity, the performance of the resulting model will be improved. In addition, augmentation also helps to avoid overfitting of the model to the collected datasets.



**Fig. 1** Data augmentation for deep learning

### 3.4 Deep Learning Model

Our proposed system is based on convolution neural networks. CNN is one of deep neural networks utilized for image data analysis. They are composed of neurons, each with various weighted inputs and biases. Every neuron gets various inputs. They add their weights up, applied under activation function and provide the result in the output nodes. A set of neurons make up a layer. The first layer, convolution layer is the heart of all layers, where major computation occurs. Filters of various dimensions are applied on images. After applying the activation function, it is fed into pooling layer.

This system focuses on the technique called “Transfer Learning.” Transfer learning [11] is the art of using pre-trained model on some other data to solve the problems in hand. Our feature extractor uses pre-trained weights of VGG 16 [12] architecture for the base layers. VGG 16 architecture composed of 16 layers, where the convolution layers, pooling layers, fully connected layers are repeating over and over again and finally provides the output. Experimental evidences show that VGG 16 has high accuracy in classifying different classes of images. As this system attempts to detect three different types of cancer that are in different location of body, this architecture provides satisfying results in detection and classification.

### 3.5 Training and Fine-Tuning

The neural network is trained by employing the pre-trained weights of VGG 16. The images are sorted in different folders according to their classes. The folders are named as brain\_tumor, normal\_brain, lung\_cancer, normal\_lung, skin\_cancer, normal\_skin. The folder names are used as labels during training. These labels are transformed into binary vector values for processing. Then, data augmentation is applied dynamically alongside, helps narrowing overfitting of the model. The base layers except the top fully connected layers are froze, and the head layers are built. Nearly, 15 epoches are performed with over 48 iterations in one epoch and with a batchsize of 32. An epoch refers to one cycle through the full training dataset. It is equal to one forward pass and one backward pass over the entire network. Then, the

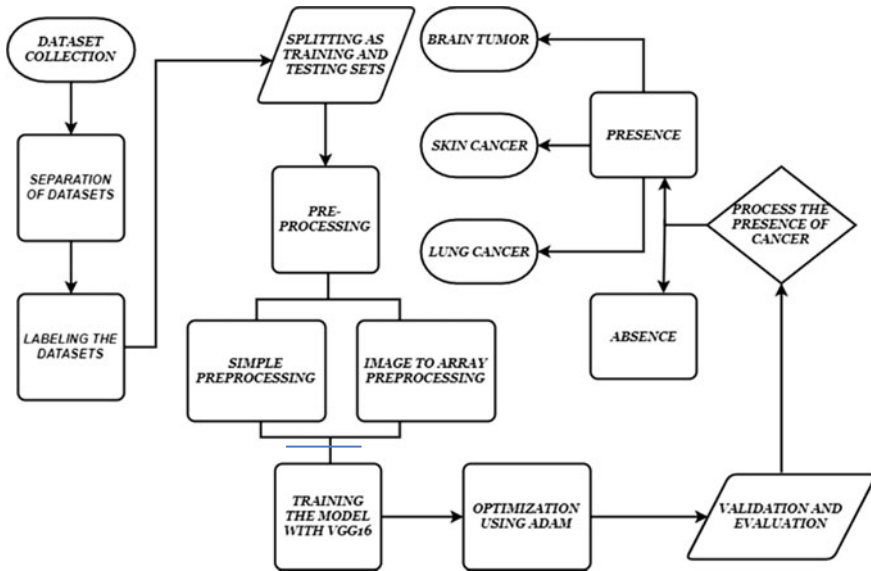


Fig. 2 Flow diagram of the proposed methodology

results are observed. Later, the lower layers are unfroze, and once again 10 epochs are performed. Unlike before, the entire network is trained now. This benefits in increasing the accuracy. This method is called fine-tuning; once training is done, the model is validated for results. While compiling the model, optimizers such as Adam, SGD are used to reduce losses and to speed up the learning process. The model thus developed is imported and applied for disease diagnosis.

### 3.6 Flow Diagram

Fig. 2 shows the pictorial representation of the entire process involved in the development of a model for automatic diagnosis of cancer in brain, lungs, and skin.

## 4 Results and Discussion

### 4.1 Performance Metrics

Accuracy of the model alone is not sufficient to evaluate its performance. Performance of the developed system is also assessed using other metrics such as precision, recall, fi score, and support. They are explained below:

**Table 1** Results at the end of fine-tuning

Classes	Precision	Recall	f1-score	Support
Brain tumor	1.00	1.00	1.00	27
Brain normal	1.00	1.00	1.00	122
Lung cancer	0.99	1.00	1.00	96
Lung normal	1.00	1.00	1.00	89
Skin cancer	0.99	1.00	0.98	28
Skin normal	1.00	0.96	0.99	32
Avg/Total	<b>0.99</b>	<b>0.98</b>	<b>0.98</b>	<b>394</b>

- $ACCURACY = TP + TN / TP + TN + FP + FN$
- $PRECISION = TP / TP + FP$
- $RECALL = TP / TP + FN$
- $F1\ SCORE = 2 * (Recall * Precision) / (Recall + Precision)$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. After fine-tuning the model, the results obtained are tabulated.

## 4.2 Training Results

The results of the proposed transfer learning method, using VGG 16 algorithm is shown in Table 1. As mentioned before, 25% of the data are used for testing and validation. The accuracy obtained is about 0.981, which is approximately 98%. Accuracy indicates how close the results are to the "true" values, whereas precision indicates how they are close to each other. Precision of the model built is also high. The support value indicates the actual number of true positives in each class of testing data. The trained model provides reliable results. Hence, it could be used for disease diagnosis.

## 5 Conclusion and Future Scope

Thus, the proposed methodology makes use of transfer learning in neural networks and detects the presence of cancer in brain, skin, and lungs effectively, using the scan images. When a scan image is passed as input to the system, the presence or the absence of the disease is presented as result (output). Thus, this technique is effective and helps in early detection of the disease. As transfer learning is adopted, the time taken to train the model is comparatively less. Also, the accuracy of the results obtained is highly dependable. In future, in addition to different types, a deep learning model to identify different stages of cancer in particular region of the body could possibly be developed, as an advancement to this proposed system.

## References

1. O'sher K, Nash R (2015) An introduction to convolution neural networks. Neural and Evolutionary Computing, Cornell University
2. Chaudhary K, Poirion OB, Lu L, Garmire LX (2018) Deep learning based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res* 24(6):1248–1259
3. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: *IEEE conference on computer vision and pattern recognition*, pp 258–255
4. Werghi N, Donner C, Taher F, Al-Ahmad H (2012) Detection and segmentation of sputum cell for early lung cancer detection. In: *19th IEEE international conference on image processing*, pp 2813–2816
5. Grzegorzczak TM, Meaney PM, Kaufman PA, diFlorio-Alexander RM, Paulsen KD (2012) Fast 3-d tomographic microwave imaging for breast cancer detection *IEEE Trans Med Imaging* 31(8):1584–1592
6. Harangi B, Baran A, Hajdu A (2018) Classification of skin lesions using an ensemble of deep neural networks. In: *40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp 2575–2578
7. Shukran MAM, Ahmad NSM, Ramli S, Rahmat F (2019) Melanoma cancer diagnosis device using image processing techniques. *Int J Recent Technol Eng (IJRTE)* 7(5S4):490–494
8. Li Y, Zhang LI, Chen H, Yang N (2019) Lung nodule detection with deep learning in 3D thoracic MR images. *IEEE Access* 7:37822–37832
9. Noreen N, Palaniappan S, Qayyum A, Ahmad I, Imran M, Shoaib M (2020) A deep learning model based on concatenation approach for the diagnosis of brain tumour. *IEEE Access* 8:55135–55144
10. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning. *Computer vision and pattern recognition*, Cornell University
11. Wang J, Chen Y, Yu H, Huang M, Yang Q (2019) Easy transfer learning by exploiting intra-domain structures. *Machine Learning*, Cornell University
12. Simonyan K, Zisserman A (2015) Very deep convolution networks for large scale image recognition. In: *Computer vision and pattern recognition*, Cornell University

# Automated Surveillance Security System Using Facial Recognition for Homes and Offices



Gunjan Bharadwaj, Shubham Saini, Anmol, Ajeet Chauhan,  
and Puneet Kumar

**Abstract** Video cameras are the most popular way of security used by the majority of people in homes and offices. These systems require a human being for video surveillance so that if some unusual activity happens, appropriate action can be taken. Usually, cameras are installed to monitor the unusual activities which may be a threat to the workplace. In the case of home security, the major reason to install video cameras is to monitor that no unknown person should enter the home without permission. This demand for home security systems has been utilized to provide automated surveillance security system which can specifically be used in such places to monitor the presence of an unknown person in the premises so that the people using this system are alerted using notification. For identification of unknown face, our system has been trained on some known faces using LBPH face recognizer and then the threshold is applied to decide whether the face is known or unknown, finally an alarm is generated if the face is declared unknown. The above system works well on good quality videos and images.

**Keywords** Home security · Video automated surveillance · Face recognition · Face detection

## 1 Introduction

The automatic security systems are in great demand in today's era. Automated surveillance security systems are added in places nowadays to monitor unusual activities. Many techniques are being developed which can automatically detect such unusual activities instead of manually monitoring the videos. Therefore, automatic video surveillance systems have got major advantages of eliminating the need for manual surveillance. CCTV is one of the most used hardware for security systems in

---

G. Bharadwaj (✉) · S. Saini · Anmol · A. Chauhan · P. Kumar  
Department of Computer Engineering and Applications, GLA University, Mathura, India  
e-mail: [gunjan.bhartiya@gla.ac.in](mailto:gunjan.bhartiya@gla.ac.in)



homes, offices and public places. One obvious demand for CCTV in homes/offices is that there is a known set of people who are going to enter the house. This demand has been utilized for home security systems to provide automated surveillance security system.

Face Recognition appears to have a lot of applications in computer science and therefore, stays at the heart of today's AI systems. Many researchers across the world are focusing on applications of facial recognition in video surveillance [1, 2]. Various other approaches have used facial recognition and shown its importance in video surveillance applications [3, 4]. There exist various face recognition methods developed by various researchers. Face recognition algorithms are known to extract the meaningful geometric features from an image, represent them and perform classification on them. Facial recognition algorithms use the most intuitive features and process them for human identification.

An approach can be built which uses automatic detection of unknown faces entering homes/offices. For the detection of unknown faces, LBPH facial recognition is used [5]. The face of the known person will be able to match who is entering the premises with an existing database of known faces whereas unknown faces will not be able to find a match with the database.

Section 2 describes the proposed method, Sect. 3 describes the implementation and includes the results generated by the proposed method followed by a conclusion in Sect. 4.

## 2 Proposed Method

As stated earlier, an approach is intended to build which can be utilized for automated security surveillance at places where there is a known set of people who can safely enter the premises whereas an unknown person can be identified.

The whole process of the proposed system can be divided into three major steps; the first step is to construct a database of known faces with multiple images for each known face. The images are taken from different angles. The next step is to use face detection on the images stored in the database and then use the LBPH face recognizer for the training of the system. Next and final step is to test the built trained model to recognize the faces whether known or unknown. The confidence of a face recognizer is low in some images which are unknown to the trained model whereas high for those faces which are known. A threshold is used to differentiate a known face from an unknown face.

The block diagram of the proposed approach is as shown in Fig. 1. Each step is explained in detail in the following subsections:

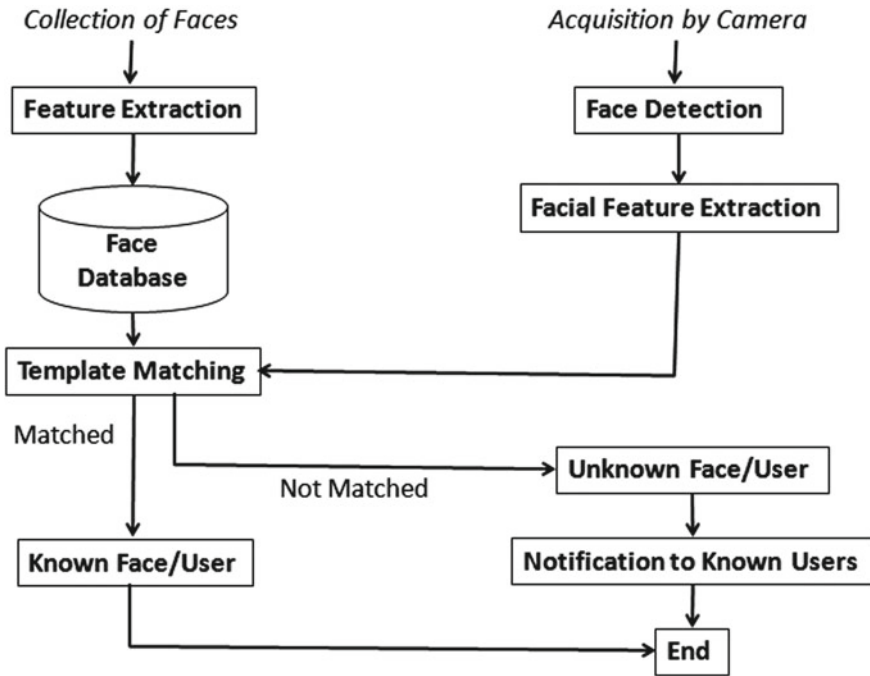


Fig. 1 Block diagram of unknown face identification using facial recognition

### 2.1 Face Detection

This feature comes when a person will come in the range of the camera. His/Her face will be detected by the system and this detection will be done based on a pattern of the face. Haar Cascade classifier is used for this purpose [6]. The eyes, nose, mouth and face pattern will help the system to detect the face. The output of face detection will be the input for feature extraction.

### 2.2 Feature Extraction

After face detection, specific features will be extracted from the output of face detection. A specific analysis will be done to find a specific pattern of the face. This will help the system to recognize the face whether it is known or unknown this will be decided in further steps.

### ***2.3 Face Database and Template Matching***

Feature extraction is followed by template matching. The detected face will be matched with our Face Database. In this database, first of all, those faces will be stored who will be having the permission of IN and OUT to the premises. Only the administrator has the permission to add/edit/delete the faces from the database. The system will be trained by using these faces from the database. The matching is done based on the specific information of different parameters that includes eyes, nose, mouth and face pattern. LBPH face recognizer will be used since it uses both texture and shape information based on local binary patterns; histograms are extracted for small regions and converted into a single feature vector.

### ***2.4 Result***

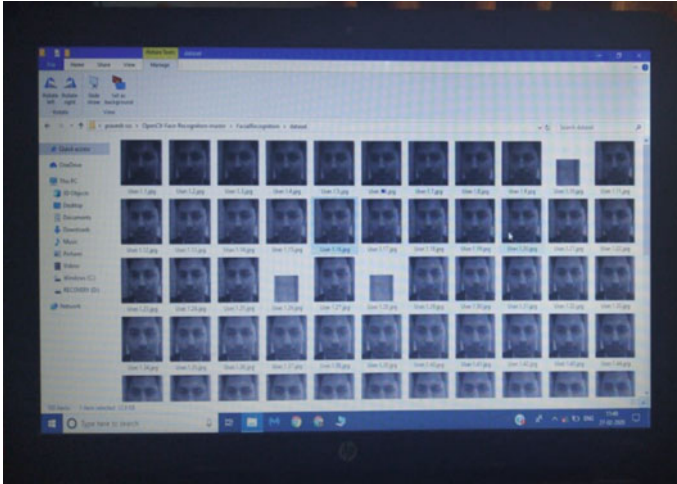
If the feature vector of the input face image is matched with the face database then the detected face is known to the system. If the information is unmatched then the system will trigger a notification to the owner of the premises.

## **3 Implementation and Result Analysis**

The proposed model has been designed on the OpenCV framework using Python 3+. Scanning of faces is done through the in-built webcam. All information of the detected face is transferred into binary format. The transformed representation marks for their respective categories and maximum marked category is responsible for the detection of the identity of the face. The designed model works efficiently. Following steps represent the working of the proposed model and finally the resulting output by the system.

### ***3.1 Input Acquisition***

The first step starts with the image acquisition in which webcam is used to take the image as an input. The webcam captures the entering entity in the premises. The faces will be detected from the webcam video and then the information will be retrieved from the cropped image that includes the face of the entity. The information extracted in this step will be provided to the next step.



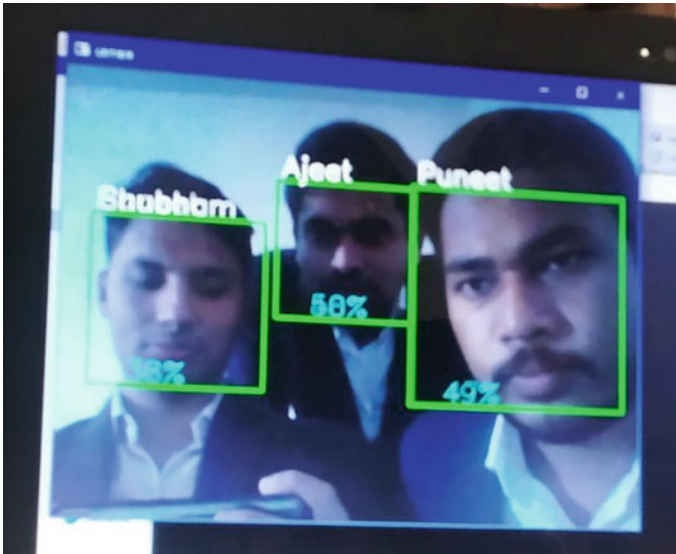
**Fig. 2** Training data

### ***3.2 Training***

Training of the faces is done by taking different angles and observations of the cropped image provided by the image acquisition step. Every known member of the premises sits in front of the camera at the time of training to add their faces in the database. The camera observes different angles and observations of the sitting member (Fig. 2). The algorithm of feature extraction is run on each of these faces using LBPH face recognition as explained in Sect. 2.3. These feature vectors of each known face are added to the system's known faces database set and will further be used to match the identity of the person entering the premises and can thus result in efficient identification of the entrant.

### ***3.3 Face Recognition***

After completion of training, our system is tested with any of the faces trained in the previous steps. Now a new face is streamed in the video. The face is detected from the video by Haar Classifier (Fig. 3) [7]. Then, features for the input face are constructed which called as face recognition. This face feature vector generated using LBPH classifier will further be used to match with the existing feature vectors of know faces in the database.



**Fig. 3** Detection and recognition of faces in a database

### **3.4 Results**

As soon as the face of the entrant is detected, the facial feature vector is analyzed and matched with the existing feature vectors of known faces in the database. If the entrant's face is matched with the existing person's face, the system will take no action. But if the entrant is unmatched, it will send a notification to the owner about the suspicious entrant by triggering a notification. This notification will be sent to one of the known users in the form of SMS or alarm based on the user's choice. The system is capable of finding known or unknown faces as shown in Figs. 3 and 4 respectively.

## **4 Conclusion**

The proposed system requires only a webcam and database storage which makes it financially effective in terms of manpower and time as well as easing the maintenance. The system is highly efficient in processing real-life videos. The system triggers the owner if any unknown entrant is on the premises as suspicious. This method can also be built to customize itself for a public place or in crime investigation or other applications as a future scope. The proposed method can only be applied at places where the place is known to have a common set of known people and an unknown person can be a threat to the premises. Face occlusion is a common phenomenon



**Fig. 4** Non-trained faces identified as unknown

while using facial recognition with the video camera, so as part of improving the quality of results, some robust methods of facial recognition can be used. Also, the expressions and angles of the entity can be detected with more accuracy and more observations.

## References

1. Cottrell GW, Fleming MK, Face recognition using unsupervised feature extraction. Submitted to INNC-90, Paris
2. Kanade T (1973) Picture processing system by computer complex and recognition of human faces. Ph.D. Thesis, Department of Information Science, Kyoto University
3. Zhang D, An P, Zhang H (2018) Application of robust face recognition in video surveillance systems. *Optoelectron Lett* 14(2):152–155
4. Mileva M, Burton AM (2019) Face search in CCTV surveillance. In: *Cognitive research: principles and implications*, vol. 4, no. 37
5. Suma SL, Raga S (2018) Real time face recognition of human faces by using LBPH and Viola Jones algorithm. *Int J Sci Res Comput Sci Eng* 6(5):06–10
6. Sander AS, Object detection using haar-cascade classifier. In: *Institute of computer science, University of Tartu*
7. Varun A, Garg A, Kririka A, Garg A (2016) Face recognition using haar-cascade classifier. *J Emerg Technol Innov Res* 3(12)
8. Gonzalez RC, Woods RE (2008) *Digital image processing*, 3rd edn. Prentice Hall, New Jersey

# Strategies for Boosted Learning Using VGG 3 and Deep Neural Network as Baseline Models



K. S. Gautam, Vishnu Kumar Kaliappan, and M. Akila

**Abstract** Deep Convolutional Neural Network learns with various levels of abstraction and has made a revolution among learning algorithms. Deep Learning currently plays a vital role in object classification, natural language processing, genetics, and drug discovery. The deep learning unravels the patterns by computing gradients to minimize the loss function and based on it the internal parameters are tuned to compute the layer-wise representation. The Deep Convolutional Neural Network has revolutionized computer vision and image processing. The work cross dissects the deep convolutional neural network to light its learning mechanism and extensively experiments parameter tuning to facilitate intelligent learning. The work comes up with a roadmap to build and train a deep convolutional neural network after extensive experimentation using CIFAR 10 dataset. The work also comes up with near-optimal hyperparameters that effectively generalize the learning of Neural Network. The performance of the Deep Neural Network is also evaluated with hypertension dataset gathered from the health department. On experimentation, could be inferred that the proposed approach gives comparatively higher precision. The accuracy of the proposed approach is found to be 90.06%.

**Keywords** Deep convolutional neural network · Intelligent learning · Parameter tuning · Dropout · Data augmentation · Batch normalization · Learning rate · Variance

---

K. S. Gautam (✉)

Department of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India

e-mail: [gautam2ping@gmail.com](mailto:gautam2ping@gmail.com)

V. K. Kaliappan · M. Akila

Department of Computer Science and Engineering, KPR Institute of Engineering and Technology Coimbatore, Coimbatore, India

e-mail: [vishnukumar.k@kpriet.ac.in](mailto:vishnukumar.k@kpriet.ac.in)

M. Akila

e-mail: [akila.m@kpriet.ac.in](mailto:akila.m@kpriet.ac.in)

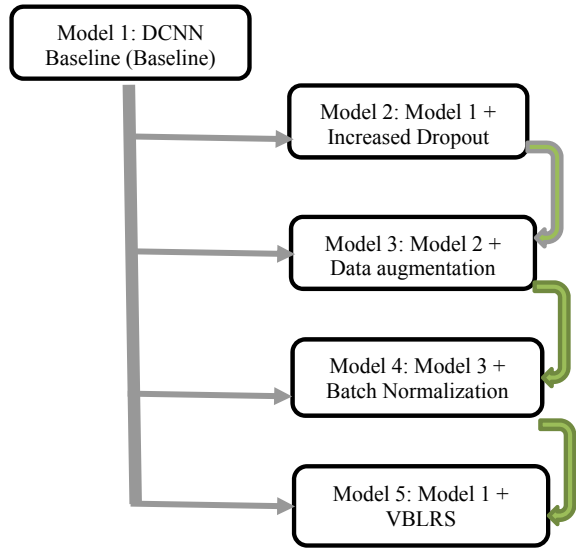
# 1 Introduction

The conventional machine learning approach works with feature engineering and demands manual intervention and expert intelligence for model building. Whereas the feature engineering is comparatively low in deep learning as the need for feature descriptors are not required. The deep learning offers flexibility by representation learning. Here the machine acquires the raw data and discovers multiple representations. The deep learning offers a generic general-purpose learning procedure to learn representations. Initially, the machine identifies the light and dark pixels and from the detected light and dark pixels, the system identifies the edges and possible shapes. From simple shapes, the machine identifies many complex shapes and at last the machine learns the region of interest. In other words, can say that the machine identifies a shallow layer, middle layer, and deep layer that is a step by step improvement from a low level to high-level structure. For a machine to learn layer-wise representation effectively the internal parameters should be tuned. In other words, fixing near-optimal parameters is dealt with that contribute to minimal loss and higher classification accuracy. Additional focus is also given to the learning rate because if it tunes smaller the network loses its learning ability. The work performs a series of experimental analyses on CIFAR 10 [1] using various parameters and hyperparameters including learning rate scheduler and the parameters are evaluated and its corresponding value that boosts performance. With the proven near-optimal parameters and hyperparameters of boosted VGG3, The work also attempts to model an intelligent machine using a neural network that could detect hypertension. The data have been gathered regarding hypertension directly from the hospitals and Primary Health Centres. As per the suggestion from the physicians, the features that contribute to the hypertension are Age, Sex, Marital Status, District, Locality, Religion, Prior hypertension occurrence of both the individual and family, Complication of hypertension, Prior occurrence of diabetes Mellitus of both individual and family, Height, Weight, BMI, Pregnant status, Waist Circumference, Systolic BP, Diastolic BP. The reason for choosing Neural Network is that they don't have restrictions towards any input variables like other learning algorithms and they can model nonlinear and complex relationships [2].

The following sections explore the approaches tailored (as shown in Fig. 1) to bring new models that show significant improvement in performance. A baseline model is built in Sect. 2. Section 3 discusses the baseline model built with increased dropout. Section 4 extends the model built in Sect. 2 by adding data augmentation. Section 5 extends the model built in Sect. 3 by adding a batch normalization technique. Section 6 extends the model built in Sect. 5 with a proposed strategy named variance-based learning rate scheduling (VBLRS). Section 7 builds a neural network and uses the parameters and hyperparameters used to build a boosted VGG3 model and checks how the afore fixed parameters generalize the leaning.



**Fig. 1** Proposed roadmap for boosted learning



## 2 Baseline Model Investigation

CIFAR 10 dataset is widely used to evaluate computer vision algorithms. The dataset has 10 classes as shown in Fig. 2. The dataset has sixty thousand colored images of size  $32 \times 32$ .

The CIFAR 10 dataset has been loaded by Keras [2] and 50,000 images are used for training and 10,000 images are used for testing. As shown in Fig. 3, a plot with the first nine images have been created to check the resolution of the image and could infer that all the images are in low resolution and some demand clarity in representation.

As the images are pre-segmented, they are loaded directly and the ten classes in the image are represented by integers from 1 to 10 by using one-hot encoding [3]. After one hot encoding, the class image appears to be a 10-dimensional binary vector. The pixel values need to scaled for modeling and to accomplish this the pixel values are scaled between 0 and 1. The pixels within the range 0–255 are converted to float and then they are divided by its maximal value so that the training and test data are normalized as shown in Eqs. 1 and 2.

$$g = \frac{f(x)}{255} \tag{1}$$

$$g' = \frac{f(y)}{255} \tag{2}$$

The pixel value for both train ( $x$ ) and test data ( $y$ ) are converted into float  $f(x)$  and  $f(y)$  respectively and they are scaled as down as shown in Eqs. 1 and 2.

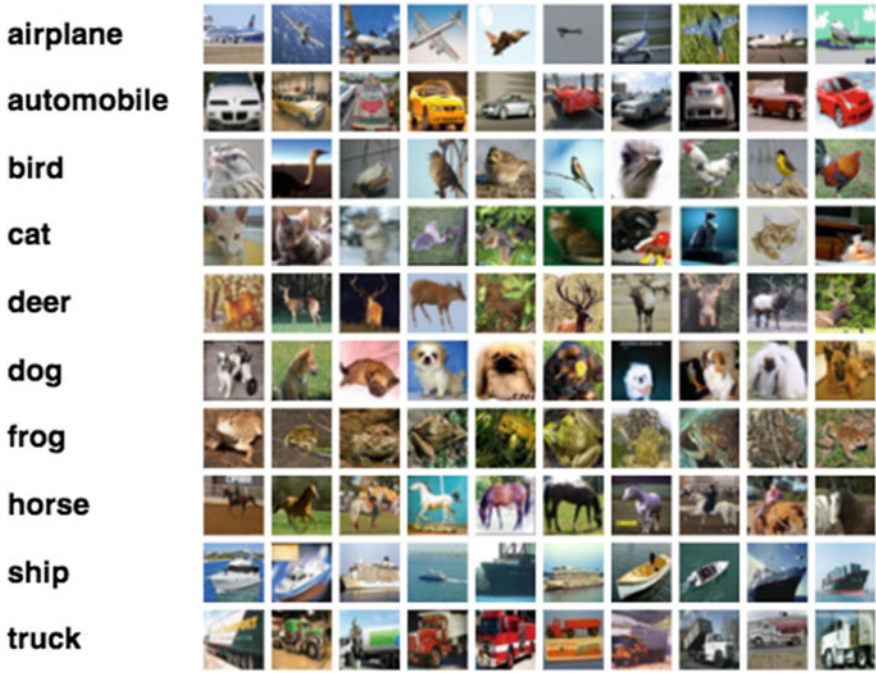


Fig. 2 Class labels in CIFAR 10 dataset

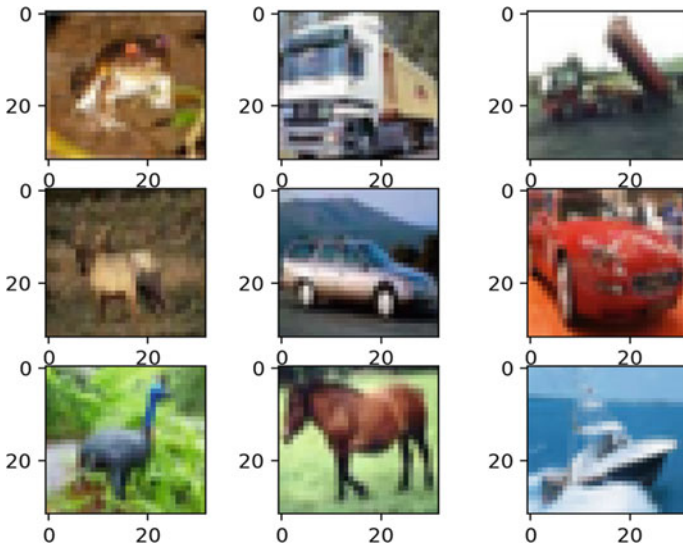


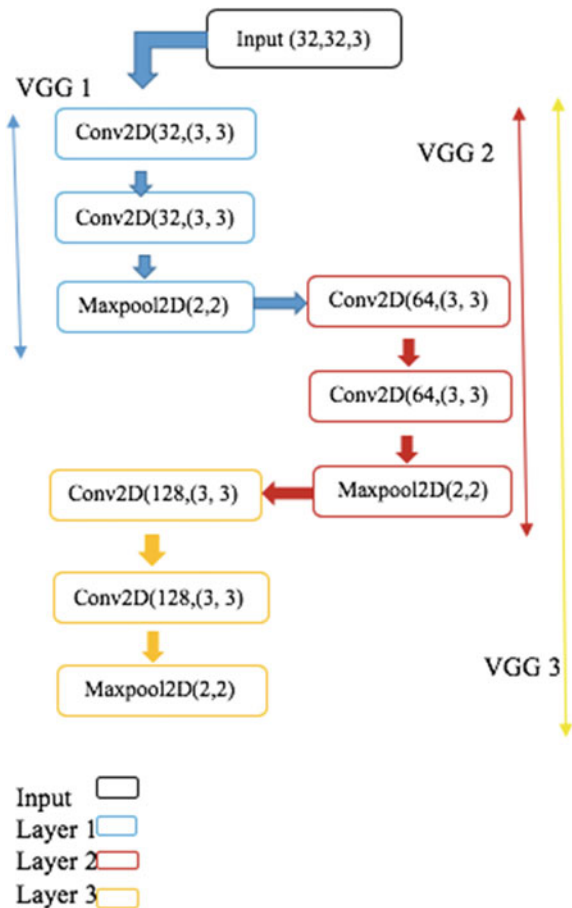
Fig. 3 Sample images IN CIFAR 10 dataset

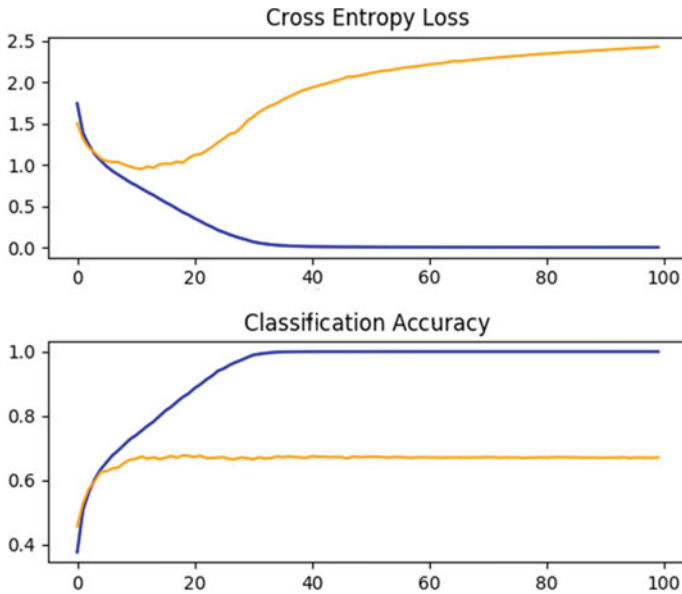
Model 1–VGG 3

The base model uses a triple block VGG architecture. The VGG [4] model has been chosen as the preliminary model based on its modularity, extensibility, and its performance in the ILSVRC 2014 competition. The architecture involves a three-layer stack and each layer has 2 convolutional filters of size  $3 \times 3$  followed by a max-pooling layer as shown in Fig. 4.

Each layer in the model uses Relu activation function with the weight initialization [5]. The afore built model is considered to be a feature descriptor of the model and this model ends with a classifier that predicts the class. The feature map at the last layer is flattened and given as input to the fully connected layer the output layer using the SoftMax activation function. The fully connected layer has 10 nodes as had 10 different classes. Learning rate as 0.001, momentum as 0.9, and epochs as 100 is fixed which are proven to be better by practice [6]. As it is a multi-class

Fig. 4 Extensible VGG architecture





**Fig. 5** Learning curves of VGG 1 (baseline attempt model)

classification problem the model uses categorical cross-entropy loss function. To find the apt baseline model for investigation the experimental analysis is done with a single, double, and triple VGG stack and the results are shown in Figs. 5, 6 and 7. Table 1 shows the inferences made from the experimentation

### 3 Model 2—VGG 3 (Baseline and Dropout)

To address the issue of overfitting, regularization has been adopted in the upcoming models. To slow down the convergence in the upcoming models the works use data augmentation, learning rate schedulers, and change in batch size. The work uses dropout as one of the regularization techniques as it takes control of the randomly muted nodes. In our work, after each max pooling, dropout is added and retain 80% of the nodes. In our work, the dropout value is 0.2 and the addition of dropout has given us a significant raise in performance and has addressed overfitting. From Figs. 8 and 9 it is inferred that the model has convergence between 32 and 52 and after it there is no learning.

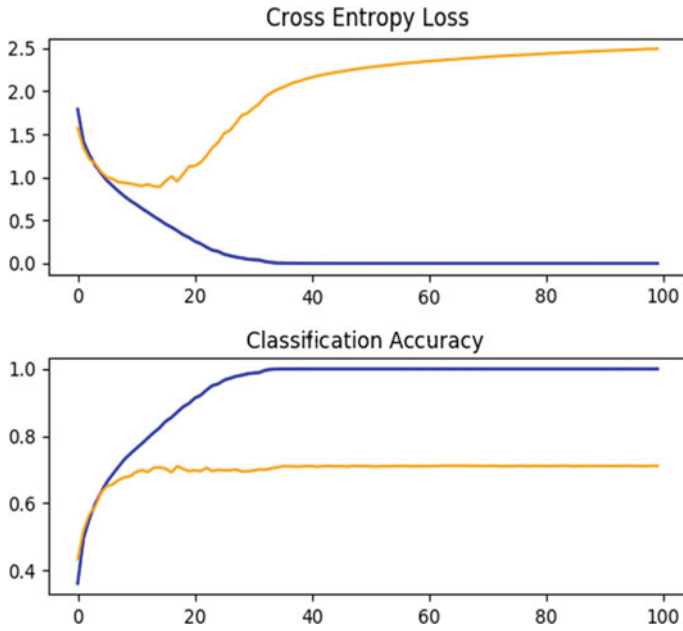


Fig. 6 Learning curves of VGG 2 (baseline attempt model)

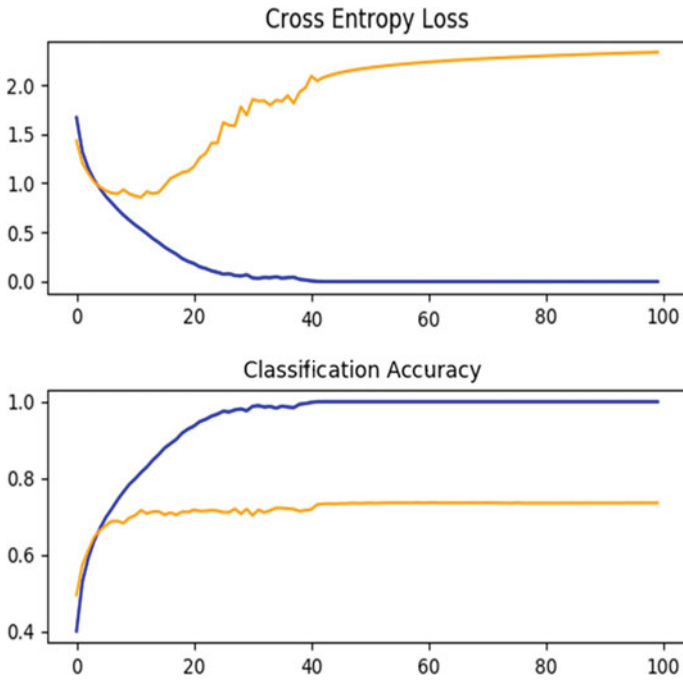
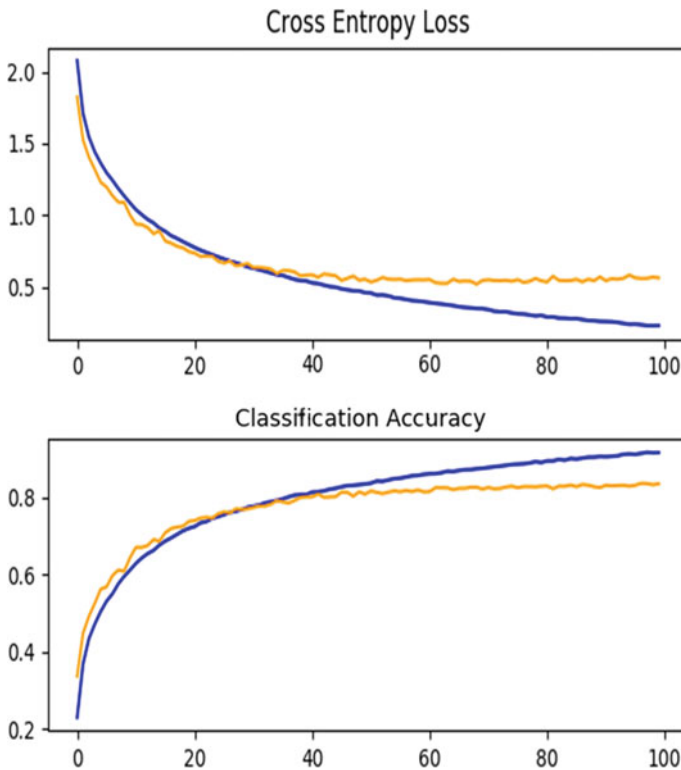


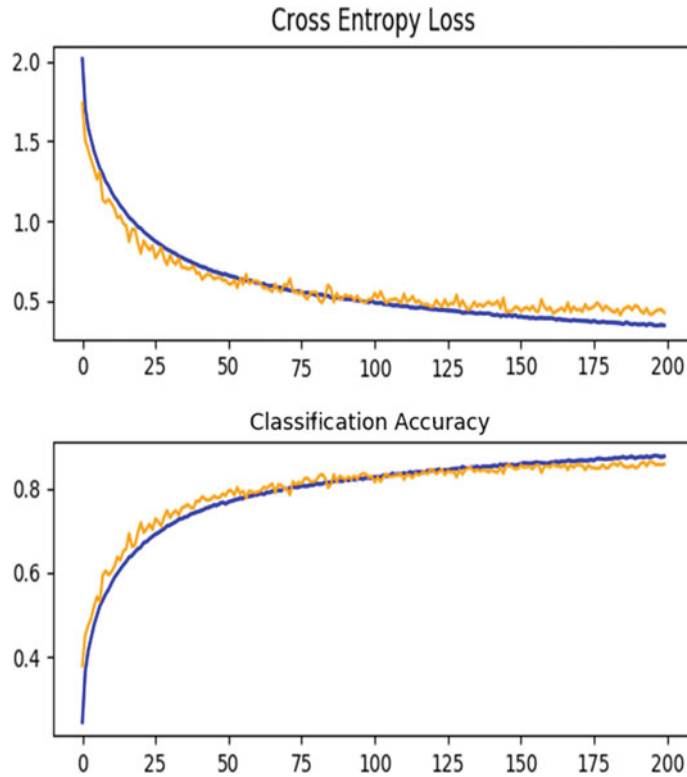
Fig. 7 Learning curves of VGG 3 (baseline attempt model)

**Table 1** Inferences from experimentation (VGG1, 2 and 3)

S. No.	Inferences
1	All the experimented VGG models start learning with accuracy more than 68%
2	From 14th to 22nd epochs the VGG models suffer from profuse overfitting
3	All the VGG models start learning after 35th epoch
4	On stacking more VGGs, the accuracy increases and this trend continues. In other words, on increasing the depth of the VGG model the accuracy increases
5	The model demands techniques that could slow down the convergence to search near-global minima (to minimize the cost function)
6	The VGG models stop learning after the 39th epoch



**Fig. 8** The learning curve of VGG 3 with dropout

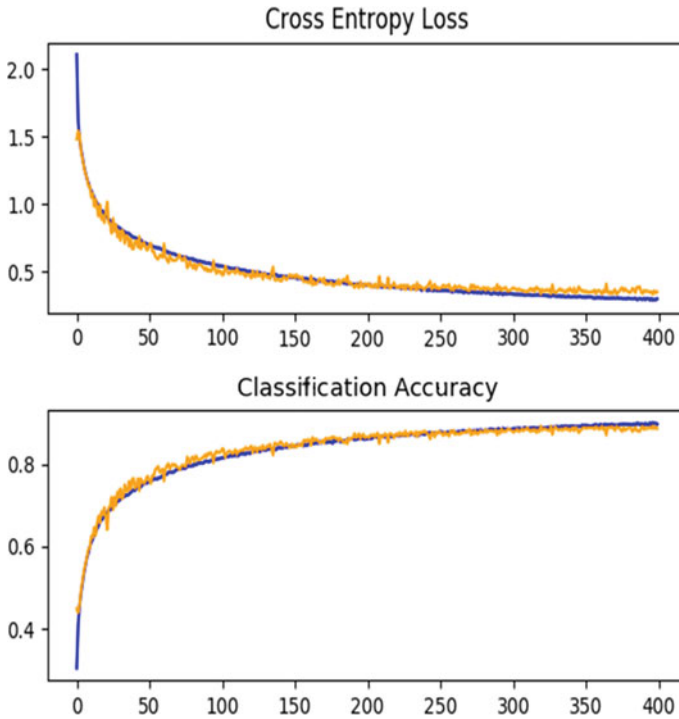


**Fig. 9** The learning curve of VGG 3 with Dropout, and Data augmentation

#### **4 Model 2—VGG 3 (Baseline, Dropout, and Data Augmentation)**

**Data augmentation** techniques involve creating a replica with a minor modification, for ensuring a generalized feature learning. As the images in the CIFAR 10 are small, extreme augmentation techniques that intrude nearly new image shall be avoided. Techniques such as minor zooming, minor shifts, minor flips shall be used. On evaluating the performance of the system, could be inferred that performance of the model is boosted (Fig. 10).

The number of epochs has been increased to 200 to check the possibilities of boosting classification accuracy. On analyzing the learning curve, inferred that the magnitude of the performance boost is similar to the addition of dropout to the baseline model.



**Fig. 10** The learning curve of VGG 3 with dropout, data augmentation and batch normalization

## 5 Model 2—VGG 3 (Baseline, Dropout, Data Augmentation, and Batch Normalization)

Further, another attempt has been taken to boost the classification accuracy. The epochs have been increased from 200 to 400 and batch normalization is added. The dropout is also increased by 0.2 across the layers. On analysing the learning curve, could infer that the model has started learning from 10th epoch to 400th epoch and the learning curve goes on increasing. Thus, the rate of leaning is made higher by tuning the parameters.

In the second approach, a learning rate scheduler has been formulated, in such a way that the amount of variance that the SGD is undergoing is inversely proportional to the learning rate. The mathematical formulation for the variance-based learning rate scheduler is given below in Eq. (2) and from the equation, could infer that the learning rate is made dynamic based on the inputs from variance.

$$\text{Learning rate} \propto \frac{1}{\text{Variance}} \quad (2)$$



**Table 2** Variance based learning rate scheduling

Variance	Learning rate
0.5112	0.15112
0.5103	0.15103
0.5099	0.15099
0.5097	0.15097
0.5096	0.15096
0.4918	0.14908
0.4693	0.14693
0.4128	0.14128
0.3279	0.13279
0.3391	0.13391
0.2921	0.12921
0.1872	0.11872
0.1567	0.11567
0.1455	0.11455
0.1248	0.11248

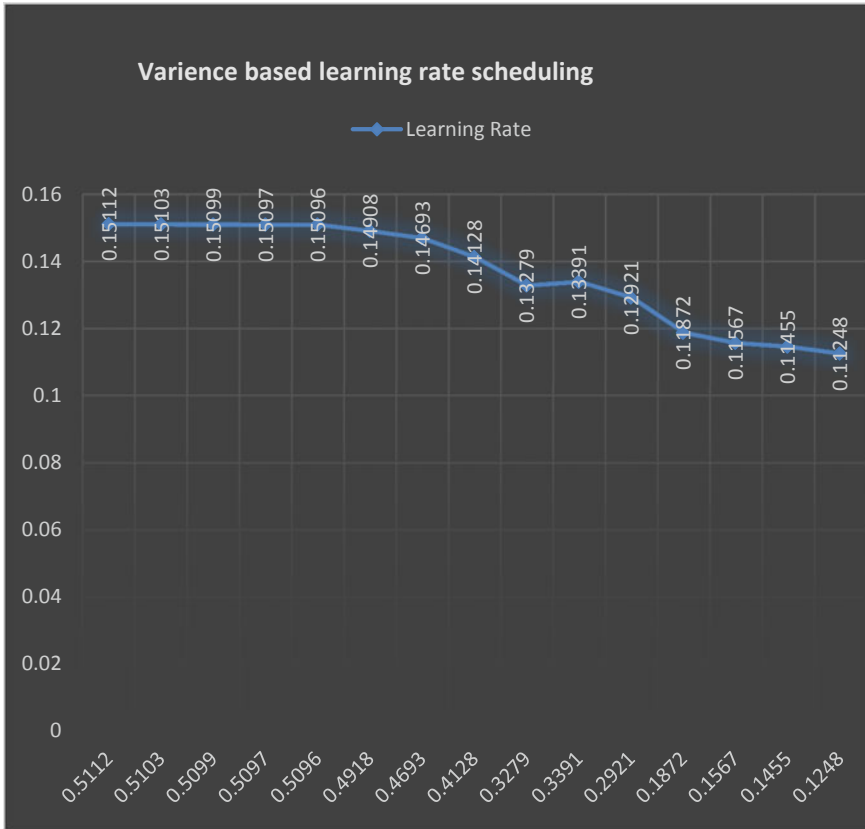
With Eq. (2) as a base, a refined version of the equation is constructed in such a way that when the variance increases the learning rate should lower down accordingly and vice versa. In the proposed Variance based learning rate scheduler, the learning rate is fixed based on the direction of the variance.

$$\text{Learning Reate} = \frac{1}{1 + \text{Variance}} \quad (3)$$

The initial learning rate is set as 0.1, Table 2, shows the learning rate fixed for the set of recorded variances. Equation 4 shows how the learning rate is being updated based on variance.

$$\text{Learning Reate} = \frac{1}{1 + 0.5112} = 1.5112 \quad (4)$$

The relationship between the learning rate and the variance is plotted in Fig. 3 and the X-axis in the graph is the set of variance and Y-axis is the learning rate that is scheduled based on the variance. From the graph, could infer that when the variance decreases the learning rate also decreases (Fig. 11).



**Fig. 11** Variance-based learning rate scheduling

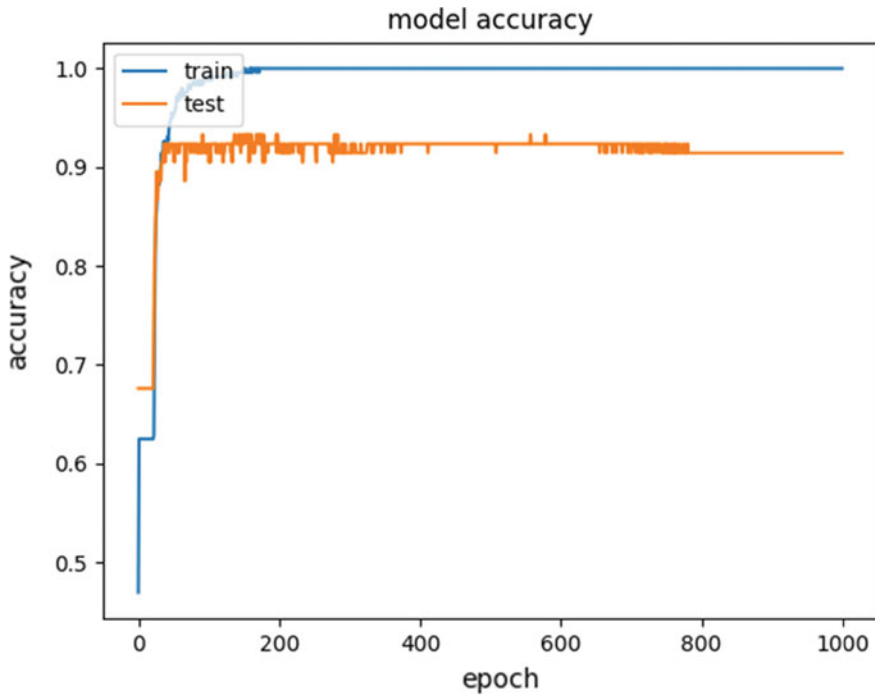
Pseudocode: Variance based learning rate scheduler

Learning rate → LR → 0.01  
 Expected Outcome → X  
 Actual Outcome → x'  
 Variance → Var

```

For each Var compute LR
{
    Var → |x-x'|
    Learning Reate =  $\frac{1}{1 + Var}$ 
}
    
```

The accuracy and the loss for the model that uses a variance-based learning rate scheduler are shown in Figs. 12 and 13 respectively. The validation accuracy drifts above 0.9 at the 51st epoch, which is considered to be one of the fast convergences without compromising accuracy. The model accuracy of the proposed approach is



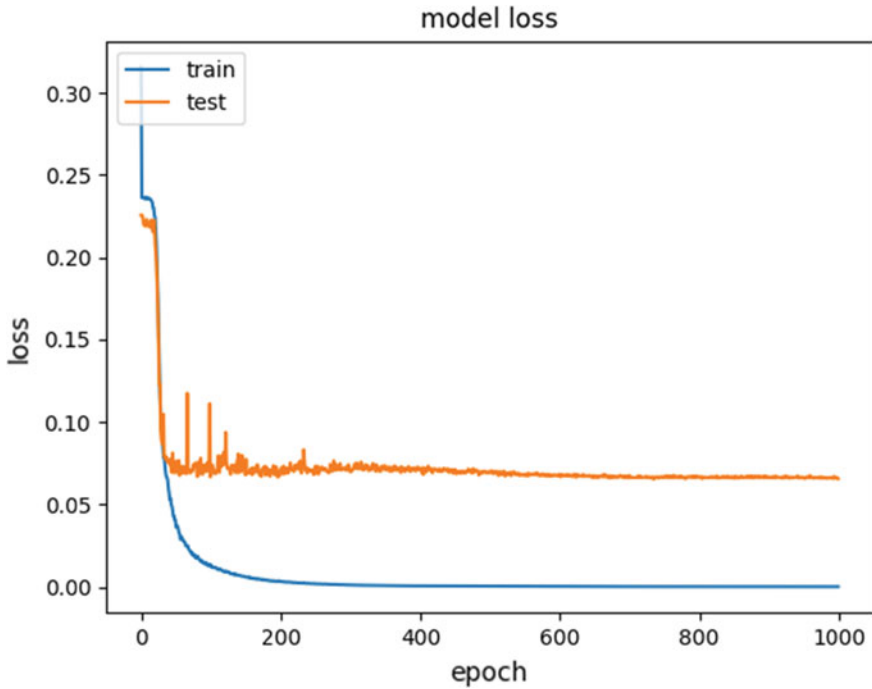
**Fig. 12** Model accuracy of the proposed approach (dropout, data augmentation and batch normalization and variance based learning rate scheduler)

0.9376. The performance of the models built is shown in Fig. 14 and could infer that the performance of the system increases when the parameters are tuned as per the proposed roadmap. There has been a constant boost in accuracy from 69.011% to 91.031%.

## 6 Deep Neural Network

The deep neural network used has a single input layer, two hidden layers, and one output layer. A sequential model is created and all the neurons are connected from one layer to another. The feature vectors of dimension 18 are given as input to the hidden layer 1 and the weights are randomly initialized for eight neurons. The hidden layer 2 has six neurons and the output layer has one neuron that could reveal the status of hypertension.

A sequential model is chosen to add layer one at a time and the first layer has 4 parameters such as many neurons, input dimension, the distribution of weights, the activation function. The next layer has six neurons with uniform initialization of weights and have chosen Relu [3] as the activation function to overcome the

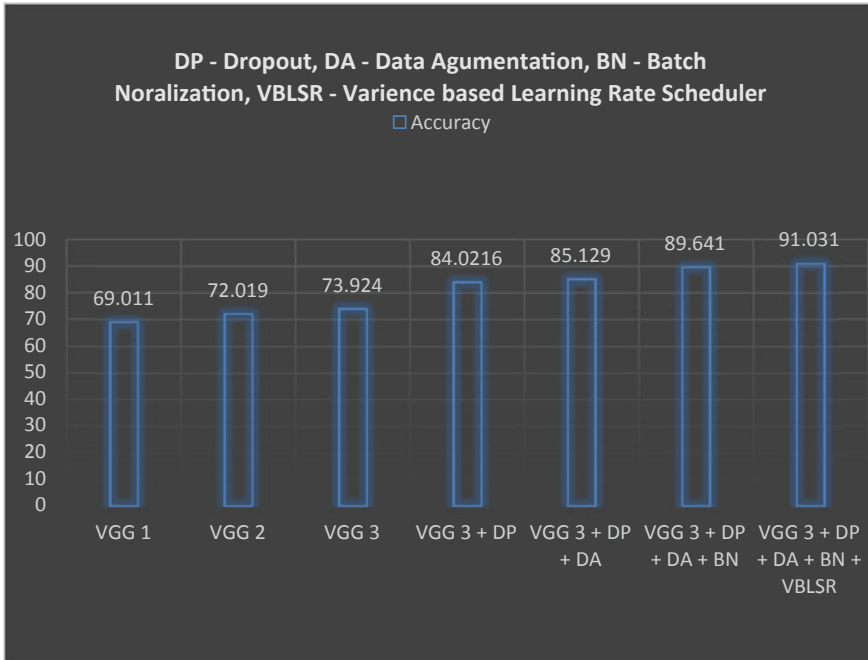


**Fig. 13** Model loss of the proposed approach (dropout, data augmentation, and batch normalization and variance-based learning rate scheduler)

vanishing gradient problem. Relu has been used in our work because it doesn't trigger the neurons at a time. Relu checks the output of the linear transformation and if it is less than zero the neuron gets deactivated. The output layer has four neurons with a sigmoidal activation function. As the aforementioned problem belongs to multi-class classification, categorical cross-entropy has been chosen as a loss function and the learning algorithm used is Stochastic Gradient Descent. In overall absolute mean error is used as a loss function. The number of iterations involved during the training process is 200.

Python is used as a programming platform with Keras [6] in the front end and Tensor flow [7] in the back end. The computing device used for experimentation is MacBook Air with 8 GB RAM. The system is developed using Spyder Integrated Development Environment and the average running time for all the execution platforms is shown in Table 4. The model summary of the network created that acquires unoptimized feature vectors are input as shown in Table 2 and could infer that the network has 229 parameters. The model summary of the network is shown in Table 3.

600-row vectors are used for experimentation and out of it 420 feature vectors are used for training the network and rest 180 are used for testing. Out of 180 feature vectors, the proposed deep neural network detects 148-row vectors and fails to classify 32-row vectors. After feature vector optimization, the proposed approach



**Fig. 14** Performance of the parameter and hyperparameter tuned model (comparative analysis)

**Table 3** Network model summary after feature optimization

Model: "sequential\_2"

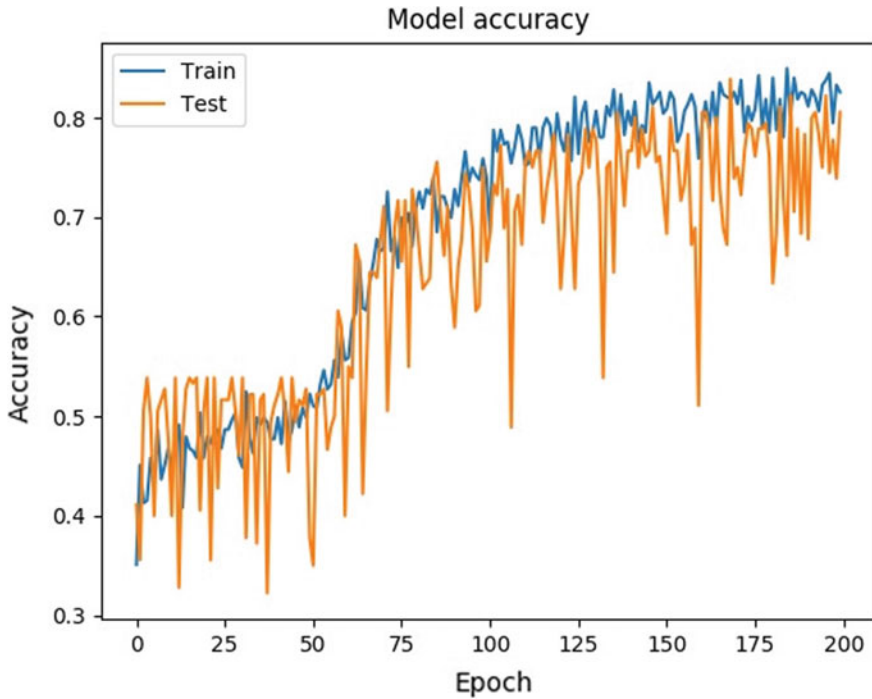
Layer specifications	Output shape	Parameter
Dense layer 1	(Nil, 8)	48
Dense layer 2	(Nil, 6)	54
Dense layer 3	(Nil, 1)	7

Total number of parameters: 109, Trainable parameters: 109 Non-trainable parameters: 0

correctly classifies 174 features and fails to classify 2 feature vectors. Based on the aforementioned results, the accuracy of the proposed approach is calculated in Table 4.

**Table 4** Performance evaluation of the existing and the enhanced approach

Algorithm	True positive	False positive	Accuracy	False discovery rate
Neural network (with near-optimal parameters and hyperparameters)	174	6	86.01	3.2
(Proposed)				

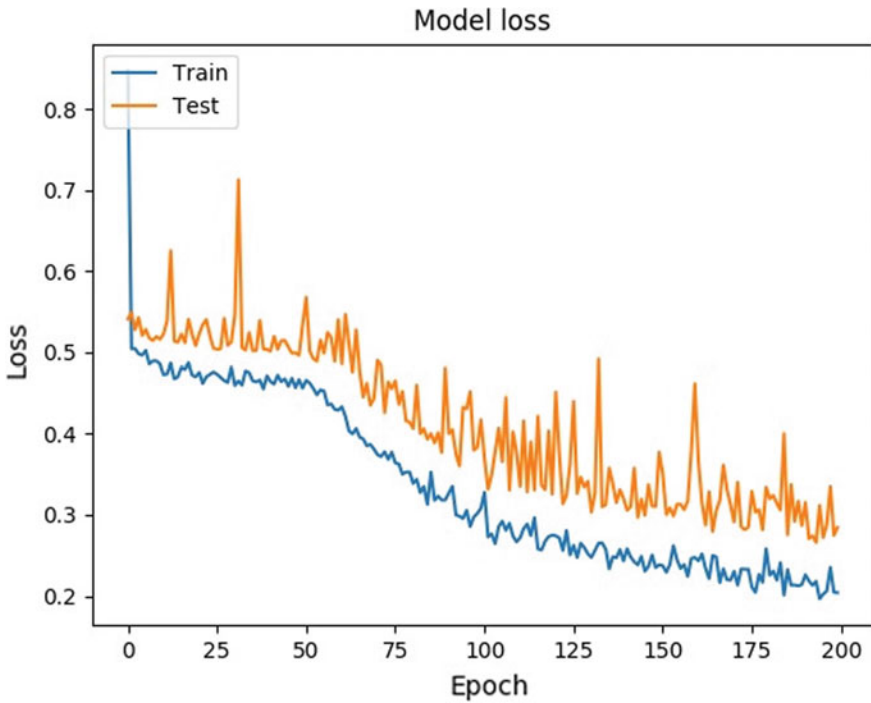


**Fig. 15** Accuracy of the feature optimized deep neural network

The performance of the existing Neural Network is evaluated. While model building both the networks took 200 epochs for convergence. From Fig. 4, could infer that the accuracy of the existing approach is 0.55 and the model loss is calculated in categorical cross-entropy and it is 0.59 as shown in Fig. 5. From Fig. 5 could infer that the difference between the expected and actual outcome is more than half. Once the optimized features are given as input to the Deep Neural Network, could infer that the network converges at the 151st epoch with an accuracy of 86.06% and categorical cross-entropy 0.21 as shown in Figs. 15 and 16 respectively. In the proposed approach the loss function or categorical cross-entropy is comparatively lower than the existing architecture.

## 7 Conclusion

The performance of the baseline, parameter, and hyperparameter tuned models are evaluated. From the experimental analysis could infer that there is a gradual growth in learning when the parameters and hyperparameters are tuned as per the proposed



**Fig. 16** The loss function of the feature optimized deep neural network

strategy. In future, the dropout can be started exploring by random values and positioning, as don't know where the assigned value and position of dropout is optima. The work also proves that the parameters and hyperparameters generalize the learning of the mode with performance up to state of art.

## References

1. Gautam KS, Thangavel SK (2019) Video analytics-based facial emotion recognition system for smart buildings. *Int J Comput Appl* 1–10
2. Gautam KS, Kumar TS (2016) Discrimination and detection of face and non-face using multi-layer feedforward perceptron. In: *Proceedings of the international conference on soft computing systems*, Springer, New Delhi, pp 89–103
3. Agarap AF (2018) Deep learning using rectified linear units (relu). arXiv preprint [arXiv:1803.08375](https://arxiv.org/abs/1803.08375)
4. Haykin S (2010) *Neural networks and learning machines*, 3/E. Pearson Education India
5. Jingwei T (2020) Genetic algorithm for feature selection. <https://www.mathworks.com/matlabcentral/fileexchange/71547-geneticalgorithm-for-feature-selection> MATLAB Central File Exchange. Retrieved Mar 23 2020

6. Chollet F (2015) keras
7. Theano Development Team. Theano: a python framework for fast computation of mathematical expressions



# Integrating ICT in Communicative Language Teaching in Bangladesh: Implementation and Challenges



Nafis Mahmud Khan and Khushboo Kuddus

**Abstract** The methods and approaches to language teaching and learning have always been changing with time from the Grammar Translation Method to the Communicative Language Teaching in order to be compatible with the globalized world. The governments across the world have realized the importance of implementation of ICT in education and especially in the Foreign Language Teaching and Learning. Researchers and policy makers also emphasize on the importance of integration of Information and Communication Technology in English Language Teaching. The status of ICT in education in Bangladesh is at the emerging stage where the adoption differs from the implementation and hence, it leads to a significant gap between the adoption and the actual implementation and outcome of ICT in education. Therefore, the present study aims to discuss the implementation of CLT in Bangladesh and to explore the impact of ICT in CLT. Further, it also attempts to highlight the challenges in the integration of CLT and in the incorporation of ICT in CLT.

**Keywords** Communicative language teaching (CLT) · Information and communication technology (ICT) · Language teaching and learning · Challenges · Initiatives

---

N. M. Khan · K. Kuddus (✉)  
School of Humanities (English), KIIT Deemed to be University, Bhubaneswar 751024, Odisha, India  
e-mail: [khushboo3133@gmail.com](mailto:khushboo3133@gmail.com)

N. M. Khan  
e-mail: [nafismahmud53@gmail.com](mailto:nafismahmud53@gmail.com)

N. M. Khan  
Division of Research, Daffodil International University, Dhaka, Bangladesh

## 1 Introduction

Information and Communication Technology (ICT) has changed the process of teaching and learning and has brought a revolution in the education sector. Now the methodology and procedures of teaching and learning are undergoing significant changes concerning curriculum and pedagogy. Instead of traditional teaching aids such as blackboard, chalk and duster, classroom teachers have started using modern tools like smart boards, multimedia projectors, speakers, etc. Classrooms have become more student-centric, unlike earlier times. Currently, ICT has become the most useful tool in promoting and providing the quality of education in the world. Innovation in the methodology of teaching through ICT can enhance students' creativity, experience, students' critical thinking capability, productivity, and make a great impact on students' future careers. Almost all the developing countries are trying to adopt ICT into their education systems. The ministries of Education across the world have already agreed to the point of integrating ICT in their education systems. Bangladesh has also made a vision of "Vision 21" to improve the quality of education and integrating ICT into its education system [1].

Initiatives of implementing ICT in education have been taken massively. Despite, introducing a number of initiatives and programs for the implementation of Communication Language Teaching (CLT) in education, the current status of CLT in Bangladesh is not satisfactory. The educators and the policymakers observe several inhibitions in the implementation of ICT in Education, particularly in Communicative Language Teaching. There have been several studies to identify the inhibiting factors in implementing ICT in education but going through all the empirical researches is quite a time-consuming for educators. Therefore, literature pertaining to ICT in education in Bangladesh and its implementation in CLT need literature reviews which would further analyse and integrate the results of the empirical research. In addition to this, this study would provide educators and policymakers with the information related to ICT in education in Bangladesh and challenges in implementing ICT in CLT in Bangladesh without going for an extensive study on it. Therefore, the purpose of the study is to inform educators and policymakers about the emergence of ICT in education in Bangladesh and the challenges of implementing ICT in education, especially in CLT.

## 2 Methodology

The rapid advancement of ICT and its implementation in education has distracted the educators and the policymakers from focusing on identifying the challenges of its implementation in education in general and in CLT in particular. Further, an extensive study of the emergence of ICT, its implementation in education and the challenges

in its implementation would be a time-consuming process for educators and policymakers. Therefore, the present study attempts to address the above-mentioned problems in the form of the following research questions:

What is the present condition of ICT in education in Bangladesh?

What is the status of implementation of CLT in Bangladesh?

What are the challenges in implementing ICT in Education in Bangladesh?

What are the challenges in integrating ICT in CLT?

What are the measures taken to integrate ICT in CLT and education as a whole in Bangladesh?

## ***2.1 Data Collection***

Data was collected from the empirical studies conducted on identifying the status of ICT in education and the challenges of ICT in education and ICT in CLT in Bangladesh, published in books, peer-reviewed journals and proceedings. The keywords used for data collection were; ICT, CLT, language teaching and learning, ICT in education and challenges, ICT in CLT and issues, teachers perception of ICT in education, students perception of ICT in education. The databases that were used for data collection include; Google Scholar, Research Gate, Educational Resources Information Center (ERIC), JSTOR, Conference Proceedings and Book Chapters.

## **3 Findings**

As a result of reviewing the studies, the following major themes related to; the emergence of ICT in education, emergence and implementation of ICT in CLT, challenges of CLT in education and challenges in implementing ICT in CLT emerged.

### ***3.1 Emergence of ICT in Bangladesh***

Bangladesh has an agenda to become a “Digital Bangladesh” [1] and is determined to ensure quality Education. In light of this, ICT division has announced its policy ‘National ICT Policy 2009’ and based on this policy ‘National Education Policy’ has been rephrased in 2010. It is giving much emphasis on the use of ICT in improving quality education. For this, UNESCO has been providing its maximum support to the Government of Bangladesh (GoB) to adopt ICT in Education. After realizing the potential of ICT in Education, different private organizations and NGOs are extending their hands in integrating ICT as an innovative approach to Education. BRAC (Bangladesh Rural Advancement Committee) and Grameen Bank particularly have taken some initiatives in using ICT for education such as In-service secondary

teachers' ICT training program, Gonokendros (Union Library), and Computer-Aided learning. In addition to this, National Education Policy (NEP) 2010, has been working on; the promotion of ICT enabled teaching and learning, professional development of teachers to enable them to use ICT in teaching, ICT literacy for students, introducing ICT enabled education-related services and, also ICT use in education administration.

### ***3.2 ICT in Education: Bangladesh***

The integration of ICT in Education has brought a significant change particularly, in English Language Teaching (ELT) as discussed in [2–4]. To stimulate the changes, the Bangladesh Government has taken several initiatives to implement the use of ICT in classroom teaching. Especially, the Government has already focused on ICT based ELT in Education rather than only ICT literacy. The National Education Policy (NEP) 2010 emphasizes the use of ICT to improve the quality of education. At the same time, NEP 2010 identifies some strategies for using ICT in all the level; primary, secondary and higher secondary.

## **4 CLT at a Glance**

Communicative language teaching was introduced in Europe in the 1970s to communicate with a large number of new migrants and workers from different countries. According to Sandra, in the 1970s, the educators and practitioners of ELT took initiatives in Europe and North America to develop students' communicative skills [5]. Hyme claims that CLT came to light in the 1970s and its entrance into the English language curriculum started in the 1980s [6]. Another theorist argued with his view regarding communicative activities and emphasizing the ability to use language for different purposes. Thus, Widdowson, provided a pedagogical influential analysis of communicative competence of language in which four dimensions are identified: (a) discourse competence (b) grammatical competence, (c) sociolinguistic competence and (d) strategic competence [7].

Now, CLT has been widely used as a method of language teaching throughout the world since 1990. CLT has gone through numerous changes in the last 50 years. According to Widdowson, the changes of ELT can be divided into three phases such as traditional method (1960s), classical communicative language teaching (1970s–1990s), and current communicative language teaching (1990s up-present) [7].

## ***4.1 Emergence and Implementation of CLT in Bangladesh***

Followed by the independence of Bangladesh in 1971, through the amendment to the constitution in 1972 Bangla became the 'official language', and national language both for communication and educational intuitions. Soon after the act of 'The Bengali language Introduction act 1987', all the spheres and all government offices started using Bangla as an official and formal language [8]. Consequently, Bangla got new status officially into society and school, college and other government intuitions. Gradually, English started losing its demand as a second language and finally, was placed as a foreign language.

Though inconsistency was the common feature in education particularly, in ELT field in Bangladesh, after realizing the value and effectiveness of communication in language teaching in 1996, the Ministry of Education (MoE) shifted to communicative language teaching approach, replacing the traditional Grammar Translation Method (GTM) based language teaching [9, 10]. Thus, since 2001, NCTB has been publishing textbooks with introducing the communicative approach in teaching and learning for every level of various streams of education [11]. During 1990–1995 OSSTTEB, a UK-based donor-funded for the teacher training project but eventually, they forced the British council to implement CLT in Bangladesh. However, it can be claimed that OSSTTED failed to modify as OSSTTED's teacher selection was very slow and finally, the project remained incomplete and ended abruptly after three years [12]. Later, in 1997–2008 another project named English Language Teaching Improvement Project (ELTIP) was taken jointly funded by the UK government and currently, a project named English in Action (EIA, 2010) is introduced funded by the UK Department for International Development (DFDL) that aims to boost economic development in Bangladesh through improving English Language Teaching [13].

So from the above discussion, it is clear that from the very beginning initiatives have been taken massively to introduce and to implement communicative language teaching successfully in the secondary and higher secondary level. Though the country has already experienced CLT more than a decade, the students of the secondary schools are still lagging to achieve desired proficiency in English [14] and the impact of the methodological change has been under research [15].

## ***4.2 Implementing CLT in Education***

According to Haider et al. Bangladesh is one of the biggest primary Second Language English learning country in the world [15]. But, during the British rule (1757–1947), English did not get much priority in primary education. After the liberation war of Bangladesh, in 1992 English was made a compulsory subject in the primary curriculum [15]. The current status of English in the national curriculum is at stake. Despite having inconsistency in education policy and planning particularly, in the field of ELT in Bangladesh, realizing the value and effectiveness of communication

in language teaching, in 1996 the Ministry of Education (MoE) shifted to communicative language teaching approach, replacing the traditional Grammar Translation Method (GTM) [10]. In 1995 National Curriculum and Text Book (NCTB) revised the Secondary English curriculum and shifted to communicative language teaching CLT [11]. Thus, since 2001 till now, the communicative approach has been implemented in teaching and learning for every level of various streams of education.

The higher secondary education is considered a major tier of Education in Bangladesh as it is the pathway to study abroad as well as working in different foreign multinational and national companies [10]. So, to be proficient in communication for different purposes such as higher education and to communicate with the native and non-native speakers, admittedly, learning English is an indispensable part for the learners of 16–17 years age group. After realizing the value and effectiveness of communication in language teaching in 1996, the Ministry of Education (MoE) offered English as a compulsory subject in the curriculum between the years 1–12 [10], and finally, shifted to communicative language teaching approach replacing the traditional Grammar Translation Method (GTM) based language teaching.

### ***4.3 Challenges in Implementing CLT in Bangladesh***

The CLT approach aims to develop learners' communicative competence so that learners can use the language in a meaningful communicative context. After realizing the importance of CLT, in 1996 the Ministry of Education (MoE) shifted to the communicative language teaching approach. It not only started spending a lot of money to implement and improve the CLT approach in Bangladesh but also aimed to find incongruity between the expectation and the result [16]. Researchers, show the problems in implementing CLT in the EFL context, especially Bangladesh. Islam et al. also find hardly any research which confirms the success of CLT in Implementation. So it can be claimed that CLT has experienced various challenges in its implementation in primary school, secondary school, and higher secondary level. The literature shows that challenges faced by Primary, Secondary and Higher Secondary level in the implementation of CLT are more or less the same.

#### **4.3.1 Challenges in Primary School**

Although Communicative language teaching has been practised in the primary schools of Bangladesh for many years, still a wide range of gap is observed between the policy of CLT and the execution. Haider et al. found the lack of learner-centered approach, the limited students' participation in the classroom learning, unavailability of classroom materials [10]. Though CLT supports participatory engagements of the students but insignificant participation is observed in the classroom settings. Lack of teachers' training facilities is another hindrance to implementing CLT in a classroom setting [17]. Most of the time, teachers start their teaching career without having any

previous formal training [18]. The absence of language proficiency and very poor knowledge about language teaching among the teachers are also responsible for the failure of CLT. Lack of sufficient teaching aids and classroom size also hinder CLT implementation [18, 19]. Teacher-student ratio, unavailability of the teachers, the overburden of the classes, lack of lab facilities for listening practice are the challenges in implementing CLT at the primary school level.

### 4.3.2 Challenges in Secondary Level

To develop communicative competence, in 2001 the secondary English course of Bangladesh was reformed [10]. Though CLT has already been introduced in secondary schools, the flavor and the fragrance of the Grammar translation method still exist. Students are forced to develop the reading and writing skills to do better results in the examination where listening and speaking are neglected. Rehman et al. also asserts that the GT method is a great constraint to implement CLT in classroom teaching [18]. On the other hand, many rural teachers of secondary schools in Bangladesh still believe and favor teaching students through the GT method instead of CLT [19]. There is an acute deficiency in curriculum and syllabus designing and testing systems. Trisha et al. reports that, though the present syllabus focuses on language learning through interaction, reality works reverse [20]. They also show the disharmony between the present syllabus and the testing system at the secondary school level. Moreover, Sultana et al. asserts the dissimilarities between the curriculum, syllabus designing and the validity of the testing system of the English language in a public examination at the secondary level [21].

Another hindrance to implementing CLT at the Secondary level is the lack of trained teachers [22]. In most cases, English teachers start their careers without any ELT qualifications. Most of them have a post-graduate degree in English literature which is not enough for language teaching. On the other hand, training facilities for rural secondary school English teachers are not available [9]. Therefore, they are not aware enough about the principles and techniques of CLT and as a result, they do not apply and practice it within the classroom settings. Sometimes suitable English teachers are not chosen for training rather fewer skills teachers get the opportunity [19].

The low proficiency skills and knowledge of the teaching of the teachers is a great hindrance in implementing CLT [16]. Most of the time teachers are not practising English in classroom teaching. Though technology is mostly recommended for communicative language teaching, Chaudhury et al. pointed out that, in the secondary schools both in rural and urban, students are deprived of using technology [23]. Only the blackboard is being used for language teaching. Ahmed also considered insufficient teaching aids as the source of difficulty in implementing CLT in Classroom settings [11]. Ansarey found a similar result in his study of rural secondary schools. Apart from this, he asserts, lack of motivation and confidence of the teacher of the rural secondary schools are the great problems in implementing CLT in secondary schools [19].

**Table 1** Challenges of implementation of CLT

Students	Teachers
<ul style="list-style-type: none"> <li>• Lack of cooperation</li> <li>• Priority on accuracy more than fluency</li> <li>• Only focus on grade</li> <li>• Different level of proficiency, sometimes very poor</li> <li>• Lack of motivation</li> <li>• Lack of confidence</li> <li>• No interest in classroom participation</li> <li>• Fear of making mistakes</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of quality and well trained teacher</li> <li>• Lack of subject-based training</li> <li>• Lack of clear understanding of CLT</li> <li>• Lack of motivation</li> <li>• Teacher-centered classroom</li> <li>• Teacher's workload</li> <li>• Using blackboard, chalk and duster.</li> <li>• Lack of professional degree</li> <li>• Lack of confidence</li> </ul>

### 4.3.3 Challenges of CLT in Higher Secondary School

The current status of CLT implementation in higher secondary level is still in the doldrums. The principle of CLT always focuses on student-centered learning which follows a process-oriented syllabus. However, instead of giving priority to learner-centered learning, teachers are giving priority to rote memory to secure a good grade in the exam. Islam pinpointed in his research that most of the time, teachers have to remain under the pressure of completing the syllabus within the given time and that leads them to turn from process-oriented syllabus to product-oriented syllabus which plays a reverse role in the implementation of CLT in the college level [22]. Abedin also noted the same problem as a factor in implementing CLT at the college level [24]. The very common challenges in implementing CLT at the college level which has been reported by the researchers are a lack of logistic support and inappropriate class size [25].

Challenges faced by the teachers, students, and other obstacles in the CLT integration at primary, secondary, and higher secondary level schools are mentioned in Tables 1, 2 and 3.

## 5 Challenges in Implementing ICT in CLT: Bangladesh Context

Although the Bangladesh Government is still committed to implement ICT in education, there are some challenges which are working as an impediment in its implementation in classroom settings. One of the major challenges faced by teachers is the lack of knowledge about ICT tools. CLT always prefers technology-based autonomous learning. So to make the English class successful, integrating ICT in language teaching means to teach through ICT materials like using a computer, whiteboard, audio devices, internet, television, etc. But unfortunately, in most cases,



**Table 2** Miscellaneous challenges related to CLT

<ul style="list-style-type: none"> <li>• Classroom management difficulties</li> <li>• Examination is not CLT based</li> <li>• Large classroom size</li> <li>• Lack of subject-based training</li> <li>• Lack of teachers' guide and instruction</li> <li>• Lack of CLT materials and textbook</li> <li>• Lack of pair work, group work practice in classroom sittings</li> </ul>	<ul style="list-style-type: none"> <li>• Exam-oriented assessment system</li> <li>• More achievement-oriented rather than performance</li> <li>• Little impact on development problem</li> <li>• Conflict within the management</li> <li>• Corruption in administration</li> <li>• Time constraints to complete the syllabus</li> </ul>
--	---

**Table 3** Challenges related to CLT itself

<ul style="list-style-type: none"> <li>• Discrepancies between the curriculum and exam</li> <li>• Gap between the principle of CLT and practices</li> <li>• Inconsistency in ELT practice</li> <li>• Cultural conflict between the nature of the student and nature of CLT</li> <li>• Lack of Administrative support to CLT</li> <li>• Priority on only reading and writing skills</li> <li>• Methodological problem</li> <li>• Priority on product-oriented syllabus rather process-oriented</li> </ul>
--

it is found that teachers are unaware of ICT devices and have very little knowledge about the usage of the ICT tools [26]. Lack of qualified subject-based teachers is also another cause of not using ICT devices properly in the classroom. In most cases in Bangladesh, most of the teachers do not require any degree in ELT except honors and masters. Teachers from different disciplines come to teach the English language [11]. To develop learners' language proficiency skills, it is obvious to have ICT equipment like computer, audio aids, multimedia and Language Labs in the academic institutions. But technological tools are scarce in the educational institutions and so they are not being used in the classroom [27]. Therefore, to use technology efficiently, the presence of hardware and software must be ensured. So, directly or indirectly, the lack of sufficient funds is also a great hindrance in implementing ICT in language teaching. Sometimes due to the lack of appropriate trainers, most of the training programs fail to achieve their objectives.

In addition, Bangladesh has a shortage of teacher and they are overburdened with their work. Apart from teaching, the teacher has to do many other administrative jobs also. As a result, they do not get enough time to design, develop and integrate ICT in the teaching-learning situation [28]. So lack of time is also a great constraint to integrate ICT in the Communicative Language Teaching. Another great challenge of integrating ICT in CLT classroom is teachers' attitude and belief towards ICT. Teachers with positive learning skills can learn the necessary skills in implementing it in the classroom [29, 30]. So, to use technology successfully in the classroom

depends on the attitude of the teachers towards ICT. Based on the extensive study, identified problems are categorized and depicted through Tables 4, 5 and 6.

Further, Fig. 1 explicitly explains the status of integration of ICT in education in Bangladesh. The figure clearly explains the fact that only 38% of secondary level schools use computer laboratories for teaching and learning. The given figure explains that the country struggles with the infrastructure facilitating the implementation of ICT.

**Table 4** Challenges related to the teachers and students in implementing ICT in CLT

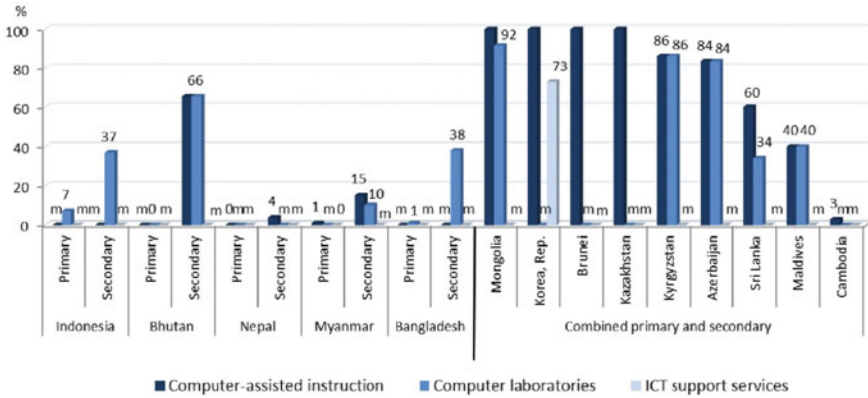
Students	Teachers
<ul style="list-style-type: none"> <li>• Lack of Proficiency</li> <li>• Lack of motivation</li> <li>• Technophobia</li> <li>• Lack of confidence</li> <li>• Poor attitude towards ICT based autonomous learning</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of ICT knowledge and training</li> <li>• Paucity of ELT teacher</li> <li>• Lack of motivation</li> <li>• Less confidence to use ICT materials</li> <li>• Technophobia</li> <li>• Teachers workload</li> <li>• Attitude towards ICT</li> <li>• Knowledge gap between teaching pedagogy and techno-oriented teaching pedagogy</li> </ul>

**Table 5** Miscellaneous challenges related to ICT

<ul style="list-style-type: none"> <li>• Classroom management difficulties</li> <li>• Lack of sufficient fund</li> <li>• Lack of clear vision</li> <li>• Poor socio-economic condition</li> <li>• Lower status of women</li> <li>• Lack of teachers' guide and instruction</li> <li>• Lack of ICT materials</li> <li>• Conflict within the management</li> <li>• Corruption in administration</li> <li>• Lack of time</li> </ul>
--

**Table 6** Similar challenges in the implementation of ICT in CLT: Bangladesh

<ul style="list-style-type: none"> <li>• Lack of ICT knowledge</li> <li>• Paucity of ELT teachers</li> <li>• Inadequate Classroom materials</li> <li>• Lack of sufficient fund</li> <li>• Lack of resource person</li> <li>• Teachers and learners motivation</li> <li>• Lack of proficiency of English in both students and teachers</li> <li>• Lack of clear vision</li> <li>• Poor socio-economic condition</li> <li>• Teachers' workload</li> <li>• Lack of time</li> <li>• Poor infrastructure development</li> <li>• Corruption in Administration</li> </ul>
--



**Fig. 1** Computer-assisted instruction, computer laboratories, and ICT support services by level of education, 2012. Notes m = missing data. Source UNESCO Institute for Statistics [31]

## 6 Discussion

From the above review, it can be easily stated that like CLT, ICT also faces some severe common problems from the very beginning of its implementation in Bangladesh in all academic streams: primary, secondary and higher secondary schools. The most common factors of integrating ICT in CLT are the lack of proper training facilities for the language teachers and the unavailability of teaching and learning aids in the institutions. Both factors demand huge funding which is difficult to manage by the government solely. Though various NGOs and other semi-government organizations, private sectors are contributing by organizing seminars, workshops, and training for the instructors and teachers but not proportionally adequate for such an overpopulated and economically packed county like Bangladesh. Therefore, the lack of funding is a great impediment in implementing ICT integrated Communicative language teaching in Bangladesh. Apart from this, overloaded classes and the pressure of other administrative works deter teachers to give enough time to prepare language class materials. The overload of work also decreases the motivation level of the teachers to work overtime to prepare ICT integrated CLT lesson plans. On the other hand, the policy and the principles of technology-based teaching pedagogy to facilitate learning and teaching are not much emphasized. Moreover, the lack of language-oriented teachers is another important factor that inhibits implementing ICT in language learning.

A sufficient amount of time is required to make a synergy between students and teachers through technology and to get real-time effects of the classroom. Literature invariably shows that the time allocated for each class is insufficient to integrate ICT in language teaching. So lack of time is another major common factor found in ICT based language teaching and learning. A few social factors like poor socio-economic condition in such a developing country like Bangladesh is also a great hindrance in ICT integrated CLT. Poor social status of women is one of the major challenges

as women are deprived of getting proper training facilities. Teachers' and learners' attitude to technology-oriented pedagogy is another factor as teachers still prefer to teach their students through traditional method rather than CLT. In addition to that, techno-phobia is also common among the teachers in using technology in language classrooms and is another important reason for not implementing ICT in language classrooms.

There are several barriers to the successful implementation of ICT in education as well as in English language teaching. However, Table 7, exhibits some of the important initiatives undertaken by the government and the non-government organizations of Bangladesh as a remarkable and progressive effort for the effective adaptation of ICT in education as a whole. These government and non-government organisations have implemented various innovative initiatives to address the challenges that inhibit the integration of ICT in education. The initiatives mainly focus on the following objectives:

- Improve the coordination and collaboration between ICT in education initiatives and the education sector.
- Provide technical training to teachers to help them integrate ICT in the teaching and learning process.
- Facilitate teachers with better pedagogical support for teachers to integrate ICT in class,
- Provide electronic resources to teachers and learners in the form of video-recorded lessons and ICT-mediated resources
- Monitor and investigate the incorporation of ICT in education.

## 7 Conclusion

This study has shown the current status of ICT in education in Bangladesh. It has also delineated the history of the CLT approach, the current status of CLT and the challenges in implementing CLT in Bangladesh. Further, it has exhibited the fact that the integration of ICT in CLT is full of challenges and the implementation observe challenges from both teachers' and students' end in the form of; lack of motivation, lack of ICT literacy, lack of sufficient training, lack of time, poor infra-structural development, a paucity of ELT teachers, techno-phobia and many more. It has also been seen that though teachers are one of the stakeholders of the education system, their role is more important than the others in the integration of technology in teaching and learning process. Further, the paper has also examined the challenges faced by the teachers and learners in Bangladesh in both ICT implementation and CLT. Therefore, their training needs proper planning and execution. Though many obstacles are working against the implementation of ICT in CLT in Bangladesh, its integration in English language teaching can play a significant role in keeping the promise of making 'Digital Bangladesh'. Hence, priority must be ensured equally

**Table 7** Initiatives undertaken by the Bangladesh Government and NGO's to reduce the challenges in implementing ICT in education and CLT

Initiatives adapted	Full form	Funding agencies	Functions
(a2i) programme	Access to Information	Funded by United Nation Development Programme (UNDP) and the United States Agencies for International Development (USAID) and run by the PMO	<p><b>Shikkhok Batayon:</b> It develops a huge amount of e-content with the help of teachers and teacher educators</p> <p><b>Muktupaath:</b> It provides several professional development courses for teachers</p> <p>It has produced talking books for students with disabilities, specifically for visually impaired children</p> <p>It also works on infrastructure and human resource development</p>
BANBEIS	Bangladesh Bureau of Educational Information and Statistics	Government of Bangladesh	<p>It produces training modules and conducts training programmes on ICT</p> <p>At the same time, BANBEIS is also working on the development of the first e-library in Bangladesh</p>
BOU	Bangladesh Open University		It is a national resource which produces text, audio and video programmes for the learners
EIA programme	English in Action programme	UK government	It focuses on the professional development of teachers using authentic audiovisual resources

(continued)

**Table 7** (continued)

Initiatives adapted	Full form	Funding agencies	Functions
ARBAN	Activity of Reformation of Basic Needs	It is a non-government development organisation	<b>ICT and English Club</b> (It is supported by the British Council and Dnet.) The members of the club learn English through ICT
BNNRC	Bangladesh NGOs Network for Radio and Communication	It is a national networking body	It provides training in basic English using community radio It produces episodes for radio broadcasting to engage both teachers and students It develops handbooks on English language learning for teachers It also publishes English language learning materials through local newspapers so that the common people can get access to English learning resources
Dnet: ICT-based Education Initiatives		Dnet is a non-profitable social enterprise	It provides training in ICT by effectively deploying multimedia content

Source [32, 33]

to the successful implementation of ICT integration in Communicative language teaching.

## References

1. Khan M, Hasan M, Clement CK (2012) Barriers to the introduction of ICT into education in developing countries: the example of Bangladesh. *Online Submission* 5(2):61–80
2. Dash A, Kuddus K (2020) Leveraging the benefits of ICT usage in teaching of English language and literature. In: Satapathy S, Bhateja V, Mohanty J, Udgate S (eds) *Smart intelligent computing and applications*, vol 160. *Smart innovation, systems and technologies*. Springer, Singapore, pp 225–232

3. Chatterjee B, Kuddus K (2014) Second language acquisition through technology: a need for underdeveloped regions like Jharkhand. *Res Scholar Int Referred e-J Literacy Explor* 2(2):252–259
4. Kuddus K (2013) Role of mass media and technology in fostering English language acquisition. In: *Proceedings of the UGC national seminar on English language acquisition development and environment*, pp 107–116
5. Savignon SJ (1987) Communicative language teaching. *Theor Pract* 26(4):235–242
6. Hymes DH (1972) On communicative competence. In: Pride JB, Holmes J (eds) *Sociolinguistics selected readings*. Harmondsworth Penguin
7. Widdowson HG (1987) *Teaching language as communication*. Oxford University Press
8. Banu R, Sussex R (2001) English in Bangladesh after independence: dynamics of policy and practice. In: Moore B (ed) *Who's centric now? The present state of post-colonial Englishes*. Victoria, Oxford University Press, South Melbourne, pp 122–147
9. Rahman MM, Pandian A (2018) A critical investigation of English language teaching in Bangladesh: unfulfilled expectations after two decades of communicative language teaching. *English Today* 34(3):43–49
10. Rahman M (2015) Implementing CLT at higher secondary level in Bangladesh: a review of change management. *J Edu Pract* 6(2):93–102
11. Ahmed S (2012) English language teaching at secondary school level in Bangladesh: an overview of the implementation of communicative language teaching method. *J Engl Lang Teach* 2(3):16–27
12. Obaidul Hamid M (2010) Globalisation, English for everyone and English teacher capacity: language policy discourses and realities in Bangladesh. *Current Issues Lang Plann* 11(4):289–310
13. Seargeant P, Erling E (2011) The discourse of 'English as a language for international development': Policy assumptions and practical challenges. In: Coleman H (ed) *Dreams and realities: developing countries and the English language*. British Council, London, pp 248–267
14. Afroze R, Kabir MM, Rahman A (2008) English teachers' classroom practices in rural secondary schools: an exploration of the effect of BRAC training. *Bangladesh Edu J* 7(1):07–16
15. Haider Md, Chowdhury TA (2012) Repositioning of CLT from curriculum to classroom: a review of the English language instructions at Bangladeshi secondary schools. *Int J English Linguis* 2(4):1–12
16. Maksud A, Walker AL (2014) 'Bogged down' ELT in Bangladesh: problems and policy: investigating some problems that encumber ELT in an EFL context. *English Today* 30(2):33–38
17. Rahman MS, Karim SMS (2015) Problems of CLT in Bangladesh: ways to improve. *Int J Edu Learn Dev* 3(3):75–87
18. Ansarey D (2012) Communicative language teaching in EFL contexts: teachers attitude and perception in Bangladesh. *ASA Univ Rev* 6(1):1–18
19. Alam MM (2018) Challenges in implementing CLT at secondary schools in rural Bangladesh. *IIUC Stud* 13:93–102
20. Trisha BJ, Akter M, Yusufi IN, Munna AS (2017) Dissonance between syllabus and testing: reason of weak efficiency in English at SSC level. *Curr Edu Res* 1(1):1–5
21. Sultana N (2018) Test review of the English public examination at the secondary level in Bangladesh. *Lang Test Asia* 8(1):1–16
22. Islam SMA (2015) Language policy and practice in secondary school contexts in Bangladesh: challenges to the implementation of language-in-education policy. Aalborg Universitetsforlag. Ph.d.-serien for Det Humanistiske Fakultet, Aalborg. Universitet. <https://doi.org/10.5278/vbn.phd.hum.00006>
23. Chowdhury R, Kabir AH (2014) Language wars: English education policy and practice in Bangladesh. *Multilingual Educ* 4(1):1–21
24. Abedin M (2012) The present mode of teaching in the ELT classes at the higher secondary level in Bangladesh: is it the practice of CLT or disguised GTM? *Stamford J English* 7:1–15
25. Ghasemi B, Hashemi M (2011) ICT: Newwave in English language learning/teaching. *Procedia Soc Behav Sci* 15:3098–3102

26. Shahnaj P (2013) Integrations of ICT in education sector for the advancement of the developing country: some challenges and recommendations—Bangladesh perspective. *Int J Comput Sci Inform Technol* 5(4):81–92
27. Afshari M, Bakar KA, Luan WS, Samah BA, Fooi FS (2009) Factors affecting teachers' use of information and communication technology. *Int J Instruct* 2(1):77–104
28. Ihmeideh FM (2009) Barriers to the use of technology in Jordanian pre-school settings. *Technol Pedagogy Educ* 18(3):325–341
29. Kuddus K (2018) Emerging technologies and the evolving roles of language teachers: an overview. *Lang India* 18(6):81–86
30. Kuddus K (2013) Changing roles of teachers with ICT in teaching and learning of English as second and foreign language. *ELT Weekly*. <http://eltweekly.com/2013/05/changing-roles-of-teachers-with-ict-in-teaching-and-learning-of-english-as-second-and-foreign-language-by-khushboo-kuddus/>
31. Information and communication technology (ICT) in education in Asia a comparative analysis of ICT integration and e-readiness in schools across Asia. [http://uis.unesco.org/sites/default/files/documents/information-communication-technologies-education-asia-ict-integration-e-readiness-schools-2014-en\\_0.pdf](http://uis.unesco.org/sites/default/files/documents/information-communication-technologies-education-asia-ict-integration-e-readiness-schools-2014-en_0.pdf)
32. Towards a national policy on open educational resources in Bangladesh. [http://oasis.col.org/bitstream/handle/11599/2740/2017\\_COL\\_Towards-National-Policy-OER-Bangladesh.pdf?sequence=1&isAllowed=y](http://oasis.col.org/bitstream/handle/11599/2740/2017_COL_Towards-National-Policy-OER-Bangladesh.pdf?sequence=1&isAllowed=y)
33. Innovative strategies for accelerated human resource development in South Asia information and communication technology for education special focus on Bangladesh, Nepal, and Sri Lanka. <https://www.adb.org/publications/innovative-strategies-ict-education-bangladesh-nepal-sri-lanka>



# Emerging Role of Intelligent Techniques for Effective Detection and Prediction of Mental Disorders



Priti Rai Jain and S. M. K. Quadri

**Abstract** It has been established and accepted that mental disorders pose one of the most prominent health challenges worldwide. Information retrieval from mental health data which may be explicit in electronic health records and clinical notes or may be implicit in social media postings has vast potential to detect, distinguish and diagnose the status of mental health of individuals and aid in managing this problem. This paper summarizes some recent studies that apply the state of art Artificial Intelligence (AI) techniques to mental health data. The paper summarizes that newly emerging AI technologies hold a decent promise and can be leveraged to predict, assist in the diagnosis and management of mental disorders. The role of AI in this area becomes particularly important in a scenario where there is a worldwide dearth of qualified professionals who can deal with mental health disorders, where the cost of these services is high and the people suffering from these problems often refrain from availing these services due to social stigma associated with it.

**Keywords** Artificial intelligence · Deep learning · Mental health · Depression · Social media · Twitter · EHR · NLP · CNN · DNN

## 1 Introduction

India, China, and USA are the countries by having the maximum number of people affected by mental ailments like anxiety, bipolar disorder, and schizophrenia, etc. A study by the World Health Organization (WHO) for the NCMH (National Care of Medical Health), claims that 6.5% Indians suffer from some Serious Mental Illness (SMI) or another in rural as well as in urban areas alike. Currently, over 300 million

---

P. R. Jain (✉) · S. M. K. Quadri  
Jamia Millia Islamia, New Delhi, India  
e-mail: [prtirai.jain@mirandahouse.ac.in](mailto:prtirai.jain@mirandahouse.ac.in)

S. M. K. Quadri  
e-mail: [quadrismk@jmi.ac.in](mailto:quadrismk@jmi.ac.in)

people across the world suffer from one particular mental disorder namely ‘depression’ [1]. As per WHO, India has the maximum number of individuals suffering from a depressive disorder in the world, followed by China and the USA. Depression is a psychiatric ailment whose symptoms include low mood and aversion to activity. This ultimately affects the day-to-day functioning of a person in both professional and personal spheres of life [2]. “India’s National Mental Health Survey, 2016 found that close to 14% of India’s population requires active mental health intervention. Although there exist efficient measures and therapies, there is a large scarcity of experts in such as psychologists, psychiatrists, physicians, and nurses in the area of mental health. India has 0.75 Psychiatrists per 100,000 people, while there are about 6 psychiatrists per 100,000 people in countries with higher per capita income [3]. AI tools have the potential to find undiagnosed mental health issues and speed up the required treatment. Thus, AI solutions are coming at the most opportune time. AI can be used as a powerful diagnostic as well as a therapeutic tool for mental illness [4]. Even though mental disorders are among the most enervating diseases, it is a common notion that a mental ailment can only happen to a person who is mentally weak or a person who has a lot of time and money. Seeking support from mental health professionals is interpreted as a sign of weaknesses [5].

Several recent studies focus on enhancing the quality of medical services by information retrieval from biomedical and healthcare data. This data is generated from clinical reports, clinical decision support systems, doctor’s notes, laboratory information systems, readings from monitoring devices, pathological reports, radiological images—X-rays, MRI, fMRI, CT Scans, PET Scans, electronic health records (EHR), data from hospital billing records, medical claims, wearable body sensors, genomics, genetics, etc. [6]. Online social media platforms such as Instagram, Facebook, WhatsApp, Twitter, WeChat, etc., have also become important sources of information since people can quickly post information on such sites. The ubiquity of smartphones and the Internet has exponentially increased information sharing on these sites. The posts on such sites in the form of text, images, audios, and videos are a valuable source of health information and can be used for better management in mental healthcare [7].

Healthcare datasets are large and quite complex to deal with using the currently available tools. Upcoming AI techniques can be used to gain insight and to solve present health issues; information about which may be latent in the gigantic data which is diverse and complex. Analyzing this data can help in predicting undiagnosed patients, grouping patients having similar health issues, developing accurate health profiles of individuals, predicting personalized preventive therapies, obtaining frequent patterns from healthcare databases to find relationships between health-disease-drugs, etc. It may also aid in making efficient healthcare policies, constructing better drug recommendation systems and predicting health insurance frauds, etc. Beyond doubt the relevance of results so extracted would depend on the effectiveness of methods used to organize, represent, process and interpret this data [7].

Healthcare workflows involve several complex steps and sub-steps. The very first being data acquisition or data collection since it involves privacy and anonymizing.

Other major challenges are data transformation and pre-processing needs, specific considerations like maintaining data quality that is necessary to ensure accuracy and reliability of results [8]. Subsequent steps like data modeling, information retrieval and visualization also have their specific considerations and complexities. Existing solutions in this area have a close coupling to the specific problem that they are designed to solve and have domain-specific adaptations. Thus, these solutions are not generic. The primary reason behind such dependencies is the nature of healthcare data.

“Predictive mental health analysis means moving from the patient description (hindsight) and investigation of statistical group differences (insight) to models capable of predicting current or future characteristics (foresight) for individual patients” [9]. Predictive data analytics in mental health and psychiatry promises to transform clinical practice in psychiatry. Personalized treatment optimization could maximize adherence and minimize undesired side effects. The three main application areas of predictive analytics identified by [9] are: (i) improving the prevalent psychiatric ‘trial-and-error’ strategy; (ii) differential diagnoses that are an essential aid whenever there is an ambiguous clinical scenario; (iii) models predicting individual risks that can significantly improve patient management, e.g., in case of prodrome detection of schizophrenia (Prodrome refers to early symptoms and signs of illness preceding the disease’s characteristic manifestations. Prodrome detection is important as it indicates the onset of disease much before diagnostically specific symptoms take shape.). Thus, preventive measures can be effectively targeted in cases where the mental disorder is still in its initial stages or the ailment is yet to begin setting in.

Though computerization of mental healthcare services is often quoted as impersonal, nonetheless the innate anonymity of AI applications turns out to be a positive thing in this respect. It helps suppress the pervasive stigma hindering people receiving such treatment. Mental health patients often feel awkward to divulge their problems to therapists but find themselves relatively at ease with impersonal and faceless AI tools. The lower cost of AI treatments vis-à-vis the cost of seeing a psychiatrist or psychologist is an added advantage, especially, in the scenario having a shortage of such professionals. Effective models in the area of mental health could, therefore, minimize patient suffering.

## 2 Objective

The objective of this work is to review the work done in the area of detecting, diagnosis, prediction and prevention of mental disorders. The work explores the state of art in AI in predicting a mental illness by way of information retrieval of mental health data from varied sources with special attention towards data being generated via social media (SM) platforms. The work is an effort to explore the existing tools and techniques being used by researchers, the effectiveness of these tools and techniques and their basic challenges. It is an effort to identify how this data has been dealt with

using machine learning techniques, ensemble learning methods and deep learning techniques and thus identify underlying opportunities and challenges.

### 3 Current State of Work

Abundantly increasing volume of data related to healthcare in general and mental health is attracting researchers globally. Volumes and variety of user-generated data on SM platforms are especially enticing researchers to find and interpret and use information hidden in the same for the benefit of mankind. AI techniques are being employed to estimate how this data could be used to perceive or predict mental health issues of individuals posting it. Studies show that there are clear distinctions in posting patterns of users suffering from mental illnesses and those without it. Using AI in this manner, to identify people who may be suffering from mental ailments by way of content posted by them on public forums, can supplement the traditional diagnostic methods which are based only on recollection data. The scheme has an advantage as it offers deeper insights into user's past activities, behaviors, and feelings [1].

#### 3.1 Case Studies Based on Clinical NLP

Owing to the sensitive nature and privacy-related issues of narratives of mental health patients makes them difficult to access for research purposes. Natural Language Processing (NLP) was therefore seldom applied to mental health notes till late. The 'CEGS N-GRID 2016 Shared Task in Clinical Natural Language Processing' made available the first set of neuropsychiatric records for research purposes.

The study [10] explores the effectiveness and feasibility of predicting common mental ailments based on brief text descriptions of the patient's illness history dataset from CEGS N-GRID 2016. Such a history usually occurs before the psychiatrist's initial evaluation note. The study proposes two Deep Neural Network (DNN) models—the first model is based on 'Convolutional Neural Networks (CNN)' while the second model is based on 'Recurrent Neural Networks with Hierarchical Attention (ReHAN)'. Experiments were carried out to compare these models with each other and with other models. The outcomes showed that CNN based model with the optimized threshold of probability achieved the best results. Both the models demonstrated statistically significant improvement over other models based on support vector machines (SVM) and named entities recognition (NER).

The study [11] investigates how subjective narratives of a patient's experience is of utmost importance to support the phenotypic classification in providing him/her personalized prevention, diagnosis, and intervention for a mental ailment. The researchers performed experiments using the 'Psychiatric notes from the CEGS N-GRID 2016 Challenge'. The goal was to explore methods for identifying psychiatric symptoms in the clinical text using unlabeled data. Seed lists of symptoms were

used to augment the unsupervised distributional representations. Paragraph2vec [12] (an extension of word2vec [13]) was used to produce distributional representations. Then, the semantic similarity between candidate symptoms and the distributional representations was determined to evaluate the relevance of each phrase. Interesting results shown by the experiment include: (a) distributional representations built from SM data perform better than those built using clinical data; (b) distributional representation model that was built from sentences resulted in better representations of phrases when compared to the model built with phrases alone.

The study [14] admits that even though gathering information about psychiatric symptoms is important for time appropriate diagnosis of mental ailments and personalized interventions but diversity, as well as the sparsity of these symptoms, makes it difficult for existing Natural Language Processing (NLP) techniques to automate the extraction of these symptoms from clinical texts. To explore a solution to this problem, the study uses domain adaptation (DA) methods. It used four unlabeled datasets as source domain: (i) MIMIC III which is a publicly available collection of clinical notes in ICU that is often used to generate word embeddings for clinical NLP; (ii) MEDLINE data of 2013—a corpus of article abstracts that relate to clinical issues such as diseases, medication, and treatments; (iii) Psychiatric Forum—a collection of posts on WebMD having expressions like those of patients, as recorded in psychiatric notes and (iv) Wikipedia—a well-known online crowdsourced encyclopedia having extensive coverage of topics on multiple domains. The target domain dataset was the unlabeled corpus of psychiatric notes from CEGS N-GRID 2016. The experiment applied four different strategies on these datasets for identifying symptoms in psychiatric notes: (a) only using word embeddings from a source domain, (b) directly combining data of source and target domain to produce word embeddings, (c) allocating variable weights to word embeddings, and (d) re-training the word embedding model of the source domain using a corpus of the target domain. It concluded that superior results were obtained in cases where weights were assigned to word embeddings and where word embeddings of source domain were re-trained using words from target domain when compared with the results obtained by the other two approaches (i.e. the approach using word embeddings from source domain only and the approach that combines data from the source and target domains to directly generate word embeddings). The study concludes that DL-based methods and Domain Adaptation (DA) strategies have promising potential to resolve the challenges of clinical NLP tasks.

### ***3.2 Case Studies Based on EHRs***

Electronic mental health records are a detailed and comprehensive repository of phenotypic features in SMIs. The study [15] used the pseudonymized and de-identified data from South London and Maudsley NHS Foundation Trust (SLaM)'s EHR system to represent relationships between the medical vocabulary of physical symptoms for SMI and descriptions used by clinicians. It used semantic modeling

and clustering techniques on this huge corpus of EHR data to explore if these models could be used for determining new-fangled terminology appropriate for phenotyping SMIs by using minimal prior knowledge. The experiment used a 2-stage method to derive and curate 20,403 terms extracted from EHRs using ‘English Punkt tokeniser’ from Natural Language Toolkit (NLTK) 3.0. Thereafter, a Gensim implementation of a continuous bag of words (CBOW) was used to identify the semantic resemblances of 1-word entities in bio-medical works and clinical texts with the help of k-means ++ using Scikit-Learn framework for clustering. The list of 20,403 terms was condensed to 557 significant concepts by eliminating duplicate information. These were then organized into nine categories according to different aspects of psychiatric assessment. A subset of 235 expressions from 557 was found to be expressions having clinical importance; 53 expressions out of 235 were identified as new synonyms with the existing ‘SNOMED CT (Systematized Nomenclature of Medicine—Clinical Terms)’.

The study [16] focuses on DA of sentiment analysis to mental health records by assessing multiple sentence-level sentiments using three classification models. The baseline model was a majority vote approach using the Pattern sentiment lexicon employed by [17, 18]. The second model used a fully supervised multi-layer perceptron (MLP) and the third model used semi-supervised MLP architectures. The training and test data were vectorized at the sentence level using the pre-trained Universal Sentence Encoder (USE) embedding module [19] that was designed specifically for transfer learning (TL) tasks. The experiment tuned the hyperparameters using grid search with fivefold cross-validation. It used two semi-supervised learning configurations: (i) Self-Training and (ii) K-Nearest Neighbors (KNN). The results of this study reveal that run-of-the-mill sentiment analysis tools cannot identify clinically positive or negative split.

### ***3.3 Case Studies Based on Images—Facial Appearance and Expressions***

The study [2] automates depression diagnosis using encoded facial appearance and dynamics. It uses a DNN to predict the Beck Depression Inventory (BDI)-II values from videos for this purpose. The framework uses two-streams, one aimed at capturing the facial appearance and the other aimed at capturing facial dynamics. Experiments were conducted on AVEC2013 and AVEC2014 depression databases. Deep convolutional neural networks (DCNN) and Dynamics-DCNN were used for simultaneous tuning of layers that could tacitly integrate the appearance and dynamic information. The model was used to predict depression score which was then used to diagnose depression.

The study [20] involves scrutiny of electroencephalogram (EEG) signals and continuous emotion detection from facial expressions. Emotions may be defined as a “stochastic affective phenomenon” that is produced by stimuli. It presumes videos

can bring out emotions in their audiences. A study is an approach to instantaneously detect these emotions elicited by viewers using their facial expressions and EEG signals. The facial expressions and physiological responses of the participants were noted while they were viewing a set of emotion-inducing videos. Five annotators interpreted and explained the expressions of participants' faces in the videos as negative to positive emotions valence. Likewise, stimuli video clippings were annotated on valence and arousal. "Continuous Conditional Random Fields (CCRF)" and "Long-Short-Term-Memory Recurrent Neural Networks (LSTM-RNN)" were used to detect emotions. The results conclude that facial expressions were more important in detecting emotions than EEG signals and that most of the emotionally important content in EEG features were due to the activity of face muscles [20].

### ***3.4 Case Studies Based on Information Retrieval from Social Media Posts***

Due to the persistent availability of SM and microblogs, millions of people express their thoughts, emotions and mood online. Some of these users willingly share online their struggles with mental ailments. This makes SM a valuable resource for research in mental health.

The experiment [21] studied whether day to day language contains enough indicators for predicting the probability of occurrence of future mental illness or not. Data posted by way of text messages was collated from a popular SM site Reddit. Posted messages were retrieved from the discussion groups that were related to different kinds of mental illnesses (clinical subreddits). Posts were also fetched from discussion groups related to non-mental health issues (non-clinical subreddits). It was found that words taken from an individual's post on clinical subreddits could be used to differentiate various common mental disorders (such as attention deficit hyperactivity disorder (ADHD), bipolar disorder, anxiety and depression). The study also found that words picked from posts on nonclinical subreddits such as cooking, travel, cars, etc. could also be used to differentiate types of mental disorders that individuals who posted them may be suffering from. An important implication of this was that the effect of mental illness also influences topics that may be unrelated to mental disorders. More enticingly the study found that words from nonclinical subreddits were also able to predict what could be posted to clinical subreddits in future. It thus has an important implication—that everyday language has enough indications vis-a-vis the likelihood of a person developing a mental illness in future, perhaps even before the person is aware of his/her mental health condition. The study uses clustering via DBSCAN to demonstrate that there exist themes in textual descriptions of peoples' postings on mental health issues. Another cluster analysis for depression showed that individuals subscribing to subreddit r/Depression were more inclined to talk about sadness, harm, ugliness, fantasies, life problems and medicines used in the treatment of treat depression.

The study [22] proposes AI-based models to predict the development of depression and Post Traumatic Stress Disorder (PTSD) using Twitter data and depression history for 204 users. Prognostic features determining the affect, linguistic style and context were extracted from tweets. Several supervised learning classifiers were built out of which a Random Forests classifier having 1200-trees demonstrated the best results. Stratified cross-validation methods were used for optimizing random forest's hyperparameters. State-space models, with observable data, were used by the study to estimate the status of hidden variable viz-a-viz time. A 2-state 'Hidden Markov Model (HMM)' was built to find differential changes between affected and healthy instances over time. Word shift graphs were used to compare the language adjusted happiness scores of tweets by affected and healthy users. Resulting models could successfully distinguish depressed and healthy users. A significant result of state-space time series analysis suggested that the onset of depression is detectable much before (several months) clinical diagnosis. Predictive results were replicated with a separate set of users diagnosed with PTSD. The state-space temporal model, in this case, revealed pointers to PTSD almost immediately after trauma, which could mean many months before the actual clinical diagnosis. These results suggest that a data-driven, predictive approach for early screening and detection of mental disorders holds great promise.

The study [23] used ML to infer the mental well-being of students on campus using their language in online communities in Reddit. The inferred expressions were analyzed based on their linguistic and temporal characteristics for campuses of about 100 universities. Results were then used to develop a campus-specific Mental Well-being Index (MWI) that relates to the collective mental well-being of students on campus. The study builds and evaluates a transfer learning (TL) based classification method that is claimed to detect mental health expressions with 97% accuracy.

The study [24] is aimed at detecting depression using CNNs and Recurrent Neural Networks (RNNs) and NLP. It builds word embeddings to learn feature representations for health-specific tasks. It uses embeddings average to find improved feature representation. The study proposes four models for depression detection. Three of these models use CNN (CNN WithMax, Multi-Channel CNN, Multi-Channel Pooling CNN) and one model uses RNN-bidirectional LSTM with context-aware attention. Experiments were done using two publicly available datasets—CLPsych2015 (having self-reported posts by Twitter users about PTSD and depression) and Bell Let's Talk. Results show that CNN models perform better than the RNN model.

The study [25] uses CLPsych2015 dataset to investigate whether learned user representations could be correlated with mental health status. The experiment induced embeddings for users known to be suffering from depression and PTSD with embeddings for a group of demographically matched control users. The induced user representations were evaluated for: (i) their ability to capture homophilic relations i.e. the tendency for people to maintain associations with people who are like themselves to mental health conditions; and (ii) their predictive performance in proposed mental health models. The results of the study affirm that learned user embeddings can capture relevant signals for mental health quantification.



The study [26] extracts a new large dataset ‘SMHD (Self-reported Mental Health Diagnoses)’ from publicly available Reddit corpus2 posts of users with one or more mental ailments along with matched control users. The experiment measures linguistic and psychological variables and uses text classification to provide insight into the mental health conditions of users by spotting the language used in their posts. The approach extends Reddit Self-reported Depression Diagnosis (RSDD) dataset [27] by integrating synonyms in matching patterns. The work captures self-reported diagnoses of nine different mental disorders. SMHD identifies users who may suffer from multiple disorders, thus allowing a language study of interacting mental conditions. The study trains both binary as well as multi-label multi-class classifiers using—logistic regression, XGBoost, an ensemble of decision trees, support vector machine (SVM) with tf-idf bag-of-words, shallow neural net model using supervised FastText, and CNN. It was observed that FastText performed best in terms of F1 score across all conditions except in case of bipolar disorder where SVM showed better results, and in case of eating disorder where CNN showed better results.

The study [28] uses images to disclose predictive markers of depression. Instagram photos (43,950) of 166 participants were used to extract statistical features using metadata components, color analysis, and algorithmic face detection. A suite of supervised ML algorithms was used to find the predictive capacity of models. It was found that the 100-tree Random Forests classifier performed the best. The proposed model performs better than the average success rate of a general practitioner for identifying depression. These results were valid even when the analysis was confined to messages posted on the SM platform before the depressed individuals were first diagnosed.

The study by [29] combines textual, visual, and connectivity clues on Twitter to identify the state of mental health. The experiment uses a multi-modal dataset introduced in [30], about users who had self-reported depression. It finds the depressive behavior using a lexicon of depression symptoms containing 1500 depression revealing terms created in [30]. From this, a subset of 8770 users (24 million time-stamped tweets) having 4789 control users (those who do not express depressive behavior) and 3981 depressed users were verified by two human judges. It does a detailed analysis of the visual and textual content of susceptible individuals by using their facial presentation, emotions from facial expressions, and demographic features from profile pictures. When analyzing text, the study considers an individual’s natural way of analyzing and organizing complex events that may have a strong association with the user’s analytical thinking, his/her authenticity as a measure of the degree of honesty, clout, swear words, self-references, sexual lexicon terms, etc. The study then developed a multi-modal framework by combining heterogeneous sets of features by processing textual, visual, and user interaction data. It conducted an extensive set of experiments and employed statistical techniques for comparative assessment of the predictive power of the proposed multimodal framework.

The study [31] predicts the tendency of depression by analyzing his/her postings on Instagram. It used a DL classifier to combine and evaluate three features—text, images, and behavior, to predict depression tendency. For collecting data, the

researchers crawled Instagram postings with hashtags depression, suicide, depressed. From this data pool, users who self-reported i.e. explicitly stated that they suffer from mental illness were tagged as depressed users. For finding non-depressed users, postings with hashtags relating to the word happy were crawled. Word2vec [13] was used for representing text features. Useful features for identifying depressed individuals and non-depressed individuals from images were extracted by the pre-trained VGG16 [32] model on the ImageNet dataset [33] and then fine-tuning the obtained model on the crawled dataset. The ImageNet [33] is a huge ontology of images organized according to the millions of annotated images organized as per the semantic hierarchy proposed in WordNet [34]. The work proposes that the behavior of SM users can be categorized as two features (i) social behavior and (ii) writing behavior. The social behavior includes—the time of post, frequency of posts, number of ‘likes’ that the post gets, etc. whereas the writing behavior of a post comprises of how a user writes posts for e.g. people who are depressed are inclined to use absolutist words like “always,” “complete,” and “definitely” etc. [35]. The proposed model predicts the tendency of depression of these users with an F-1 score of 82.3% using a five layered CNN. The study claims that the combination of images and text make the proposed model more rigorous than the models using only text or images.

## 4 Challenges

The challenges of mental health data are a superset of the challenges of healthcare data in general.

The medical data is very high dimensional, heterogeneous, unstructured, and multimodal [8, 36–39]. Healthcare research demands precision in the classification of diseases and their subtypes. Methods that can integrate heterogeneous data such as images, molecular data, electronic health records (EHRs), etc. with precision are an impending challenge [39]. Due to the variety in biological measurements and the imprecision in measurement by the equipment used, healthcare data may not consist of impeccably accurate measurements. It rather has estimations and that too with noise.

Lack of correctly labeled training data in adequate volumes has affected the advancement of nearly all DL applications. Moreover, for accurate results, it is crucial to look at the uncertainty in measurements that may capture noise and also keep in mind how noise in input values propagates through DNN or other models [39]. There exists a lack of appropriate documentation in locally developed tools that store this data and so there is a lot of missing information and inaccurate entries [40]. It is currently difficult to train models for rare diseases using DL due to the problem of overfitting.

Lack of standards for format definition and naming conventions makes this data prone to potential human errors when uploading it to databases. Privacy of data and legal considerations thereof, make it difficult to access data for research purposes [41].

Deep Learning models are often not interpretable [42]. A model that has achieved very high performance is sure to have recognized some important patterns in data. Professionals in the field would certainly want to know those patterns. Understanding these is not possible if the deep learning model is not interpretable and remains a black box. Moreover, when a DNN model diagnoses some medical conditions, it is important to make sure that the DNN is taking that decision based on reliable reasons because DL models are prone to adversaries. Hence, it is extremely important to understand the basis of the model's output. This is critical because it has been observed that at times these models may output confidence scores of the order of 99.99% for samples that are nothing but pure noise [39].

## 5 Opportunities

DL has extensive potential for approximating and reducing the large and complex healthcare datasets to precisely transformed predictive output. Applying AI to mental health data holds a big promise for personalizing treatment selection, foretelling, monitoring to avoid/manage relapse, detecting and preventing mental ailments before they develop into symptoms that can be clinically detected [38]. There is a need for improving clinical operations. It involves learning from huge amounts of historical data that may be unstructured in nature. There exists a need to devise more effective access to unstructured information by improving searches [6]. There is a need to retrieve related data from different resources and consolidate it into a single application that can provide auto-generated summaries to support clinical decisions. Although NLP tools have been used to find important indicators of risks from EHRs, there are not many studies that apply sentiment analysis tools to clinical data [16].

DL architectures can be applied to almost all types of data—numerical, text, audio, video, and their combinations. Thus, DL can be used for exploratory investigation and hypothesis-driven research in medicine and healthcare [8].

Apart from issues of computer security, data and information security, etc. data privacy is a big challenge. While dealing with healthcare data research team needs to ensure that the individual's privacy preferences and their personally identifiable information are protected [43].

It is increasingly important to make DL models more interpretable as they reach the state of art in performance [42]. Interpretation is vital for building trust and confidence. Currently, most DL approaches are a black box. Hyperparameters values that control the DNN architecture such as size and number of filters in a CNN, its depth, etc., are still guesswork. Trustworthy AI is the need of the hour.

## 6 Conclusion

AI-based techniques have great potential to uncover a multitude of factors that are responsible for relationships between mental health conditions, diseases, and drugs. They can conceivably solve many problems in this area, which may not otherwise be solved via traditional studies based on randomized trials alone. Analysis of the aforesaid data can help healthcare researchers in improving services to a large extent.

The anonymity of AI is a boon in a society having a pervasive stigma in the way of mental health patients and its treatment. Patients quite often feel embarrassed to reveal problems to a therapist and find themselves more at ease with AI-powered tools. The need for computational techniques in mental health is further armored by the fact that there is a shortage of doctors and professionals while the mental stress and problems are rapidly increasing.

DL models incite the interest of researchers in this domain for two reasons: (i) they require less manual feature engineering and have shown better performance than many traditional ML methods; (ii) large and complex datasets present in mental health enable training of complex DL models. Indisputably, the relevance of information extracted from these models depends on the efficacy of technique employed to represent, organize, access, and interpret the healthcare and biological datasets.

There are enormous needs and opportunity for future work in the area of detection and prediction of mental illnesses. Trustworthy and verifiable techniques are needed for handling, retrieving, consolidating and interpreting the multi-modal and multi-structured mental health data that is present in various rapidly growing traditional repositories such as clinical records and EHRs as well as contemporary platforms such as social media. AI-based techniques need to be applied for personalizing treatment, identifying risks, detecting and preventing mental ailments in pre-nascent and nascent stages and to avoid cases of relapse. Innovative techniques are required to ensure the privacy preferences of patients so that mental health data is easily sharable for various research and decision-making purposes.

## References

1. Wongkoblaph A, Vadillo MA, Curcin V (2019) Modeling depression symptoms from social network data through multiple instance learning. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci* 2019:44–53
2. Zhu Y, Shang Y, Shao Z, Guo G (2018) Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Trans Affect Comput* 9:578–584. <https://doi.org/10.1109/TAFFC.2017.2650899>
3. Garg K, Kumar CN, Chandra PS (2019) Number of psychiatrists in India: baby steps forward, but a long way to go
4. Garg P, Glick S (2018) AI's potential to diagnose and treat mental illness. *Harv Bus Rev*
5. The Live Love Laugh Foundation (2018) How India perceives mental health: TLLLF 2018 national survey report
6. Fang R, Pouyanfar S, Yang Y et al (2016) Computational health informatics in the big data age. *ACM Comput Surv* 49:1–36. <https://doi.org/10.1145/2932707>

7. Andreu-Perez J, Poon CCY, Merrifield RD et al (2015) Big data for health. *IEEE J Biomed Heal Inform* 19:1193–1208. <https://doi.org/10.1109/JBHI.2015.2450362>
8. Miotto R, Wang F, Wang S et al (2017) Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 19:1236–1246. <https://doi.org/10.1093/bib/bbx044>
9. Hahn T, Nierenberg AA, Whitfield-Gabrieli S (2017) Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiat* 22:37–43. <https://doi.org/10.1038/mp.2016.201>
10. Tran T, Kavuluru R (2017) Predicting mental conditions based on “history of present illness” in psychiatric notes with deep neural networks. *J Biomed Inform* 75:S138–S148. <https://doi.org/10.1016/j.jbi.2017.06.010>
11. Zhang Y, Zhang O, Wu Y et al (2017) Psychiatric symptom recognition without labeled data using distributional representations of phrases and on-line knowledge. *J Biomed Inform* 75:S129–S137. <https://doi.org/10.1016/j.jbi.2017.06.014>
12. Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: *Proceedings of the 31 st international conference on machine learning*, Beijing, China, pp 29–30
13. Mikolov T (2013) Distributed representations of words and phrases and their compositionality. In: *NIPS’13: proceedings of the 26th international conference on neural information processing systems*, vol 2. ACM, pp 3111–3119
14. Zhang Y, Li H-J, Wang J et al (2018) Adapting word embeddings from multiple domains to symptom recognition from psychiatric notes. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci* 2017:281–289
15. Jackson R, Patel R, Velupillai S et al (2018) Knowledge discovery for deep phenotyping serious mental illness from electronic mental health records. *F1000 Res* 7:28. <https://doi.org/10.12688/f1000research.13830.2>
16. Holderness E, Dawkwell P, Bolton K et al (2019) Distinguishing clinical sentiment: the importance of domain adaptation in psychiatric patient health records
17. McCoy TH, Castro VM, Cagan A et al (2015) Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: an electronic health record study. *PLoS One* 10:1–10. <https://doi.org/10.1371/journal.pone.0136341>
18. Waudby-Smith IER, Tran N, Dubin JA, Lee J (2018) Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. *PLoS One* 13:e0198687. <https://doi.org/10.1371/journal.pone.0198687>
19. Cer D, Yang Y, Kong S et al (2019) Universal sentence encoder for English. In: *Proceedings of the 2018 conference on empirical methods in natural language processing (System Demonstrations)*, pp 169–174
20. Soleymani M, Asghari-Esfeden S, Fu Y, Pantic M (2015) Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans Affect Comput* 7:17–28. <https://doi.org/10.1109/TAFFC.2015.2436926>
21. Thorstad R, Wolff P (2019) Predicting future mental illness from social media: a big-data approach. *Behav Res Methods*. <https://doi.org/10.3758/s13428-019-01235-z>
22. Reece AG, Reagan AJ, Lix KLM et al (2017) Forecasting the onset and course of mental illness with Twitter data. *Sci Rep* 7:1–11. <https://doi.org/10.1038/s41598-017-12961-9>
23. Bagroy S, Kumaraguru P, De Choudhury M (2017) A social media based index of mental well-being in college campuses. *Conf Hum Factors Comput Syst Proc* 1634–1646. <https://doi.org/10.1145/3025453.3025909>
24. Husseini Orabi A, Buddhitha P, Husseini Orabi M, Inkpen D (2018) Deep learning for depression detection of Twitter users. In: *Proceedings of the fifth workshop on computational linguistics and clinical psychology*, pp 88–97. <https://doi.org/10.18653/v1/w18-0609>
25. Amir S, Coppersmith G, Carvalho P et al (2017) Quantifying mental health from social media with neural user embeddings. *Proc Mach Learn Healthc* 2017:1–17
26. Cohan A, Desmet B, Yates A et al (2018) SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions, pp 1485–1497

27. Yates A, Cohan A, Goharian N (2017) Depression and self-harm risk assessment in online forums. In: EMNLP 2017—conference on empirical methods in natural language processing, pp 2968–2978. <https://doi.org/10.18653/v1/d17-1322>
28. Reece AG, Danforth CM (2017) Instagram photos reveal predictive markers of depression. EPJ Data Sci 6. <https://doi.org/10.1140/epjds/s13688-017-0110-z>
29. Yazdavar AH, Mahdavejad MS, Bajaj G et al (2019) Fusing visual, textual and connectivity clues for studying mental health
30. Yazdavar AH, Al-Olimat HS, Ebrahimi M et al (2017) Semi-supervised approach to monitoring clinical depressive symptoms in social media. Proc 2017 IEEE/ACM Int Conf Adv Soc Networks Anal Min ASONAM 2017:1191–1198. <https://doi.org/10.1145/3110025.3123028>
31. Huang Y, Chiang C-F, Chen A (2019) Predicting depression tendency based on image, text and behavior data from Instagram. In: Proceedings of the 8th international conference on data science, technology and applications. SCITEPRESS—Science and Technology Publications, pp 32–40
32. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 3rd international conference on learning representations, ICLR 2015—conference track proceedings, pp 1–14
33. Deng J, Dong W, Socher R et al (2009) ImageNet: a large-scale hierarchical image database. In: 2009 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. IEEE, pp 248–255
34. Fellbaum C (Rider U and PU) (1998) WordNet: an electronic lexical database. In: Fellbaum C (ed) The MIT Press, Cambridge, MA
35. Al-Mosaiwi M, Johnstone T (2018) In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clin Psychol Sci 6:529–542. <https://doi.org/10.1177/2167702617747074>
36. Minor LB (2017) Harnessing the power of data in health. Stanford Med Heal Trends Rep 1–18
37. Esteva A, Robicquet A, Ramsundar B et al (2019) A guide to deep learning in healthcare. Nat Med 25:24–29. <https://doi.org/10.1038/s41591-018-0316-z>
38. Rosenfeld A, Benrimoh D, Armstrong C et al (2019) Big Data analytics and AI in mental healthcare
39. Travers C, Himmelstein DS, Beaulieu-Jones BK et al (2018) Opportunities and obstacles for deep learning in biology and medicine
40. Ghassemi M, Naumann T, Schulam P et al (2018) A review of challenges and opportunities in machine learning for health
41. Wongkoblap A, Vadillo MA, Curcin V (2017) Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 19. <https://doi.org/10.2196/jmir.7215>
42. Chen IY, Szolovits P, Ghassemi M (2019) Can AI help reduce disparities in general medical and mental health care? AMA J Ethics 21:167–179. <https://doi.org/10.1001/amajethics.2019.167>
43. Ravi D, Wong C, Deligianni F et al (2017) Deep learning for health informatics. IEEE J Biomed Heal Inform 21:4–21. <https://doi.org/10.1109/JBHI.2016.2636665>

# Comparison Analysis of Extracting Frequent Itemsets Algorithms Using MapReduce



Smita Chormunge and Rachana Mehta

**Abstract** Frequent itemset mining (FIM) is among widely known and essential data analytics techniques, to discover and extract frequently co-occurring items. However, due to the massive information available online, it is difficult to extract valuable data with the help of FIM algorithms. Traditional FIM suffers from scalability, memory and computation issues. Onto this context, MapReduce framework can be used which can handle those issues along with algorithmic parallelization. In this paper, algorithms have been explored and compared majorly on two approaches for the reduction of the computation cost. The first approach for mining frequent itemset is candidate generation and data pruning; it is adopted by Apriori, equivalence class transformation and Sequence-Growth algorithm. The second approach for mining is through pattern growth, and the algorithm under this is FP growth. The parameters taken into comparative analysis are search strategy and load balancing.

**Keywords** Frequent itemset mining · MapReduce framework · Data analytics · Apriori algorithm · Parallel FP growth · Equivalence class transformation

---

S. Chormunge (✉) · R. Mehta  
Computer Science and Engineering Department, Institute of Technology, Nirma University,  
Ahmedabad, Gujarat, India  
e-mail: [smita.darandale@nirmauni.ac.in](mailto:smita.darandale@nirmauni.ac.in)

R. Mehta  
e-mail: [rachana.mehta@nirmauni.ac.in](mailto:rachana.mehta@nirmauni.ac.in)

# 1 Introduction

The increasing usage of recent applications such as social networking, Internet of things (IoT) and Internet to produce humongous amounts of data has drastically emphasized the need for data mining on a large scale. The piquing interest in data analytics and mining which focuses on extracting meaningful and efficient knowledge from such enormous data has increasingly risen. The data mining provides various techniques for knowledge extraction and data analysis like clustering, outlier analysis, frequent pattern mining, classification and many others.

Frequent mining is one of the essential and famous techniques for extracting frequently co-occurrence patterns in a given data source. It has various applications ranging from weblog analysis, market basket analysis, credit card fraud detection, intrusion detection, DNA sequence and drug discovery to more recent like finding a defect-free ratio in the semiconductor industry [5]. The patterns can be itemsets, subsequences, graphs or substructures [6]. Frequent mining plays an important role in mining association, relationships or correlations among data [7].

The wide applicability of frequent pattern mining algorithms has increased its usage in current high data-intensive environments. However, the traditional algorithms suffer from scalability, memory and computations limitations. There is a need for improvement in such algorithms that can cope with the big data scenario using environments such as parallel one. Computational parallelism is a natural approach but comes along with challenges of the partitioning of work and memory scalability, load balancing and job scheduling.

Parallel programming is divided into two subcategories based on memory storage: 1. shared memory and 2. distributed [6]. Onto this context of distributed processing, Google has developed a MapReduce framework [8]. It provides the distributed processing of such large datasets on a big cluster. Apart from distribution, it also gives robustness and fault tolerance.

This paper encompasses a comparison of various MapReduce-based algorithms in the context of their usage in frequent itemset mining (FIM) on big data. Various algorithms have been explored like Apriori, FP growth, equivalence class transformation (ECLAT) and new ones like Sequence-Growth on MapReduce.

The remaining paper is organized as follows. Section 2 gives a brief background about the frequent itemset mining model and MapReduce framework. Section 3 covers the different frequent itemset mining algorithms using MapReduce, followed by a comparative analysis. Section 4 concludes the work and provides up with future extensions.



## 2 Background

### 2.1 Frequent Itemset Mining

The frequent itemset mining is frequent pattern mining algorithms for extracting frequently co-occurrence items from a given transactional data. Two core methodologies for FIM have emerged on the line of computation cost reduction. In the first step, it aims to prune the candidate frequent itemset search space, while in the second step, it focuses on reducing the number of comparisons which are required to determine itemset supported by pattern growth. Apriori algorithm [9] that sets the base of frequent itemset mining problem is based on candidate generation and pruning. FP growth [11] is based on pattern growth. Equivalence class transformation (ECLAT) [10] algorithm is like a scalable version of the Apriori algorithm.

Frequent itemset mining problem can be formally defined as follows [5].

Let “ $T = \{t_1, t_2, \dots, t_n\}$ ” be a set of items.  $TR = (\text{trid}, X)$  is a transaction where “trid” is a transaction identifier and “ $X$ ” is a set of items under the domain of  $T$ . A transaction database  $D = \{\text{tr}_1, \text{tr}_2, \dots, \text{tr}_m\}$  is a set of transactions. The support for a given itemset  $M$  is defined by the number of transactions which contains the itemset  $M$  in it. Equation [5] for the same can be defined as:

$$\text{support}(M) = |\{\text{trid} | M \subseteq X, (\text{trid}, X) \in D\}| \quad (1)$$

An itemset is termed as frequently occurring if the support is more than the threshold. Here, threshold,  $\delta$ , or minsup is a user-specified value also called minimum support value.

It can also be carried in the vertical database format. Here, each itemset is stored along with its cover (trid-list). The vertical database ( $D'$ ) is defined as follow:

$$D' = \{(t_j, C_{ij} = \{\text{trid} | t_j \in X, (\text{trid}, X) \in D\})\} \quad (2)$$

where  $C_{ij}$  is the trid-list of  $t_j$ . The support of an itemset  $N$  can be calculated by the intersection of a trid-list of any two subsets. It can be defined as follow:

$$\text{support}(N) = |\bigcap C_{ij}| \quad (3)$$

where the intersection is over the  $ij \in N$ .

Apriori algorithm and FP-growth algorithm work on the horizontal data format, while ECLAT algorithm works on the vertical data format. These algorithms being efficient adapt well to single machines with limited memory but need to be parallelized when working with a big data environment. Even some articles aim to provide a comprehensive review of recent research efforts on deep learning-based recommender systems [14].

## 2.2 *MapReduce Framework*

MapReduce is a programming model that provides a linked implementation to generate and process large data sources [8]. It is developed by Google in order to solve distributed memory systems problems. It has become known in the community for a heavy data-intensive computation. It provides methods to create parallel programs. Hadoop [12, 13], provides an open-source implementation for MapReduce.

In MapReduce, first of all, it will partition the data into equal size blocks. Along with the partitioning, the duplication of blocks will also be created and evenly distributed to the file systems automatically. The user does not need to worry about the distribution of data. The file systems can be of Hadoop or Google. In the next phase, MapReduce will try to re-execute the crashed task. During this re-execution, the tasks which were completed will not be executed again, eventually achieving good fault tolerance. In the later phase, the throughputs are increased through the re-assignment of incomplete tasks of slow or busy nodes to idle nodes in heterogeneous clusters. The idle nodes are those who have completed their part of work.

The MapReduce follows the below-mentioned task hierarchy:

- Data partitioning, with the help of Hadoop Distributed File System (HDFS).
- Allocation of partitioned nodes to worker nodes to carry out computations.
- Task scheduling of worker nodes by the master node.
- Two major task execution: Map and Reduce.
- The problem here will be specified in a key-value pair format.
- Individual keys will be processed by map phase, giving a new and different key-value as output.
- Each of the output key-value pair's value will be grouped by key, and then the value list will be processed by reducing phase.
- In the reduce phase, the final program output will be generated or the new key-value.
- The new key-value will serve as an input to the next map iteration.

MapReduce-based algorithms can go through several iterations. The map and reduce are independent of one another and hence they can run in parallel. The master node collects the results of a map and reduce phase in a file like HDFS or GFS, then schedules its workers for the next MapReduce phase. Hadoop framework is responsible for all communication tasks. Hadoop provides distributed cache, a file caching mechanism for global data access such as when the complete content of a file is to be read by the entire map or reduce workers. As the intermediate stage, outputs are stored in the file system, failure of a map or reduce phase can be recovered from checkpoints. This avoids re-execution of the entire job when an error or exceptions occurs.

### 3 Frequent Itemset Mining Algorithms

In this section, various frequent itemset mining algorithms that use MapReduce implementation are discussed.

#### 3.1 *Parallel FP-growth Algorithm [1]*

Parallel FP growth (PFP) is the parallel modification to the classic FP-growth algorithm implemented using MapReduce [11]. The FP growth suffers from several resource challenges such as storage limitation of huge FP tree, computation distribution, costly communication between inter-dependent FP tree and application-specific support value [1].

Parallel FP growth overcomes the storage, computation distribution and communication cost using the MapReduce framework. FP growth involves two phases: 1. construction of FP tree and 2. mining frequent patterns. These two phases involve two scans of the database. The first scan is used for frequency counting and storing it into F-list. In the next scan, FP tree is generated to represent the compressed version of the database. PFP works in a 5-step process with 3 MapReduce phases. The 5 steps are sharding, parallel counting, Item grouping, parallel FP growth and aggregating. The advantages of this algorithm are high linear speedup and achieve higher parallelization than traditional FP-growth algorithm and less computational dependencies between parallel tasks. It also has some drawbacks such as lacking an efficient load balancing mechanism for MapReduce worker nodes.

#### 3.2 *Balanced Parallel FP-growth Algorithm [2]*

The parallel FP-growth (PFP) [1] algorithm does not take into consideration load balancing strategies that are highly needed when a dataset is extremely large. To overcome this limitation of the PFP algorithm, a balanced parallel FP-growth (BPFP) algorithm is developed. BPFP estimates the workload of each mining unit first and then divides these units into several groups to balance the load.

BPFP algorithm provides a transaction database and minimum support count that uses two rounds of MapReduce. Two major differences, when compared with PFP, are as follows: First, the grouping strategy is balanced, and second, the aggregating step has been eliminated to discover all frequent items rather than top-K.

BPFP algorithm is a four-step procedure with 2 MapReduce jobs given below:

- **Sharding:** It partitions the database into smaller partitions and then stores those onto  $P$  different computers (worker nodes).

- **Parallel Counting:** Parallel counting uses a MapReduce job to count the occurrence of all items. This step in BFPF is similar to one in the PFP algorithm. The output is a sorted frequency list, F-list.
- **Balanced Grouping:** In this step, items present in the F-list get divided into  $Q$  number of groups. Load on FP growth is the summation of all the load of sub-FP growths. The sub-FP growth is growth on the conditional patterns. The balanced grouping is further divided into two steps: mining load estimation step and balanced partition step.
- **Parallel FP growth:** Mapper performs group dependent transactions, while reducer forms an FP tree recursively and outputs frequent occurring patterns.

The advantages of this algorithm are effective load balancing strategy, BFPF provides good speedup compared to PFP and low standard deviation of time consumed by each local FP growth. A drawback is that searching for a space partition using a single item is not efficient.

### 3.3 Apriori-Based Algorithms

Apriori algorithm is one of the well-known and simple methods for frequent itemset mining on transactional databases. The fundamental of parallelizing apriori lies in designing appropriate maps and reduce function for support counting and candidate generation. The three different approaches to Apriori algorithms are given below.

#### 3.3.1 Single Pass Counting Algorithm [3]

Single pass counting (SPC) algorithm is the simplest apriori implementation into MapReduce. It searches out the frequent  $k$  itemsets at the  $k$ th pass of database scanning in a map-reduce stage. The map task will generate the candidates and outputs  $\langle x, 1 \rangle$  in every pass of database scanning, for each candidate  $x$  confined in a transaction. The reduce assignment will collect the candidate support counts and resultant candidates having sufficient counts as frequent itemsets for the map tasks of the next pass. The result is stored as  $L_1$  in distributed cache, where  $L_1$  is a collection of all frequent 1-itemsets. This completes the support counting procedure of phase-1 of SPC.

In phase-2, a MapReduce job is carried out, where mapper reads a frequent  $L_1$  generated candidate set,  $C_2$ , in a hash tree using the apriori-gen function [9] of traditional Apriori algorithms. The mapper then reads a transaction  $t_i$  to form all candidates using subset function [9] and outputs result  $\langle x, 1 \rangle$  for each candidate  $x$  in a transaction. The reduce function performs the same procedure as phase-1, to perform support counting of all candidates and outputs  $\langle \text{itemset}, \text{count} \rangle$  for the one above the minimum threshold to the file  $L_2$ .

The later phases of SPC are all the same as phase-2, wherein each phase-k, mapper reads  $L_{k-1}$  from distributed cache to generate  $C_k$  in a hash tree. It then performs the support counting and outputs <itemset, count> to file  $L_k$  if the count is above the minimum threshold. The advantage of this algorithm is that it achieves better parallelization than the Apriori algorithm. Drawbacks are low worker utilization in the last phases of MapReduce due to a small number of candidates. Scheduling of maps and reducing phase is overhead compared to workload, and database loading cost is high.

### 3.3.2 Fixed Passes Combined-Counting Algorithm [3]

SPC algorithm being level-wise iterates  $n$  times of the map-reduce phase when the length of frequent itemsets is  $n$ . This incurs low worker node utilization, database loading and schedule. To overcome this, fixed passes combined-counting (FPC) algorithm combines candidates from several fixed phases in SPC and performs the support counting and candidate generation in a single phase.

The phase-1 and phase-2 of FPC are similar to the SPC algorithm. Phase-3 onwards, it combines candidates from a fixed number of database passes for support counting. For a particular example of 3 fixed passes, candidates of three consecutive lengths are counted in a single phase until no more frequent itemsets can be generated. Thus, FPC counts  $C_3$ ,  $C_4$  and  $C_5$  in single phase and so on using prefix trees. The map phase of the FPC algorithm here includes candidate generation for 3 consecutive steps. The apriori-gen and subset function are the same as that of the Apriori algorithm [9]. Advantages of this algorithm are less number of map-reduce phases than SPC, better utilization of worker nodes in the last phases and less number of database scans. Disadvantages are worker overloads due to the large number of candidates generated, may run out of memory when a large number of candidates are generated and increase in the number of false-positive candidates. Inability to prune candidates at intermediate passes and execution time may exceed SPC for a large number of candidates.

### 3.3.3 Dynamic Passes Combined-Counting Algorithm [3]

FPC algorithm though lessens the number of map-reduce phases and database scans, suffers from candidate overload, large false-positive candidates, resulting in degraded performance than simple SPC algorithm. To overcome the limitations of FPC, dynamic passes combined-counting (DPC) algorithm is developed. DPC performs the same functioning as FPC, except that it dynamically combines candidates of subsequent passes by considering the workload of the worker nodes. DPC also balances the reduction in map-reduce phases and an increase in the number of pruned candidates.

The phase-1 and phase-2 of the DPC algorithm are similar to the FPC and SPC algorithm. Starting from phase-3, DPC dynamically combines candidates in several

passes. DPC first calculates the candidate threshold  $c_t$ , which indicates the number of candidates that a mapper can count in a given phase. The mapper generates and aggregates the candidates level by level until they remain below the threshold value. Once the candidates are generated, support counting is carried out using a prefix tree.

DPC uses the execution time ( $\alpha$ ) of the previous phase to adjust the candidate threshold at each map-reduce phase. The candidate threshold also depends on the longest frequent itemset generated in the last phase. Its candidate threshold can be given as below.

$$c_t = \alpha * |L_{k-1}| \quad (4)$$

DPC thus takes advantage of SPC and FPC, works well by reducing map-reduce phases and dynamically adjusting workload by combining passes of light workload. Advantages of this algorithm are dynamic workload balancing, fewer database scans, less number of MapReduce phases than FPC and SPC, efficient pruning of infrequent candidates and less execution time than SPC and FPC. It also has some drawbacks such as workload calculation can be made efficient by utilizing worker configuration and partial removal of false-positive candidates.

### 3.4 ECLAT-Based Algorithms

Apriori-like algorithms are simple and efficient for mining frequent itemsets; however, they suffer from the high overhead of communication cost and initiation cost of MapReduce jobs, and also the generation of candidate itemset and scanning of databases is expensive [4, 5]. Equivalence class transformation (ECLAT) is a frequent itemset mining algorithm that uses vertical dataset format for the reduction in map-reduce iterations and fast support computation. ECLAT passes through the prefix tree using a depth-first manner. ECLAT prunes the complete subtree if an itemset is found to be infrequent. The prefix tree generated is sorted according to the individual frequency of item in ascending order, as this helps easily prune prefix trees at lower depths. The two variants of the ECLAT algorithm that work in a parallel environment are given below.

#### 3.4.1 Dist-ECLAT Algorithm [4]

Dist-ECLAT is a parallel version of the traditional ECLAT algorithm using the MapReduce framework [11]. It allocates the search space appropriately among the worker nodes. The search space distribution here is efficient as compared to data distribution, as data distribution needs combining and counting of local frequent itemsets to prune globally infrequent itemsets, which is an expensive task.

Dist-ECLAT works in a depth-first manner; hence in memory, it keeps a limited number of candidates, even when discovering large itemsets. Also, the diffset

approach of ECLAT alleviates the memory overhead problems. The Dist-ECLAT algorithm works as follow on the vertical database format as follows:

1. Searching frequent items: The vertical database is partitioned into shards and disseminated to mappers. From the provided shard, each mapper discovers the frequent singletons. The reducer job is to collect every frequent item which does not need any additional processing.
2.  $k$ -FIs Generation: Here, the set of frequent itemset notation as  $P_k$ , for which size  $k$  are produced. It distributed frequent singletons across the mapper, every mapper searched frequent supersets of items which are of  $k$ -sized by executing ECLAT for a level of  $k$ . At last,  $P_k$  is assigned by reducer to a fresh batch of  $m$  mapper, Using Round-Robin algorithm the distribution is completed.
3. Subtree Mining: Using ECLAT, this step consists of mining of the prefix tree initializing at a prefix as of the batch which is assigned to it. Every mapper does this job without any intervention, as subtree does not require any mutual information. The advantages of this algorithm are efficient search space distribution, no need for extra communication among mappers, and there is no verification for the results of overlapping mining and provides good speedup in a distributed environment. Drawbacks are it would not fit the tid-list of the item into main memory, complete dataset needed for generating low-level frequent itemsets, complete dataset needed to mine subtrees and do not adapt well to huge data, lacks scalability.

### 3.4.2 BigFIM Algorithm [4]

The Dist-ECLAT though being the distributed version of the ECLAT algorithm does not adapt well to the big data environment and has several memory and scalability issues. To overcome the limitations of Dist-ECLAT, BigFIM was proposed. BigFIM is a hybrid of ECLAT and Apriori algorithm that works effectively in a large scale environment.

The BigFIM algorithm can be given below.

1. Generating  $k$ -FIs: To cover the issue for large tid-lists needed to be loaded in main memory, BigFIM carries out a breadth-first search in this step using Apriori algorithm. The mapper task is to output the item with a value 1 if it is contained in that transaction. A reducer will combine entirely indigenous frequencies and outputs as a global frequent itemset. The resultant frequent itemsets may be redistributed for whole mappers for the further stage of breadth-first search which acts as candidates. The same thing is repetitive for  $k$  times, to acquire the  $k$ -FIs sets.
2. To search potential extensions: Here, it generates  $(k + 1)$ -FIs using global tid-list. Local tid-list is searched by mapper for  $k$ -FIs which provides for reducing function. All local tid-lists are combined by reducer and prepare one global tid-list which gives the whole prefix set to various mappers.

3. Subtree Mining: In this stage, the mappers work on separate prefix sets. A prefix set describes a conditional database which will be fitting into memory. To mine conditional databases, the mining part employs diffsets aimed at frequent itemsets with the help of a depth-first manner. The advantages of this method are efficient load balancing strategy, work well with massive data and provide good scalability. Disadvantages are high execution time compared to Dist-ECLAT and complexity increases in data switching from one format to another.

### 3.5 Sequence-Growth Algorithm [5]

Sequence-Growth (SG) is a novel mining algorithm of frequent itemset, which works on the concept of lexicographical order to generate candidate subsets to avoid expensive database scanning frequent itemsets produced by Sequence-Growth in a breadth wide support-based method (“lazy mining”). For every MapReduce repetition, it will prune away the infrequent itemsets. Due to its memory consumption and starting time of every MapReduce task deducts significantly.

#### 3.5.1 Lexicographical Sequence Tree

The functionality of lexicographical sequence trees is the construction of candidate generation space. Lexicographical sequence tree signifies every node as a subsequence, and the height of it resembles the length of the subsequences at that level. Concatenating every node to its suffix repeatedly in a manner of Sequence-Growth extends the lexicographical sequence tree. A  $(k + 1)$  length sequence is a merging of its  $(\text{length}-k)$  base nodes and parent’s suffix item. The advantages of this approach are efficient pruning strategy, less waiting time for a process once a job is completed, less number of database scans are required and better execution time than parallel versions of traditional algorithms. A disadvantage is item suffix generation and large-1 is time-consuming.

### 3.6 Comparative Analysis of Frequent Itemset Mining Algorithms

The comparative analysis of the frequent itemset mining techniques is explored in this section.

Table 1 shows FIM algorithms comparison based on three parameters such as search strategy breadth-first search (BFS) or depth-first search (DFS), communication cost handling and load balancing features.

The parallel FP growth achieves parallelization as compared to its traditional counterpart. However, parallel FP-growth algorithm lacks load balancing capacity;



**Table 1** Comparative analysis of FIM algorithms

Technique	Underlying algorithm	Search strategy	Communication cost handling	Load balance handling
PFP [1]	FP-Growth	DFS	Yes	No
BPFP [2]	FP-Growth	DFS	Yes	Yes
SPC [3]	Apriori	BFS	Yes	No
FPC [3]	Apriori	BFS	Yes	No
DPC [3]	Apriori	BFS	Yes	Yes
Dist-Eclat [4]	Eclat	DFS	Yes	Yes
BigFIM [4]	Apriori and Eclat	BFS and DFS	Yes	Yes
Sequence-growth [5]	Sequence-growth	BFS	Yes	Yes

for the same balanced parallel FP growth can be used. BPFP is not efficient in search space partitioning. The next discussion involves widely used Apriori algorithms; they are good when it comes to finding a frequent pattern through candidate itemset generation and pruning approach. But, they suffer from initiation and communication cost of MapReduce. In ECLAT-based algorithm, Dist-ECLAT uses vertical dataset in the reducer phase and achieves good communication cost handling and efficient search space distribution. Further BigFIM helps to achieve a scalability issue of Dist-ECLAT. At last, the Sequence-Growth algorithm is discussed, which helps to achieve better initiation time and memory consumption of MapReduce as compared to other algorithms. Sequence-Growth also shows better execution time as compared to parallel counterparts of traditional algorithms.

## 4 Conclusion

Frequent pattern mining algorithms have wide applicability in various areas from industry, academic, scientific research to the medical field. However, with the advent of Internet technology, the volume of data being generated and processed has increased to meet the analysis need of such massive data, traditional frequent itemset mining algorithms fall short.

This paper has explored and compared various frequent itemset mining algorithms like FP growth, Apriori, ECLAT and Sequence-Growth based on the MapReduce framework. Among which, Sequence-Growth has shown good scalability and efficient processing in distributed environments. It also has a good possible extension in incremental behavior pattern mining.

## References

1. Li H, Wang Y, Zhang D, Zhang M, Chang EY (2008) PFP: parallel fp-growth for query recommendation. In: Proceedings of the ACM conference on recommender systems, pp 107–114
2. Zhou L, Zhong Z, Chang J, Li J, Huang JZ, Feng S (2010) Balanced parallel FP-growth with MapReduce. In: IEEE youth conference on information computing and telecommunications (YC-ICT), pp 243–246
3. Lin MY, Lee PY, Hsueh SC (2012) Apriori-based frequent itemset mining algorithms on MapReduce. In: Proceedings of the 6th international conference on ubiquitous information management and communication. ACM, pp 76–84
4. Moens S, Aksehirli E, Goethals B (2013) Frequent itemset mining for big data. In: IEEE international conference on big data, pp 111–118
5. Liang YH, Wu SY (2015) Sequence-growth: a scalable and effective frequent itemset mining algorithm for big data based on MapReduce framework. In: IEEE international conference on big data, New York, NY, pp 393–400
6. Anastasiu DC, Iverson J, Smith S, Karypis G (2014) Big data frequent pattern mining. In: Frequent pattern mining, vol 9783319078212. Springer International Publishing, pp 225–259. [https://doi.org/10.1007/978-3-319-07821-2\\_10](https://doi.org/10.1007/978-3-319-07821-2_10)
7. Goleand S, Tidke B (2015) Frequent itemset mining for big data in social media using Clust-BigFIM algorithm. In: International conference on pervasive computing (ICPC). IEEE, pp 1–6
8. Dean J, Ghemawat S (2004) Mapreduce: simplified data processing on large clusters. In: Proceedings of the sixth symposium on operating system design and implementation (OSDI), pp 137–150
9. Agrawal R, Srikant R (1995) Mining sequential patterns. In: Proceedings of the eleventh international conference on data engineering. IEEE, pp 3–14
10. Zaki MJ (2000) Scalable algorithms for association mining. *IEEE Trans Knowl Data Eng* 12(3):372–390
11. Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. In: Proceedings of the SIGMOD international conference on management of data, New York, NY, USA, pp 1–12
12. White T (2009) Hadoop: the definitive guide. O'Reilly Media
13. (2016) Hadoop: the lay of the land. Dr. Dobb's. <http://www.drdoobs.com/database/hadoop-the-lay-of-the-land/240150854>
14. Zhang S, Lina Y, Aixin S, Yi T (2019) Deep learning based recommender system: a survey and new perspectives. *ACM Comput Surv* 52(1):38. Article 5. <https://doi.org/10.1145/3285029>

# Customer Lookalike Modeling: A Study of Machine Learning Techniques for Customer Lookalike Modeling



Anna Mariam Chacko, Bhuvanapalli Aditya Pranav,  
Bommanapalli Vijaya Madhvesh, and A. S. Poornima

**Abstract** In the current online advertising scenario, lookalike methods of audience expansion are highly valued to enhance the existing advertising campaigns. Machine learning approaches are highly favored, and the typically used models involve distance-based, keyword-based and classification techniques. In this perspective, the proposed research work has incorporated a comparative study of the basic distance-based clustering model for lookalike classification against a positive unlabeled classification model. The results demonstrate a greater accuracy of the positive unlabeled classification method.

**Keywords** Online advertising · Lookalike modeling · Audience expansion

## 1 Introduction

Targeted advertising is an advertising approach that seeks to present online advertisements to the most receptive target audience by identifying the common traits based on product or service requirements as stated by the advertiser and utilizing sophisticated models built on data science and machine learning concepts. It is a fast-growing sub-domain of the IT industry that holds great significance in the world of business. Evolved from more traditional advertising methods such as television, newspaper, hoardings and so on, targeted online advertising presents businesses with the opportunity to expand into the online space and come ever closer to their target audience. This online space provides a huge platform where products can be promoted and which generates vast amounts of data that can be analyzed to understand user trends and behaviors and so provide a more personalized experience. Advertising platforms such as Google, Facebook and Yahoo undertake extensive research into discovering

---

A. M. Chacko · B. A. Pranav · B. V. Madhvesh · A. S. Poornima (✉)  
Department of Computer Science and Engineering, Siddaganga Institute of Technology, Tumkur,  
Karnataka, India  
e-mail: [aspoornima@sit.ac.in](mailto:aspoornima@sit.ac.in)

new techniques of targeting users and improving existing techniques in order to improve the overall process of an ad campaign for all the stakeholders involved: the advertising company, the platform and the end user.

The common approach to lookalike targeting is to study a seed set of users already known to be interested in the product in terms of various features such as their personal details (age, gender, location, etc) as well as their interactions and interests. Then, based on the formed understanding of customer behavior, a wider segment of users can be studied in order to identify users within this segment that conform to behavioral patterns exhibited by the seed set of users. These identified users will then be the target audience desired. Such intensive study of user data can be accomplished effectively by making use of machine learning techniques. The task is then to create a technique that can handle variable seed set sizes and variable user set sizes, i.e., fully scalable lookalike modeling techniques.

Current work that has been carried out in this field involves the use of distance-based models that determine the distance between the seed set and every user. Classification is the most commonly used machine learning method used to identify lookalikes. Another frequently used approach is building keyword and topic distributions.

In this study, we attempt to identify a method of customer lookalike modeling that is simple to understand while providing results that are an improvement over traditional out-of-box machine learning algorithms. We have utilized a publicly available dataset [9] containing 1,997,410 records of user's click data, i.e., each record contains details of a display ad like its advertiser as well as details of the user and the user's interaction with the ad. The dataset is described by 12 attributes: ad impression; the hashcode of the ad url; ad and advertiser identifiers; depth and position of the ad; query, keyword, title, description and user identifiers; and the target class click which classifies whether the user clicked the ad or not.

The seed user set ( $P$ ) contains users that have clicked on the ad and the universal set ( $U$ ) contains users for whom it must be determined whether they would potentially click on a similar ad or not. For this, our study began with a simple  $k$ -NN implementation that classified users using feature vectors of features selected using a random forest algorithm. A positive unlabeled (PU) classification model was then developed to improve the observed performance metrics. The recall,  $R$ , which represented the ability of the model to correctly classify lookalikes and  $C$ , the percentage size of  $U$  that is selected as the lookalike set, are used to measure performance.

## 2 Literature Survey

Currently, lookalike modeling is a field that is still undergoing research and has begun to be implemented in certain platforms. Lookalike modeling is based on the idea that similarities between users can be related to their probability of providing the same feedback on advertisements. Traditionally, the task of searching for a target audience is guided by a set of advertiser defined settings which are simple assumptions about

user demographics, communities or interests. Below some of the existing schemes are summarized.

The paper [1] proposes an adaptive model that utilizes information learned during current advertising efforts to explore and select relevant user features for effective categorization. The model is a reinforcement model where actions consist of adding users to the universal set  $U$  and makes use of Bernoulli Naive Bayes to classify individual users. Here, Thompson sampling is used to select a random sample of features at every iteration instead of using every feature, in an attempt to mitigate negative influence of bias factors that may not be filtered out if iterative modeling taking feedback from earlier derived models is not performed.

Paper [2] proposes a two-stage audience expansion model comprising a combination of a classifier model trained for a single user set of data with a model based on users modeled as an embedding of features. It makes use of transfer learning to utilize historical logs and lightweight feature selection to enable faster on-boarding of new user into the system. The model described uses two components: The global user embedding model and the embedding-based user-advertiser scoring model.

Paper [3] approaches feature selection using a variety of sampling methods to train an initial linear classifier and logistic regression is then used to converge to the required user set. This paper treats the problem as a prediction problem which, upon receiving a seed user set and universal user set, is transformed into a positive unlabeled learning problem. The major goal of sampling is to pick out a reliable set of negative instances. The model used in the proposed model is a factorization machine which is used in order to extract feature crosses and reduce the influence of high-dimensional sparse features.

Paper [4] is a study of an implicit lookalike modeling method which learns a user's profile based on their online behavior and uses this knowledge to predict their ad response. This relies upon two tasks: web browsing prediction, which determines a user's tendency to visit a given Web site based on their web browsing history, followed by ad response prediction, which estimates the probability of a user responding to a given ad based on the ads that user has responded to in the past. The prediction models for both tasks are learned using a joint data discrimination framework.

Paper [5] uses cookie and mobile ad identifier information along with relevant metadata to construct user identities rather than relying directly on identifiers. Off-policy evaluation techniques are used to evaluate the potential of identity powered lookalikes models while also taking into account previously studied biases.

Paper [6] provides three scoring models that attempt to score users in a way which relates directly to the business metrics that are most relevant to the advertiser in an attempt to measure the potential value of the users.

Paper [7], in an attempt to overcome constraints induced during modeling of the expanded audience, explores binary mapping from sparse, high-dimensional data to a binary address space through the use of an adversarial training procedure.

Paper [8] presents an associative classification (AC) approach for campaigns with low number of positive examples. Pairs of features are used to derive rules to build a rule-based associative classifier, with the rules being sorted by frequency-weighted log-likelihood ratio (F-LLR). The top  $k$  rules, sorted by F-LLR, are then applied to

any test record to score it. Individual features can also form rules by themselves, though the number of such rules in the top  $k$  rules and the whole rule set is very small.

### 3 Architecture and Implementation

Figure 1 describes how the selected dataset is loaded into Spark and split into test and training set of appropriate sizes. Various machine learning algorithms are used to develop suitable lookalike learning models on the training data which are then tested for accuracy. Then the developed models are then evaluated on the identified metrics, and finally, a list of lookalikes is produced.

The basic architecture used in this project is displayed in Fig. 2. A binary dataset of click-prediction data was used which was divided into three segments—positive examples ( $P$ ), unlabeled examples ( $U$ ) and a set of known positive test examples (PT). The positive examples were those users who had clicked upon the display ad. Unlabeled examples represented the set of users among which lookalikes of the users in set  $P$  were to be found. Positive test examples were the users upon which the trained model was tested for efficiency.

First, a naive, out-of-box 5-NN model was used as shown in Fig. 3. The training data was fed into a 5-NN classifier. Next, the trained 5-NN classifier was applied on the unlabeled data which classified them as having clicked or not clicked on the ad. The model proved unsuitable to be carried further based on the results observed, so a positive unlabeled classification model was then developed.

The classification model was implemented in the following way according to the architecture given in Fig. 4. The model requires both positive and negative examples in order to be trained accurately, so in order to obtain a set of users to use as negative examples, some of the members of  $U$  are randomly assumed to be negative to form a set  $N$ .

The selected set of negative examples  $N$  was then combined with  $P$  to give the training set PN. The unlabeled set  $U$  was updated as in (1) below.

$$U = U - N \tag{1}$$

The classifier, which forms the core of the model, was then trained on PN. In our study, we have implemented the model using two different classifiers: one versus rest (OVR) classifier and Naive Bayes (NB) classifier. PT was used to test the classifier and determine its  $R$ -value.  $R$  is computed by running the model with a set of known positives and determining how many it correctly classifies. When a satisfactory  $R$ -value was obtained, the classifier was then applied to the updated  $U$  in order to obtain those users that appeared similar to the users in  $P$ . This set of users was the required set of lookalikes. This model followed the algorithm seen in Fig. 5. The size of the output set as a percentage of  $U$  is calculated to give the value  $C$ .

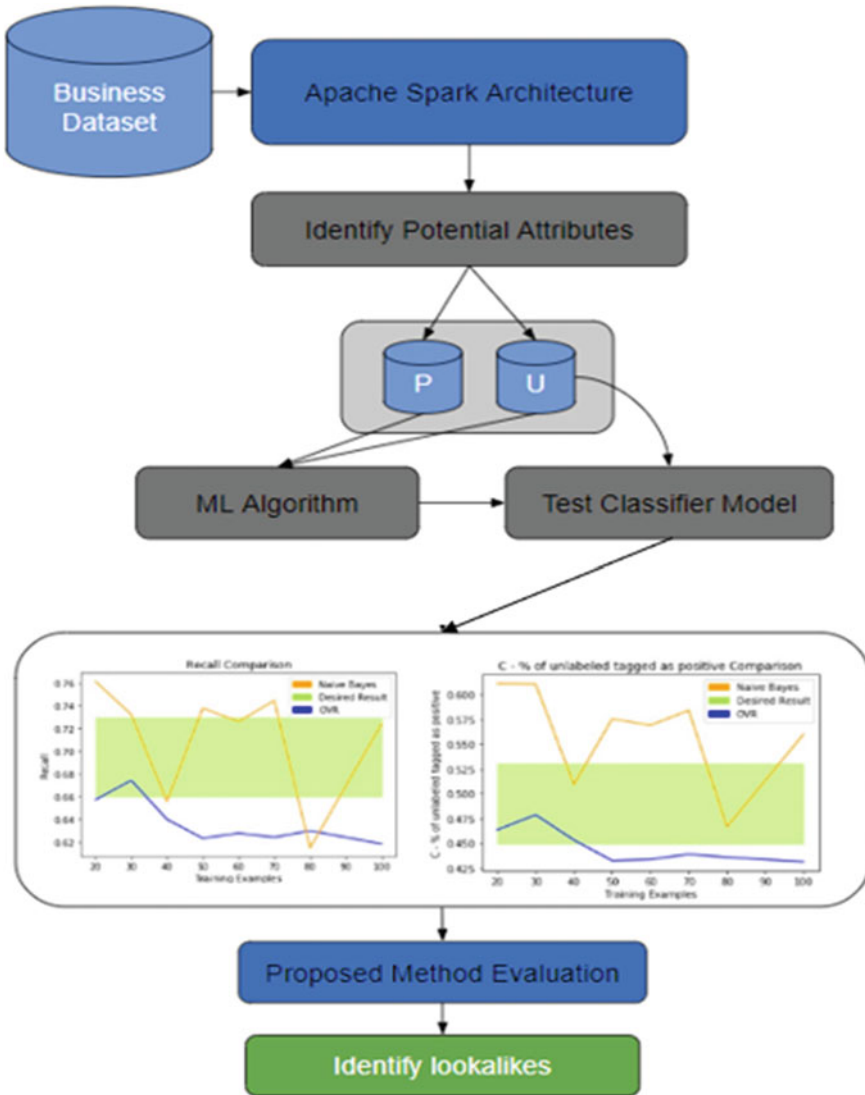
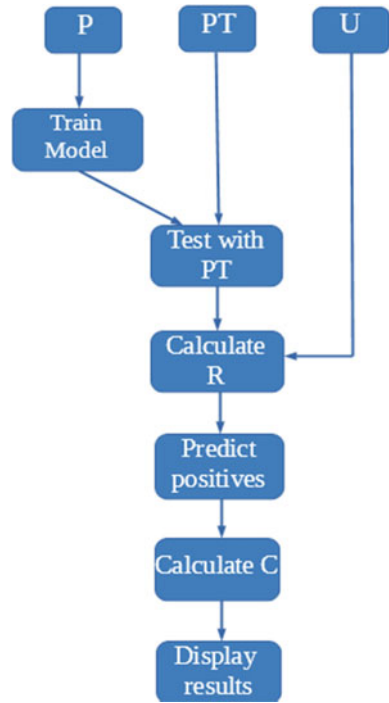


Fig. 1 Implementation architecture: acquire dataset and prepare, evaluate developed model and produce lookalike list

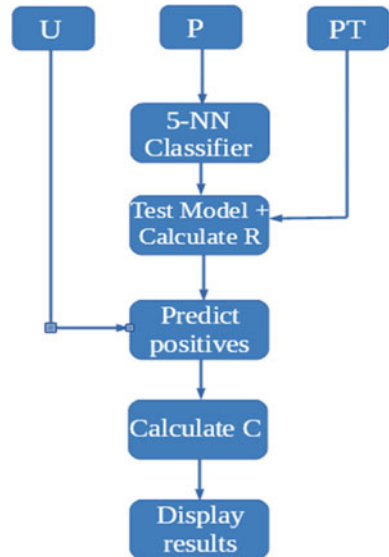
### 3.1 Tools and Technology

**Git:** Git is a distributed version control system for tracking changes and coordinating work during software development. Git supports nonlinear, distributed development and is compatible with many protocols and existing systems.

**Fig. 2** General data flow in lookalike modeling



**Fig. 3** Data Flow in 5-NN model of lookalike modeling





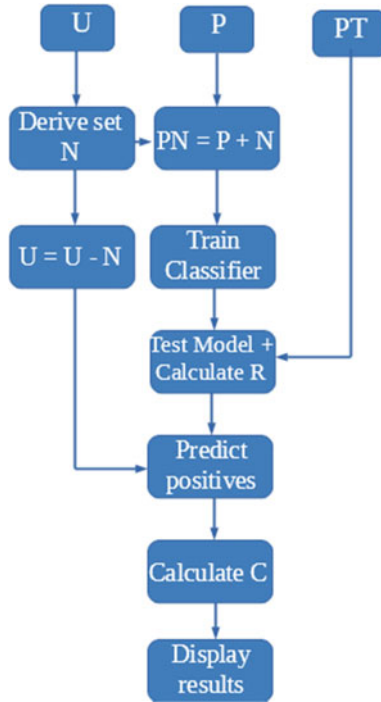


Fig. 4 Data flow in PU classification model of lookalike modeling

- P – Positive data
- PT – Positives to be tested
- U – Unlabeled data
- Let N = Assumed to be negative ( $|N|=|P|$  approx where  $|x|$  represents cardinality of a set x)
- Train Data = P + N
- Test Data = U – N
- Train the models with Train Data (P + N)
- Predict values for PT and calculate the recall R.
- R = Accuracy of the model over PT.
- Predict the Test Data (U – N)
- Get the Lookalikes L and calculate C (% of unlabeled tagged as positive)
- $C = (1 - A) * 100$   
A = Accuracy of the model over U.

Fig. 5 Algorithm of PU classification

Some characteristics of Git support nonlinear development, Distributed development, Compatibility, Efficient handling of large projects, Toolkit-based design, Pluggable merge strategies, Periodic explicit object packaging.

The extremely flexible versioning system provided by Git enhanced team collaboration and eased the process of developing and combining various parts of the solution.

**Anaconda:** Anaconda is a free and open-source distribution of the Python programming language for scientific computing. It is used in the fields of data science, machine learning, predictive analytics and so on in order to simplify package management. It is suitable for Windows, Linux and MacOS. Anaconda includes the conda package manager and 1500 packages, as well as a desktop graphical user interface called Anaconda Navigator, which serves as an alternative to using command-line commands. Anaconda comes with a number of applications, some of which are Jupyter Notebook, JupyterLab, Spyder, Glue, Orane, R Studio and Visual Studio Code. All versions of Anaconda also include the conda package and environment manager which is open-source, language agnostic and cross-platform. Conda was created to help handle conflicts during installation of packages that were not being handled by the pip package management system. Conda scans the environment to ensure that any package you are trying to install will not create conflicts with other previously installed packages and installs a compatible set of dependencies or raises a warning if this is not possible. Packages can be installed from Anaconda's central repository, from Anaconda Cloud or from a local repository. Custom packages can also be made and shared on Anaconda Cloud.

Anaconda was used as a suitable Python distribution that came with the packages necessary for implementing our machine learning solution. Its ready integration with Jupyter Notebook also proved highly conducive to the development of the solution.

**Jupyter Notebook:** Jupyter Notebook is an interactive computational environment that is web-based and used to create Jupyter Notebook documents which are JSON documents following a versioned schema. Each notebook consists of an ordered list of cells containing lines of code and the output of executing that code. Jupyter Notebook is built upon several open-source libraries like IPython, JQuery, Bootstrap and Tornado. Jupyter connects to different kernels to allow programming with different languages but connects to IPython kernel by default. Jupyter supports 49 types of kernel for languages like Python, R and so on. Kernels communicate with other Jupyter components using ZeroMQ over a network and thus they can be on the same or different machines. Kernels are not attached to specific documents and can be connected to by several clients at the same time. Kernels usually allow only one language to execute however there are some exceptions.

Jupyter Notebook, with its interactive interface and compartmentalized code organization, allowed us to develop our solution in an easy to understand and structured way.

**Apache Spark:** Apache Spark is an open-source cluster computing framework. Spark provides data parallelism and fault tolerance. Spark was developed to overcome the

limitation of MapReduce which imposed a linear data flow on distributed programs. Spark allows both iterative algorithms as well as data querying similar to databases. The iterative algorithms are typically machine learning algorithms, which was the original reason for developing Spark. For storage, Spark interfaces with a large number of existing distributed storage systems but also allows implementation of custom solutions.

As our problem statement called for the use of substantially large datasets, in order to store and utilize these datasets efficiently Spark proved very useful.

## 4 Results

Two evaluation metrics were used to determine the efficiency of the models. The first was the recall,  $R$ , which represented the ability of the model to correctly classify lookalikes. Ideal recall of a model is 100%.  $R$  is calculated as in (2) below,

$$R = A_{PT} = N_{cPT} / N_{tPT} \quad (2)$$

where  $A_{PT}$  represents the accuracy of the model over set PT,  $N_c$  is the number of correctly classified instances in set PT and  $N_t$  is the total number of classifications made for set PT (i.e., the number of instances in set PT).

The second metric is  $C$ , the percentage size of  $U$  that is selected as the lookalike set. Ideal value of  $C$  for a model is between 5 and 30%. A  $C$ -value of 100% would mean the entire unlabeled set are lookalikes of set P which suggests that the model is underfitting or that data is insufficient. A  $C$ -value below 5% percent suggests that the model is overfitting.  $C$  is calculated as in (3) below,

$$C = (1 - A_U) / 100 = (1 - N_{cU} / N_{tU}) / 100 \quad (3)$$

where  $A_U$  is the accuracy of the model over  $U$ ,  $N_{cU}$  is the number of correctly classified instances in set  $U$  and  $N_{tU}$  is the total number of classifications made for set  $U$  (i.e., the number of instances in set  $U$ ).

The 5-NN model was extremely naive and produced a recall of 100% and also the percentage of unlabeled examples classified as positive examples was 100%.

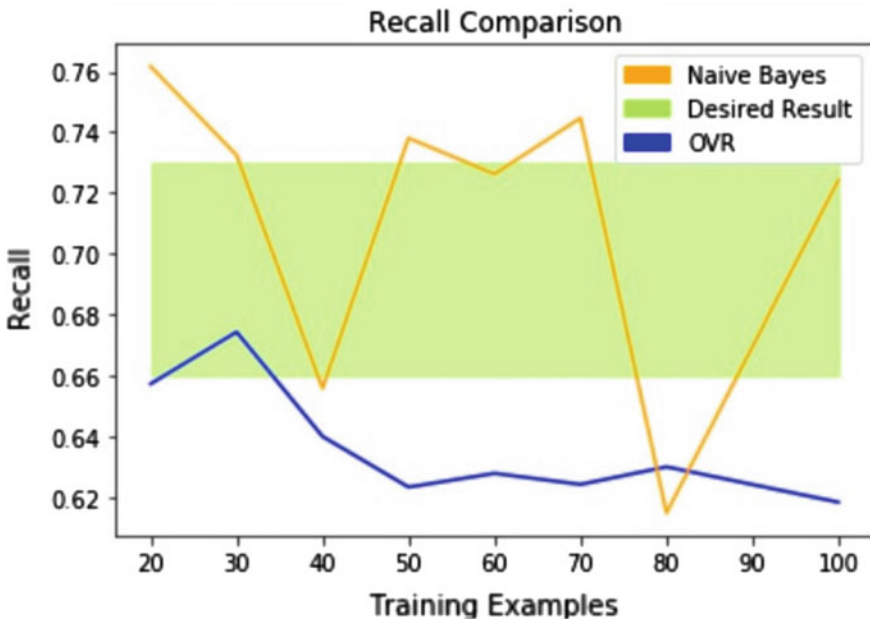
The PU classification model implemented with OVR classifier is seen to have a recall of 61.84% and  $C$ -value of 43.15%, which is a huge improvement over the naive k-NN model. The classification model implemented with NB classifier is seen to have a recall of 72.43% and  $C$ -value 55.99%.

The PU classification model described above was successively applied to larger and larger segments of the dataset to observe performance in terms of  $R$  and  $C$  and the results are summarized in Table 1. While the  $C$ -value of the classification model using NB classifier is consistently seen to be better than the  $C$ -value of OVR classifier,  $R$ -values in both classifiers are seen to vary across the different segment sizes.  $R$  of

**Table 1** Recall ( $R$ ) and % of unlabeled examples correctly classified as positive ( $C$ ) values for different segment sizes of data

% of dataset considered	OVR $R$	NB $R$	OVR $C$	NB $C$
10	0.602169	0.707742	0.406896	0.555207
20	0.657416	0.761710	0.463689	0.610922
30	0.674496	0.732478	0.478837	0.610506
40	0.640104	0.655945	0.453408	0.509401
50	0.623426	0.738126	0.432613	0.575457
60	0.627965	0.726259	0.433837	0.568972
70	0.624314	0.744718	0.439017	0.584074
80	0.630106	0.614932	0.436006	0.466803
90	0.622983	0.619882	0.435566	0.470502
100	0.618465	0.724263	0.431536	0.559973

NB classifier can be observed to be better than that of OVR classifier in 8 instances. Figures 6 and 7 show a graphical comparison of these values.



**Fig. 6** Comparison of  $R$ -values of one versus rest and Naive Bayes approaches

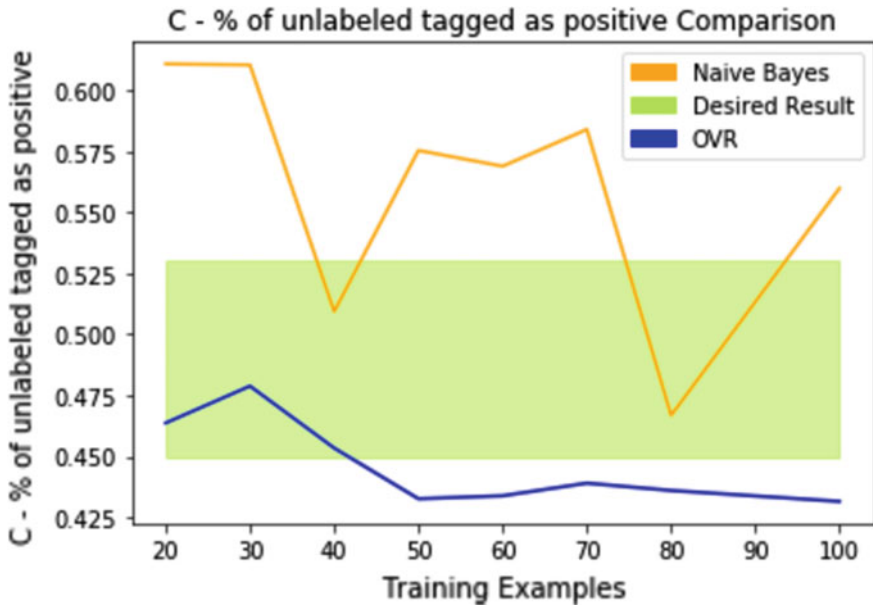


Fig. 7 Comparison of C% values of one versus rest and Naive Bayes approaches

## 5 Future Scope

The PU classification model can be observed randomly by selecting instances from  $U$ , which is to be assumed as negative examples. This creates a problem, where if certain instances that have been assumed as negative are actually positives, these do not get reflected in the final list of lookalikes. To overcome this problem, we can explore some sampling techniques like spy sampling or bootstrap sampling in order to select these instances from  $U$  that are more likely to be negative. We also see that the C-values of both PU classification variations, while good, are not within the ideal range of 5–32%. Thus, efforts can also be made by providing more training data to bring the C-values into the ideal range.

**Acknowledgements** We would like to take this opportunity to thank our institution for providing an environment conducive to carrying out this research. We thank our professor and guide Dr. A. S. Poornima for mentoring us throughout the course of the research and assisting us toward its completion. We also thank all those who encouraged us and wished us well for their support.

## References

1. Popov A, Iakovleva D (2018) Adaptive look-alike targeting in social networks advertising. *Procedia Comput Sci* 136:255–264

2. deWet S, Ou J (2019) Finding users who act alike: transfer learning for expanding advertiser audiences. In: KDD '19: proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, pp 2251–2259
3. Jiang J, Lin X, Yao J, Lu H (2019) Comprehensive audience expansion based on end-to-end neural prediction, special interest group on information retrieval eCom. France, Paris
4. Zhang W, Chen L, Wang J (2016) Implicit look-alike modelling in display ads. In: European conference on information retrieval. Springer, pp 589–601
5. Cotta R, Hu M, Jiang D, Liao P (2019) Off-policy evaluation of probabilistic identity data in lookalike modeling. In: Proceedings of the twelfth ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING (WSDM '19). Association for Computing Machinery, New York, NY, USA, pp 483–491
6. Ma Q, Wagh E, Wen J, Xia Z, Ormandi R, Chen D (2016) Score look-alike audiences. In: 2016 IEEE 16th international conference on data mining workshops (ICDMW), Barcelona, pp 647–654
7. Doan KD, Yadav P, Reddy CK (2019) Adversarial factorization autoencoder for look-alike modeling. In: Proceedings of the 28th ACM international conference on information and knowledge management (CIKM '19). Association for Computing Machinery, New York, NY, USA, pp 2803–2812
8. Mangalampalli A, Ratnaparkhi A, Hatch A, Bagherjeiran A, Parekh R, Pudi V (2011) A feature-pair-based associative classification approach to look-alike modeling for conversion-oriented user-targeting in tail campaigns, pp 85–86
9. Click Prediction data set. <https://www.openml.org/d/1218>
10. Wikipedia. [https://en.wikipedia.org/wiki/Anaconda\\_\(Python\\_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution))
11. Wikipedia. [https://en.wikipedia.org/wiki/Apache\\_Spark](https://en.wikipedia.org/wiki/Apache_Spark)
12. Wikipedia. <https://en.wikipedia.org/wiki/Git>

# Cold Start in Recommender Systems—A Survey from Domain Perspective



Rachna Sethi and Monica Mehrotra

**Abstract** In recent years, the growth of the Internet has led to the emergence of recommender systems, as embedded engines, in various domains. These engines are in a way helping tools for users, as they suggest items and services according to users' tastes and preferences. But these recommendations depend upon past user/item liking history. Non-availability of past user/item history leads to the situation of cold start. The cold-start problem is one of the major factors affecting the efficiency of recommender systems. This survey explores the cold-start problem from a domain perspective. The scenarios of cold start in different domains and techniques used to find solutions have been explored. Most of the researches have proposed solutions for e-commerce/e-business sites as they have more economic value. In recent years, deep learning techniques have gained momentum and are being used for solving cold start. Solutions for this problem in domains other than e-commerce/e-business need to be explored using deep learning methods.

**Keywords** Recommender system · Cold start · Product recommendations · Machine learning · Deep learning

## 1 Introduction

Recommender systems can be considered as intelligent systems providing suggestions for items as per the likings of the user [1]. The suggestions may be a list of items for shopping, a list of songs, or selective online news to read. So, they can be considered helping tools for the users that help in better decision making. In recent years, recommender engines have become an integral part of several application

---

R. Sethi (✉) · M. Mehrotra

Department of Computer Science, Jamia Millia Islamia, Delhi, India

e-mail: [rachna188476@st.jmi.ac.in](mailto:rachna188476@st.jmi.ac.in)

R. Sethi

Sri Guru Gobind Singh College of Commerce, University of Delhi, Delhi, India

domains like news, movies, books, music, articles, queries, social tags, jokes, restaurants, garments, financial services, life insurance, dating sites, and twitter pages. Most e-commerce applications have built-in recommender engines to boost sales.

The fundamental techniques used to generate recommendations in recommender systems are collaborative and content-based filtering [2]. The collaborative techniques generate recommendations for an active user using preferences of other users having very similar likings. Collaborative-filtering (CF) techniques are either model-based or memory-based. These techniques are based on data mining and machine-learning algorithms like the k-nearest neighbor (KNN) and other clustering algorithms, artificial neural networks, and Bayesian classifiers. The latent factor model or matrix factorization is one of the most popular techniques. In content-based filtering techniques, the building of a user profile is done by extracting features from the description of items that have already been seen or liked by the user. So, other users' profiles are not needed as they do not influence the resulting recommendation.

Knowledge-based filtering is another technique that uses explicit knowledge about users, items, and user preferences and recommends items to users by applying recommendation criteria [3]. It applies reasoning to recommend which items to which users in which context. This technique is useful in circumstances in which the two traditional recommendation (content-based and collaborative-filtering) approaches cannot be applied.

## 2 Cold Start

The efficiency of recommender systems is majorly affected by cold start. Cold start is a situation when a recommender system does not have sufficient history or information about a user or an item to make relevant suggestions [2]. There are three scenarios which may give rise to cold start [3].

### 2.1 *System Cold Start*

A system cold-start situation occurs when a new system is introduced to the user community. In this scenario, there is no interaction history for items as well as users. Thus, this cold-start problem hinders the recommender system to make relevant predictions thereby affecting the efficiency of the recommender engine.

In the case of a system cold start, using knowledge-based recommender engines might be helpful to recommend products to the users by inferring the preferences of the users. Knowledge-based recommender systems depend on the features of the items and the knowledge about how users' interests or preferences are met by these features.



## **2.2 *User Cold Start***

User cold-start scenario occurs when a new user enters the system. As no previous interactions of this user with the system (likes/dislikes) are available, the recommender system faces problems in generating predictions for this new user.

The common strategy used to overcome the user cold-start problem is recommending popular items to the users and further reducing the results with the help of the user's contextual information. The most popular items/products may be recognized by analyzing current trends and occasions. By collecting auxiliary information about the user such as the device used to access the system or the geographical location of the user, the recommender engine may be able to personalize the product/item to the user.

## **2.3 *Item/Product Cold Start***

Item or product cold-start problem occurs when a new item is introduced in the system. When a new item is introduced in the system, there are no previous ratings for this item, so it becomes difficult for the recommender engine to predict a target user for the item [4].

The item cold-start problem can also be solved using knowledge-based recommender engines that extract features from item descriptions. Item-item similarity can also be calculated and users that have liked similar items will be targeted.

# **3 Domain-Specific Cold-Start Scenarios**

Recently, recommender systems have become the building blocks for various applications as they help in identifying user needs, improve users' satisfaction, and build user trust. So, more and more organizations are looking for efficient recommender systems. Though recommender systems are being used in a large number of domains, cold-start scenarios and solutions have been proposed for specific domains only. This section attempts to explore some of those domains.

## **3.1 *e-Commerce/e-Business***

Recommender systems have become an integral part of many business applications with high economical values [5]. These applications are pioneers in bringing the concept of generating recommendations for the users, for example, Netflix's movie recommendations, Amazon's shopping items recommendations, and Pandora's

music recommendations. A famous example is Netflix, in which a user is asked to rate a set of movies upon signing on the platform to indicate their preferences and the system then recommends movies to them based on rating patterns. But a large portion of users and items are “cold” in many applications. Lack of interaction history for new users and new items leads to cold start.

### ***3.2 Crowdsourcing***

Crowdsourcing is a platform for job providers and job seekers. Job providers can search for workers with specific skills to process a specific task [6]. Similarly, job seekers can find tasks on these platforms that match their interests and skills [7]. The quality of deliverables depends on the worker-task matching. So, a recommender system is a must for a crowdsourcing platform. These recommender systems suffer the problem of cold-start workers (i.e., new developers/workers) whose task interaction data is not available. Specifically, in software crowdsourcing platforms, tasks have a short lifetime so they are generally insufficient in their data accumulations. This missing developer behavior history makes the cold-start developer problem severe.

### ***3.3 Food Journaling***

In recent years, with people becoming more health-conscious, websites recommending healthy meals are getting more and more attention. For recommending healthy meals, learning users’ preferences is a pre-requisite. On-boarding surveys and food journaling are the two main approaches used in food journaling. But they suffer from cold-start problems as they are dependent on data entered by the user. An active food journaling user makes about 3.5 entries per day [4, 8]. So, sufficient data for making recommendations will be available only after a long time [8].

### ***3.4 Location-Based Social Networks***

A social network is an abstract structure having individuals connected by different kinds of relations, such as friendship and mutual interests [9]. Location-based social networks (LBSNs) are social networks that are strengthened by localization techniques that allow users to share location-based content such as geo-tagged photos and check-ins [9]. Using LBSN, a user can “check-in” at a physical place such as a restaurant, and it would optionally include its location into their posts. As check-in is optional, data of each and every visit of all users may not be available with the system. A user might visit a few places many times and many places a few times [10] and

may not check-in every time. So, there will not be enough data for recommendations, resulting in the cold-start check-in problem.

### 3.5 News Recommendation

In the last few years, the popularity of online news has increased rapidly. News providers are now attracting more online readers than their core of regular offline readers through globally connected systems. To increase their base of readers, news websites are continuously looking for ways to provide news as per the liking of users. In general, a recommender system relies on interaction information of logged-in users collected over time. But for news recommender systems, it is very difficult to create a reliable user profile as the users are most often not logged-in. This leads to a cold-start scenario where predicting a news item for a user becomes difficult [11].

## 4 Techniques for Solving Cold Start

The research in recommender systems started in the mid-1990s [4]. This research area is comparatively new as compared to researches in other traditional information systems (e.g., databases or search engines). In recent years, the research in this area gained momentum as recommender systems provide an added advantage to customers as well as to service providers. As the cold start is a major factor affecting the efficiency of recommender systems, numerous approaches have been proposed for solving this problem. This section attempts to classify those techniques.

The cold start arises when the information required for making suggestions is not available in the system. All the solutions proposed so far for this problem have defined methods that collect the missing information. These solutions can be broadly divided into two categories [4]: Explicit and implicit. In explicit solutions, information is collected by directly interacting with the user and the user profile is built. Users may be asked to fill up a questionnaire or give a rating [12]. Implicit solutions try to have minimal interaction for understanding new user's preferences. They try to fill the missing information in different ways. The implicit techniques can be further broadly classified into the following categories:

- Adaptive Traditional Techniques

The earlier recommender systems were based on pure traditional techniques. To solve the cold-start problem, adaptive strategies like combining association rules and clustering techniques, ontological matrix factorization, have been successfully attempted since 1998. And still, researchers are working in this direction. The adaptive strategy proposed in [13, 14] achieved an improvement of 36% by combining association rules and clustering techniques over the association rule technique. An adaptive genetic

algorithm using aggregate function has been used to improve prediction accuracy in [15]. Overlapping co-cluster recommendation algorithm based on the overlapping of co-clusters of clients and products enables the discovery of interesting and useful buying patterns [16].

- Social Network-Based Techniques

Building user profiles through social media has become a major area of research for the last few years. Many social networks like Twitter, Facebook have become highly popular in recent years. The social profile of a user tells about likes, opinions, interests, dislikes, friends, and demographics of that user. People are also influenced by their friends. Hence, social profiles and social networks of users can be utilized to build their profile for solving cold-start problem. Work in this area started around 2004–05. Based on this concept, a solution that generates social interaction-based recommendations between the users on Facebook has been proposed in [17] and through opinion leaders in [18]. In [19], genetic algorithm is used to optimize social trust parameters.

- Deep Learning-based Techniques

Around 2013–14, researchers started exploring deep learning architectures to solve the cold-start problem. Deep learning methods are being used over the core traditional CF techniques without making any changes in the core model [20–22]. It has been observed that the accuracy of recommendations has increased with deep learning-based techniques. Deep learning-based recommender engines are capable of providing high-quality recommendations overcoming challenges of traditional models [23]. Multi-modality and cross-domain deep networks have been proposed in [5, 24] for cold-start recommendations. Deep learning has the advantage of making recommendations explainable [25].

Table 1 lists the different techniques proposed in surveyed papers, categorized under the above-mentioned categories. The domains and datasets used for proposed solutions have also been listed. Domain-specific distribution of solutions for cold-start problem is given in Fig. 1.

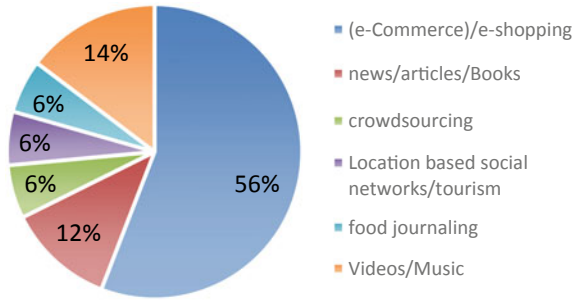
## 5 Conclusion

Recommender systems are web-based tools that generate recommendations as per the likings of the user. The advantage of these systems to service providers is that they help in identifying user needs, improve users' satisfaction, and build user trust. They are being used in various domains like News, e-Shopping, crowdsourcing, food journaling, e-tourism, and e-learning. The basis of recommendation generation is the history of user/item interaction. But the performance of the recommender system degrades in cold-start scenarios where there is insufficient user/item interaction history. Various solutions have been proposed so far for solving cold-start

**Table 1** Classification of implicit techniques

Category features	Datasets (Domains)	Reference
<i>Adaptive traditional techniques</i>		
Combining association rules and clustering technique over the association rule technique	Movielens, Jester, Netflix (e-Commerce)	[13, 14]
Implicit behavior-based solution	Books (e-Commerce)	[17]
Adaptive genetic algorithms	Yahoo!Movies (e-Commerce)	[15]
Overlapping co-clusters of clients and products	B2B-DB, CiteULike, Movielens, Netflix (e-commerce)	[16]
An ontological sub-matrix factorization-based approach, adapting matrix factorization	Groceries (e-commerce)	[1]
Topic sampling using a hybrid of clustering and classification	Zhubajie, Oschina, and Witkey (CrowdSourcing)	[7]
adaptive clustering	MovieLens (e-commerce)	[26]
<i>Social network-based techniques</i>		
Based on social interactions, opinion leaders	Books (e-commerce) Epinions.com (e-commerce)	[17, 18]
Based on directed and transitive trust with timestamps and profile similarity from the social network	MovieLens (e-commerce)	[27]
Genetic algorithm-based optimization of social trust parameters	Twitter (e-commerce)	[19, 28, 29]
<i>Deep learning-based techniques</i>		
Deep learning architectures over the core traditional CF techniques	Careerbuilder (Crowdsourcing) Amazon books, Amazon movies Douban Movies Netflix Programmable Web (e-commerce)	[20, 22, 23, 30–34]
Multi-modality and cross-domain deep network	MovieLens (e-commerce)	[5, 24]
Explainable recommendations	Amazon e-Commerce dataset	[25]
Deep learning for youtube recommendations	Youtube videos (Videos)	[35]
Deep learning and social networks	Epinions, Flixster (e-commerce)	[36]
DropoutNet	CiteULike, ACM RecSys 2017 (e-commerce)	[37]

**Fig. 1** Domain-specific distribution of solutions of cold-start problem



problem but it is still an open problem. The proposed techniques work efficiently but independently.

From a domain perspective, the e-commerce sites have more economic values; therefore, most of the researches for cold-start problem have been done in this domain. It was observed that more than 50% of proposed solutions, collected for the survey, belonged to the e-commerce/e-business domain. But the cold-start scenario is more severe in other domains, for example, in the news recommender system, the user is most often not logged-in, so the creation of a reliable user profile is really difficult. Every time a user visiting the news site is treated as a cold user, irrespective of how many times he/she has visited the site before. Also, datasets for these domains are very limited. Most of the work has been done on movie recommender systems using the movie lens dataset.

The techniques proposed in various researches can be broadly classified into three categories: Adaptive traditional techniques, social network-based techniques, and deep learning techniques. Deep neural networks for recommendation are in the early stages. Deep learning techniques have shown the capability of overcoming challenges of traditional models. Multi-modality, cross-domain recommendations, and explainable recommendations seem to be possible with these techniques.

So future work may include

- (1) finding solutions for cold-start problems in other domains (News/Food Journaling/LSBN) using deep learning techniques,
- (2) creating more datasets for other domains, and
- (3) making recommendations explainable.

## References

1. Jones AM et al (2017) An ontological sub-matrix factorization based approach for cold-start issue in recommender systems. In: International conference on current trends in computer, electrical, electronics and communication (ICCTCEEC), 8–9, September 2017, Mysuru, India
2. Liphoto M et al (2016) A survey on recommender systems. In: Third international conference on advances in computing, communication and engineering: ICACCE 2016: proceedings, 28–29 November 2016, Durban, South Africa

3. Selva Rani B, Ananda Kumar S (2019) Mitigating cold start problem in a personalised recommender system. *Int J Innov Technol Explor Eng (IJITEE)* 8(7):212–216. ISSN 2278-3075, ISSN: 2278-3075
4. Gope J, Jain SK (2017) A survey on solving cold start problem in recommender systems. In: *IEEE international conference on computing, communication and automation (ICCCA 2017)*, 5–6 May 2017. <https://doi.org/10.1109/ccaa.2017.8229786>
5. Sun M et al (2018) A multi-modality deep network for cold-start recommendation. *Big Data Cogn Comput* 2(1):7. <https://doi.org/10.3390/bdcc2010007>
6. Aldahri E et al (2015) Towards an effective crowdsourcing recommendation system: a survey of the state-of-the-art. In: *Proceedings—9th IEEE international symposium on service-oriented system engineering, IEEE SOSE 2015 Institute of Electrical and Electronics Engineers Inc*, pp 372–377. <https://doi.org/10.1109/SOSE.2015.53>
7. Yang Y et al (2017) Cold-start developer recommendation in software crowdsourcing: a topic sampling approach. In: *Proceedings of the international conference on software engineering and knowledge engineering, SEKE. Knowledge Systems Institute Graduate School*, pp 376–381. <https://doi.org/10.18293/SEKE2017-104>
8. Yang L et al (2017) Yum-Me: a personalized nutrient-based meal recommender system. *ACM Trans Inform Syst* 36:1. <https://doi.org/10.1145/3072614>
9. Zheng Y et al (2009) GeoLife2.0: a location-based social networking service. In: *Proceedings—IEEE international conference on mobile data management*, pp 357–358. <https://doi.org/10.1109/MDM.2009.50>
10. Gao H et al (2014) Addressing the cold-start problem in location recommendation using geo-social correlations. *Data Min Knowl Discov* 29(2):299–323. <https://doi.org/10.1007/s10618-014-0343-4>
11. Trevisiol M et al (2014) Cold-start news recommendation with domain-dependent browse graph. In: *RecSys 2014—proceedings of the 8th ACM conference on recommender systems. Association for Computing Machinery, Inc*, pp 81–88. <https://doi.org/10.1145/2645710.2645726>
12. Aggarwal CC *Recommender systems*
13. Sobhanam H, Mariappan AK (2013) A hybrid approach to solve cold start problem in recommender systems using association rules and clustering technique
14. Vizine Pereira AL, Hruschka ER (2015) Simultaneous co-clustering and learning to address the cold start problem in recommender systems. *Knowl Based Syst* 82:11–19. <https://doi.org/10.1016/j.knosys.2015.02.016>
15. Hassan M, Hamada M (2017) Improving prediction accuracy of multi-criteria recommender systems using adaptive genetic algorithms. In: *Intelligent systems conference (IntelliSys)*
16. Vlachos M et al (2019) Addressing interpretability and cold-start in matrix factorization for recommender systems. *IEEE Trans Knowl Data Eng* 31(7):1253–1266. <https://doi.org/10.1109/TKDE.2018.2829521>
17. Bedi P et al (2015) Handling cold start problem in recommender systems by using interaction based social proximity factor. In: *International conference on advances in computing, communications and informatics (ICACCI)*, 10–13 August. SCMS, Aluva, Kochi, Kerala, India
18. Mohammadi SA, Andalib A (2017) Using the opinion leaders in social networks to improve the cold start challenge in recommender systems. In: *3th international conference on web research (ICWR)*, 19–20 April 2017. IEEE, Tehran, Iran
19. Alahmadi DH, Zeng XJ (2016) Twitter-based recommender system to address cold-start: a genetic algorithm based trust modelling and probabilistic sentiment analysis. In: *Proceedings-international conference on tools with artificial intelligence, ICTAI. IEEE Computer Society*, pp 1045–1052. <https://doi.org/10.1109/ICTAI.2015.149>
20. Wei J et al (2017) Collaborative filtering and deep learning based recommendation system for cold start items. *Exp Syst Appl* 69:1339–1351. <https://doi.org/10.1016/j.eswa.2016.09.040>
21. Ying H et al (2016) Collaborative deep ranking: a hybrid pair-wise recommendation algorithm with implicit feedback. In: *Bailey J et al (eds) Advances in knowledge discovery and data mining. Springer International Publishing*. <https://doi.org/10.1007/978-3-319-31750-2>

22. Yuan J et al (2015) Solving cold-start problem in large-scale recommendation engines: a deep learning approach. In: IEEE international conference on big data, 5–8 December 2015, Washington DC, USA
23. Tilahun B et al (2017) A survey of state-of-the-art: deep learning methods on recommender system. *Int J Comput Appl* 162(10):17–22. <https://doi.org/10.5120/ijca2017913361>
24. Elkahky A et al (2015) A multi-view deep learning approach for cross domain user modeling in recommendation systems. In: WWW 2015—proceedings of the 24th international conference on world wide web. Association for Computing Machinery, Inc, pp 278–288. <https://doi.org/10.1145/2736277.2741667>
25. Chen X et al (2019) Visually explainable recommendation. In: Verma M, Ganguly D (eds) SIGIR 2019—42nd international ACM SIGIR conference on research and development in information retrieval. Association for Computing Machinery, Inc, pp 1281–1284. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>
26. Sanjeevi SG et al (2015) An adaptive clustering and incremental learning solution to cold start problem in recommendation systems. *Int J Comput Sci Electron Eng (IJCSEE)* 3(1). ISSN 2320-4028 (Online)
27. Reshma R et al (2016) Alleviating data sparsity and cold start in recommender systems using social behaviour. In: International conference on recent trends in information technology (ICRTIT). IEEE
28. Mueller J (2017) Combining aspects of genetic algorithms with weighted recommender hybridization. In: ACM international conference proceeding series. Association for Computing Machinery, pp 13–22. <https://doi.org/10.1145/3151759.3151765>
29. Pradeep K et al (2018) Comparative analysis of recommender systems and its enhancements
30. Bai B et al (2020) DLTSR: a deep learning framework for recommendations of long-tail web services. *IEEE Trans Serv Comput* 13(1):73–85. <https://doi.org/10.1109/TSC.2017.2681666>
31. Bansal T et al (2016) Ask the GRU: multi-task learning for deep text recommendations. In: RecSys 2016—proceedings of the 10th ACM conference on recommender systems. Association for Computing Machinery, Inc, pp 107–114. <https://doi.org/10.1145/2959100.2959180>
32. Zhang S et al (2019) Deep learning based recommender system: a survey and new perspectives. <https://doi.org/10.1145/3285029>
33. Zhang S et al (2017) Deep learning based recommender system: a survey and new perspectives. <https://doi.org/10.1145/3285029>
34. Zhou W et al (2018) Deep learning modeling for Top-N recommendation with interests exploring. *IEEE Access* 6:51440–51455. <https://doi.org/10.1109/ACCESS.2018.2869924>
35. Covington P et al (2016) Deep neural networks for YouTube recommendations. In: RecSys 2016—proceedings of the 10th ACM conference on recommender systems. Association for Computing Machinery, Inc, pp 191–198. <https://doi.org/10.1145/2959100.2959190>
36. Deng S et al (2017) On deep learning for trust-aware recommendations in social networks. *IEEE Trans Neural Netw Learn Syst* 28(5):1164–1177. <https://doi.org/10.1109/TNNLS.2016.2514368>
37. Volkovs M et al (2017) DropoutNet: addressing cold start in recommender systems. In: 31st conference on neural information processing systems (NIPS 2017), Long Beach, CA, USA



# Image Classification Using Machine Learning Techniques for Traffic Signal



Satthi Reddy, Nishanth, Praharsha, Dattatreya Dash, and N. Rakesh

**Abstract** This paper discusses the application of image classification that can be used in controlling a car, unmanned automatically, to pause, move, take deviation as and when any obstacles come on its way, also follow traffic signals. Our work will focus on making cars to avoid accidents, unlike the manual ones that are prone to accidents. The reasons may be many lapses on the part of the driver physically, mentally or sudden medical emergencies, etc. As part of this work, CNN classifiers have been used to perform image classification. Using CNN there was accuracy of 90%. Further, this accuracy may be increased to over 95%. This work involves the use of OpenCV that controls the hardware such as Raspberry Pi, four DC motors, Raspberry Pi cam, Arduino Uno.

**Keywords** Convolutional neural network · Computer vision · Transfer learning

---

S. Reddy (✉) · Nishanth · Praharsha · D. Dash · N. Rakesh  
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita  
Vishwa Vidyapeetham, Bengaluru, Bangalore 35, India  
e-mail: [satthireddy1997@gmail.com](mailto:satthireddy1997@gmail.com)

Nishanth  
e-mail: [raghughambir4189@gmail.com](mailto:raghughambir4189@gmail.com)

Praharsha  
e-mail: [harshavakkalanka.110@gmail.com](mailto:harshavakkalanka.110@gmail.com)

D. Dash  
e-mail: [dattaroyalchallengers@gmail.com](mailto:dattaroyalchallengers@gmail.com)

N. Rakesh  
e-mail: [n\\_rakesh@blr.amrita.edu](mailto:n_rakesh@blr.amrita.edu)

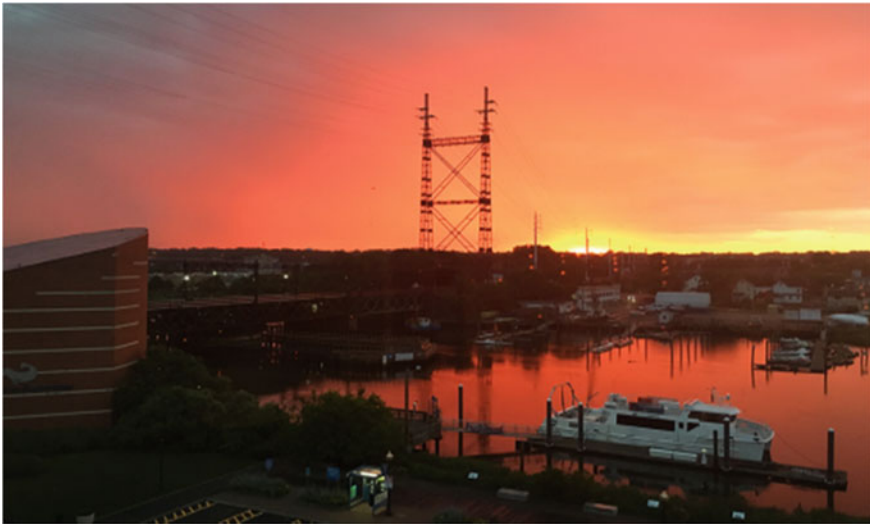
# 1 Introduction

The basic building blocks of an image are pixels. Every image consists of a set of pixels. There is no finer granularity than the pixel [1, 2]. Normally, a pixel has information about the “color” or the “intensity” of light in an image. If an image is represented as a collection of grids, then every unit of a grid is itself a pixel.

For example, in Fig. 1, the image has a resolution of  $1000 \times 750$ , meaning it has 1000 pixels horizontally and 750 pixels vertically. An image can be conceptualized as a (multidimensional) matrix. In this case, our matrix has 1000 columns (the width) with 750 rows (the height). Overall, there is  $1000 \times 750 = 750,000$  number of pixels in the image. Usually, images are represented in two ways:

1. Grayscale/single channel
2. Color

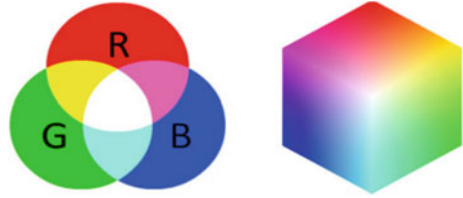
In Fig. 2, it can be seen that a grayscale image/single channel has pixels with scalar values ranging from 0 to 255, where zero corresponds to “black” and 255 being “white.” The gray color has some scale from 0 to 255. So if 0 is chosen then it means it is going to the darker side and if 255 is chosen then it is going to a lighter side. As shown in the above gradient, it can be seen that there are 0 and 255. The value 0 shows a darker side and value 255 shows the lighter value of the color.



**Fig. 1** Example image

**Fig. 2** Grayscale image



**Fig. 3** RGB color model

In the color pixel, the images are in the form of RGB color space. Pixels in the RGB color space are no longer a scalar value like in a grayscale/single channel image—instead, the pixels are represented by a list of three values: one value for the red component, one for green and another for blue. To define a color in the RGB color model, values for each of the pixel component R, G, B are defined. The amount of value determines the color.

There are three color channels in the RGB, and each color channel has ranged from 0 to 255 means a total of 256 colors shades, in which the value of pixel 0 shows there no color and values of 255 shows full pixel values (Fig. 3).

The values of the pixel must be in between [0, 255] range and must be an unsigned 8-bit integer value.

As the neural network is built, the preprocessing of the image is done by performing mean subtraction or scaling, that require converting the image to a floating-point data type. Keep this point in mind as the data types used by libraries loading images from disk (such as OpenCV) will often need to be converted before the application of learning algorithms to the images.

## 2 Methodology

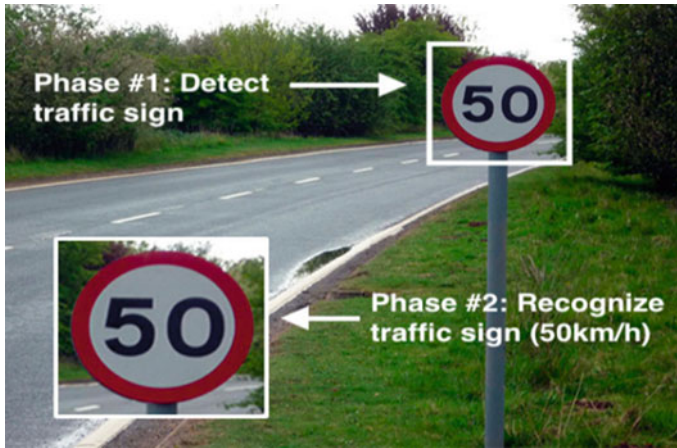
In the first part of the project, traffic sign classification is being explained and described how it will recognize the signal. Also, the datasets that are going to be used for the custom traffic sign classification are explained.

The different code structure and models are explained which are created for the classification. The model class, TrafficSignNet, which is a convolution neural network (CNN) [3–7], is used for the training time. This method will give us the best classification with high accuracy along with how it can be used for more custom data [8] for better results in future.

### What is the traffic sign classification?

Traffic sign classification teaches the machine about the meaning of the road signboard along with the speed limit as shown in Fig. 4 automatically.

This technology of classification [9–11] is used to teach automatic cars because as there are no drivers the machine has to know what is the meaning of coming signboard on the roadside and at which speed the car has to move. As shown in



**Fig. 4** Traffic sign classification

Fig. 5 of dataset images, there are different kinds of traffic signs available and will be trying to classify the similar type of dataset which is available on the Kaggle group site.

The dataset **GERMAN TRAFFIC SIGN RECOGNITION BENCHMARK** contains German traffic signals, having nearly forty classes and fifty thousand images. The number of images is increased to one lac so that more accurate results are got.

In the project classification of an image, two major steps involved are given below:

**Localization** The localization means [12, 13] wherein the whole image of our object is placed. In the whole part of an image, our required object can be in any place.

**Recognition** After localization [14] which place our object is in the image is known and can be found out our object for classification.

**Facing challenges with the GTSRB (dataset)** The difficulties in this dataset of German traffic signs are explained below:

**Poor image quality** The quality of the dataset image is very poor, as even human not able to see exactly what this image.

Also, viewpoint variation [15] also affects the dataset image, as a similar image is small and sometimes the same is very large to easily recognize as shown in Fig. 6.

Figure 6 displays a visualization of a number of these factors of variation. The variation can be,  $t$  in an image how different images were captured. Sometimes images are captured closely and far by the camera. But both image objects are same. This also affects the image classification. For example, no matter the angle in which the Raspberry Pi is captured, but still it is Raspberry Pi.

Scale variation has to account as well, have you ever ordered a tall, grande, or venti cup of coffee from Starbucks? Technically they are all the same—a cup of coffee, but different sizes.



Fig. 5 Traffic sign dataset

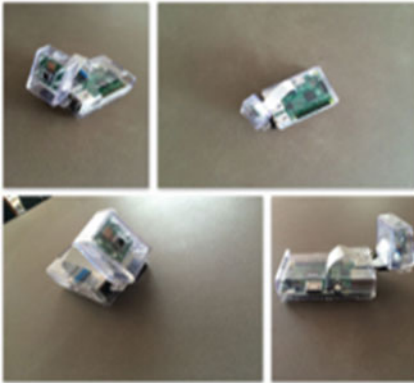
Furthermore, that same venti coffee will look dramatically different when it is photographed up close and when it is captured from farther away. Our image classification methods must be tolerable to these types of scale variations.

One of the hardest variations to account for is deformation. The character of television series Gumby is elastic, stretchable, and capable of contorting his body in many different poses and can look at these images of Gumby as a type of object deformation—all images contain the Gumby character; however, they are all dramatically different from each other.

The image classification should also be able to handle occlusions, where large parts of the object that should be classified are hidden from view in the image. On the left, we have a picture of a dog. And on the right, we have a photograph of the same dog, but notice how the dog is resting underneath the covers, occluded from our view. The dog is still clearly in both images—just more visible in one image than



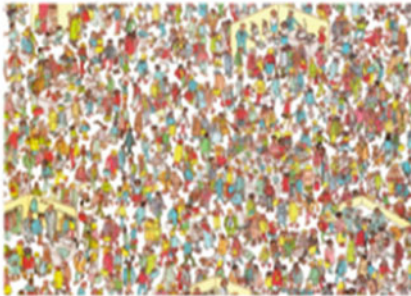
**Viewpoint Variation**



**Occlusion Variation**



**Background Clutter**



**Scale Variation**



**Deformation**



**Illumination Variation**



**Intra-class Variation**



**Fig. 6** Viewpoint variation during classification

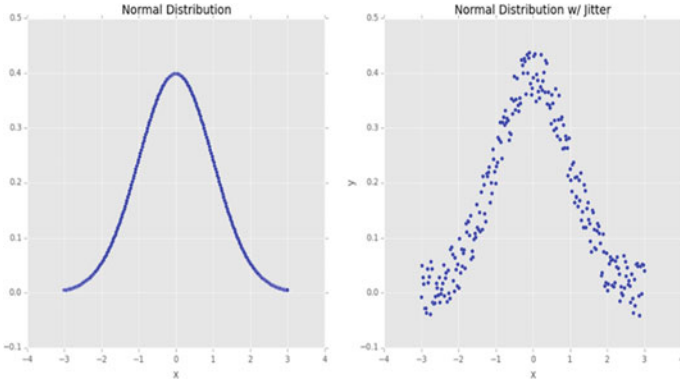


Fig. 7 Normal distribution

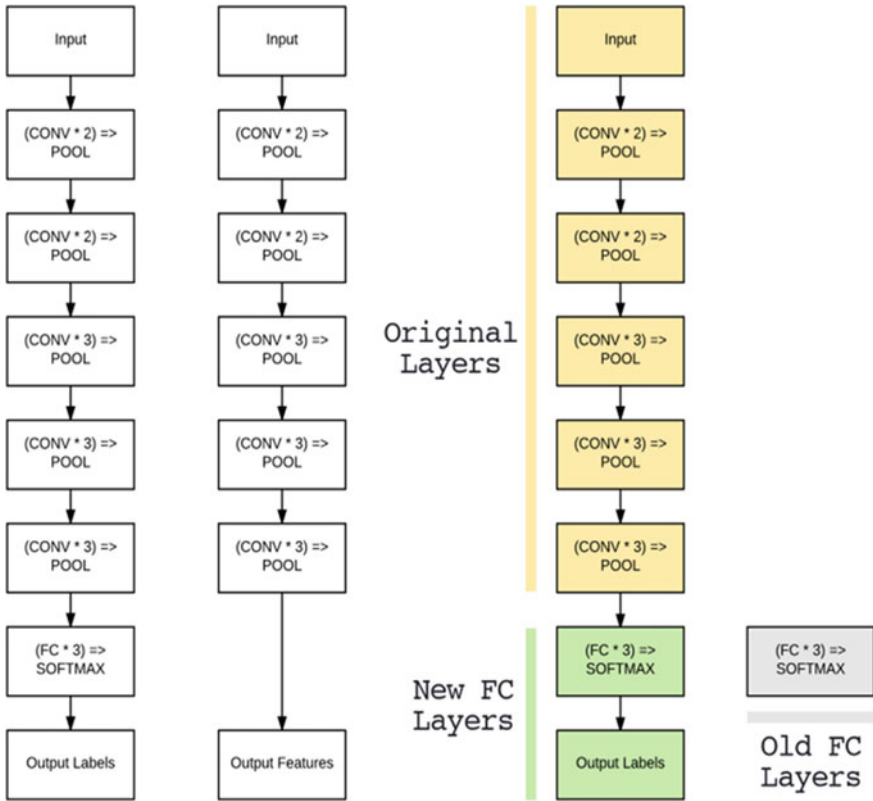


Fig. 8 Transfer learning

the other. Image classification algorithms should still be able to detect and label the presence of the dog in both images.

Apart from the challenges of deformations and occlusions mentioned above, there is a need to handle the changes in illumination. Take a look at the coffee cup captured in standard and low lighting. The image on the left was photographed with standard overhead lighting while the image on the right was captured with very little lighting and is still able to examine the same cup—but based on the lighting conditions, the cup looks dramatically different (nice how the vertical cardboard seam of the cup is visible in the low lighting conditions, but not the standard lighting).

Continuing down the line there should be an account for background clutter.

Finally, there is intra-class variation. The canonical example of intra-class variation in computer vision is displaying the diversification of chairs. From comfy chairs that are used to curl up and read a book, to chairs that line the kitchen table for family gatherings, to ultra-modern art deco chairs found in prestigious homes, a chair is still a chair—and our image classification algorithms must be able to categorize all these variations correctly.

### 3 Data Augmentation

Data augmentation is the technique which helps to create more datasets from the original datasets without affecting the original datasets by applying random jitters, and the main reason to apply the data augmentation is to increase the generalizability of the model and then it will be able to learn more robust features. This technology did not apply to the testing side. The left image is without the augmentation and right side image with augmentation which proves the generalizability of our networks (Fig. 7).

### 4 Transfer Learning and Fine Tuning

Transfer learning is an art of tune model which is pre-trained on images. The fine-tune method that already trained on the given dataset is applied. Typically, these networks are state-of-the-art architectures such as VGG, ResNet and Inception which are trained on imagenet dataset. In the left is the original VGG16 architecture and in the middle is the final FC layer of the VGG16 which is removed and treated as feature extraction. Right-side image removed the FC layer and added a new FC head (Fig. 8).



```
[INFO] evaluating network...
```

	precision	recall	f1-score	support
Speed limit (20km/h)	0.94	0.98	0.96	60
Speed limit (30km/h)	0.96	0.97	0.97	720
Speed limit (50km/h)	0.95	0.98	0.96	750
Speed limit (60km/h)	0.98	0.92	0.95	450
Speed limit (70km/h)	0.98	0.96	0.97	660
Speed limit (80km/h)	0.92	0.93	0.93	630
End of speed limit (80km/h)	0.96	0.87	0.91	150
Speed limit (100km/h)	0.93	0.94	0.93	450
Speed limit (120km/h)	0.90	0.99	0.94	450
No passing	1.00	0.97	0.98	480
No passing veh over 3.5 tons	1.00	0.96	0.98	660
Right-of-way at intersection	0.95	0.93	0.94	420
Priority road	0.99	0.99	0.99	690
Yield	0.98	0.99	0.99	720
Stop	1.00	1.00	1.00	270
No vehicles	0.99	0.90	0.95	210
Veh > 3.5 tons prohibited	0.97	0.99	0.98	150
No entry	1.00	0.94	0.97	360
General caution	0.98	0.77	0.86	390
Dangerous curve left	0.75	0.60	0.67	60
Dangerous curve right	0.69	1.00	0.81	90
Double curve	0.76	0.80	0.78	90
Bumpy road	0.99	0.78	0.87	120
Slippery road	0.66	0.99	0.79	150
Road narrows on the right	0.80	0.97	0.87	90
Road work	0.94	0.98	0.96	480
Traffic signals	0.87	0.95	0.91	180
Pedestrians	0.46	0.55	0.50	60
Children crossing	0.93	0.94	0.94	150
Bicycles crossing	0.92	0.86	0.89	90
Beware of ice/snow	0.88	0.75	0.81	150
Wild animals crossing	0.98	0.95	0.96	270
End speed + passing limits	0.98	0.98	0.98	60
Turn right ahead	0.97	1.00	0.98	210
Turn left ahead	0.98	1.00	0.99	120
Ahead only	0.99	0.97	0.98	390
Go straight or right	1.00	1.00	1.00	120
Go straight or left	0.92	1.00	0.96	60
Keep right	0.99	1.00	0.99	690
Keep left	0.97	0.96	0.96	90
Roundabout mandatory	0.90	0.99	0.94	90
End of no passing	0.90	1.00	0.94	60
End no passing veh > 3.5 tons	0.91	0.89	0.90	90
accuracy			0.95	12630
macro avg	0.92	0.93	0.92	12630
weighted avg	0.95	0.95	0.95	12630

```
[INFO] serializing network to 'output/trafficsignnet.model'...
```

Fig. 9 Training result

### 5 Result

After the training, 95% accuracy is being got. Training results are shown in Fig. 9. In Fig. 10, showing all prediction results which took 25 random images from test data, 25 images have an accurate label with actual sign images (Figs. 11 and 12).

### 6 Conclusion

With the use of the image classification, a vehicle can easily detect the image and automatically takes appropriate action either stopping and turn to a safer side. Further,



Fig. 10 Prediction result

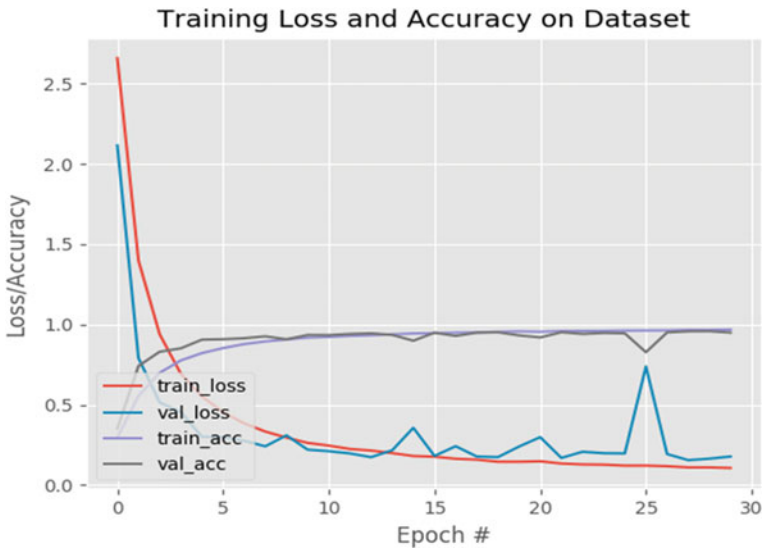
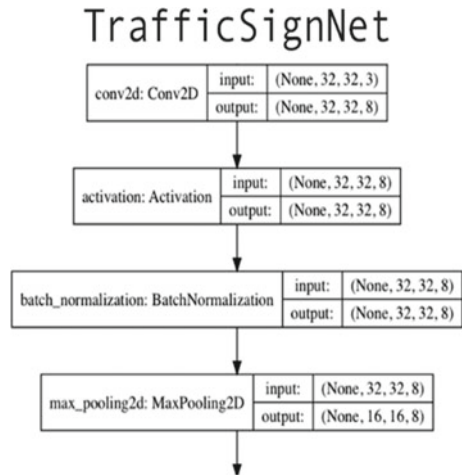


Fig. 11 Training and loss graph

**Fig. 12** Our model of CNN

it collects traffic light information and either halts or starts, to move on automatically. Once the destination information is conveyed, the car halts once it reaches the specified destination point. Hence, this way of automated controlling accidents may be avoided.

## References

1. Rosebrock A PyImageSearch Gurus. <https://pyimagesearch.com/pyimagesearch/>
2. Jyotsna C, Amudha J Deep learning model for image classification. <https://www.amrita.edu/publication/deep-learning-model-image-classification>
3. Sowmya V, Jaiswal D Image classification using convolutional neural networks. <https://www.amrita.edu/publication/image-classification-using-convolutional-neural-networks>
4. Nandagopalan S, Suryanarayan AB, Sudarshan TS, Chandrasekhar D, Manjunath CN SQL based cardio vascular ultrasound image classification. <https://www.ncbi.nlm.nih.gov/pubmed/23819259>
5. Jenson JR, Jungho IM, Hardin P, Jenson RR. [https://sk.sagepub.com/reference/hdbk\\_remotesense/n19.xml](https://sk.sagepub.com/reference/hdbk_remotesense/n19.xml)
6. Xin M, Wang Y (2019) Research on image classification model based on deep convolution neural network. J Image Video Proc 2019:40. <https://doi.org/10.1186/s13640-019-0417-8>
7. Yim J, Ju J, Jung H, Kim J (2015) Image classification using convolutional neural networks with multi-stage feature. In: Kim JH, Yang W, Jo J, Sincak P, Myung H (eds) Robot intelligence technology and applications 3. Advances in intelligent systems and computing, vol 345. Springer, Cham. [https://doi.org/10.1007/978-3-319-16841-8\\_52](https://doi.org/10.1007/978-3-319-16841-8_52)
8. Pandey S, Muthuraman S, Shrivastava A (2018) Data classification using machine learning approach, pp 112–122. [https://doi.org/10.1007/978-3-319-68385-0\\_10](https://doi.org/10.1007/978-3-319-68385-0_10)
9. LeCun Y et al (1998) Efficient backProp. In: Neural networks: tricks of the trade. This book is an outgrowth of a 1996NIPS Wrokshop. Springer-Verlag, London, UK, pp 9–50. ISBN:3-540-65311-2. <http://dl.acm.org/citation.cfm?id=64554.668382>
10. Nalineswari D, Rakesh N (2015) Link budget analysis on various terrains using IEEE 802.16 WIMAX standard for 3.5 GHz frequency. In: 2015 IEEE international conference on electrical, computer and communication technologies (ICECCT), Coimbatore, India

11. Suryanarayanan S, Manmadhan S, Rakesh N (2017) Design and development of real time patient monitoring system with GSM technology. *J Cases Inform Technol* 19:22–36
12. Jmour N, Zaylen S, Abdelkrim A Convolutional neural networks for image classification
13. Loussaief S, Abdelkrim A Machine learning frame work for image classification. <https://ieeexplore.ieee.org/document/7939841>
14. Hassari S, Ejbali R, Zaied M Supervised image classification using deep convolutional wavelet networks
15. Panigrahi S, Nanda A, Swamkar T Deep learning approach for image classification. <https://ieeexplore.ieee.org/document/8588970>

# Deep Domain Adaptation Approach for Classification of Disaster Images



Anuradha Khattar and S. M. K. Quadri

**Abstract** In the last decade, emergency responders, government organizations and disaster response agencies have certainly acknowledged that microblogging platforms such as Twitter and Instagram can be an important source of actionable information at the time of disaster. However, researchers feel that analyzing social media data for disasters using supervised machine learning algorithms is still a challenge. During the first few hours of any disaster when labeled data is not available, the learning process gets much delayed. The first objective of this study is to explore the domain adaptation techniques by making use of labeled data of some previous disaster along with the abundance of unlabeled data that is available for the current disaster. The second objective is to apply these domain adaptation techniques on disaster-related imagery data from the microblogging platforms since imagery data has largely been unexplored as compared to textual content. To achieve these objectives, domain adaptation methods would be applied to classify the images of an ongoing disaster as informative versus non-informative. This study, for the first time, proposes a semi-supervised domain adaptation technique where the classifier is trained on three types of data, labeled data of the previous disaster, unlabeled data of current disaster and a small batch of labeled data of current disaster. Our experiments have been performed on Twitter images corresponding to three disasters. The experimental results show that there is an improvement in the accuracy of the classification model if a small batch of labeled target images is also added along with the unlabeled target images and labeled source images at the time of training. The experiment aims to make the best use of the labeled data of a previous disaster to analyze the current disaster without any delay for better response and recovery.

---

A. Khattar (✉) · S. M. K. Quadri  
Jamia Millia Islamia, New Delhi, India  
e-mail: [anuradha.khattar@mirandahouse.ac.in](mailto:anuradha.khattar@mirandahouse.ac.in)

S. M. K. Quadri  
e-mail: [quadrismk@jmi.ac.in](mailto:quadrismk@jmi.ac.in)

**Keywords** Disasters · Domain adaptation · Deep learning · CNN · Social media · Microblogs · Twitter

## 1 Introduction

The possibility of exploiting social media data for managing disasters was first envisioned when an earthquake of 7.0 magnitudes struck Haiti in 2010. Haitians took to social media platforms (Twitter, Facebook, Blogs, etc.), and text messages were used to share critical situational information. Haiti earthquake could be thought of as the path-breaking event which led to the realization that data generated by microblogged communication could alter the way disaster information has so far been accessed and analyzed. It is observed that it was only after the 2010 earthquake at Haiti and the 2011 earthquake at Tohoku that the researchers started developing and handling systems for disaster data posted on social media platforms which can play an important role in managing disasters [1]. Vieweg et al. [1] were among the first few researchers who studied the communication on Twitter that took place during the Grassfires at Oklahoma in 2009 and the floods at Red River in 2009. Imran et al. [2] discussed methods for automatic extraction of valuable “Information Nuggets” from the tweets posted on Twitter. Saptarshi Ghosh et al. [3] did a study for extracting syntactic low-level features from tweets based on natural language processing, such as parts-of-speech and presence of explicit types of words. Neha Pandey and Natarajan [4] proposed a technique using the semi-supervised concept of machine learning to extract situation awareness information and also created an interactive map that was used to locate the affected areas during a disaster. Nguyen et al. [5] proposed a neural network-based method that uses the data of some previous event when no labeled data for the current event is available and achieves good results. Moumita Basu et al. [6] introduced the concept of automatically matching tweets that are need-based with appropriate availability-based tweets. This may play an important role in the coordination of post-disaster relief operations. Hemant Purohit et al. [7] were the first ones to work on the problem of identifying serviceable social media requests and prioritizing them. They applied logistic regression for developing the automatic classifier and used the Word2Vec toolkit of Google for training the model. Canon et al. [8] took the public opinion from the e-participatory toolkit, Google forms and related articles on the web to study the preparedness of the communities that are prone to disasters. They used bidirectional recurrent neural network (BRNN) to automatically assign the qualitative responses collected into their appropriate categories for better pre-disaster management.

All of the abovementioned research work was mainly based on the text data that is available on social media sites at the time of any crisis. Recently, it is realized that with the advances in the use of smartphones, a lot of imagery content is also available at social media along with the plain text messages [9]. Even within minutes of the onset of a disaster huge amount of multimodal content is posted on various microblogs which if handled immediately can be very helpful for emergency responders. With

the advances in the field of computer vision, processing of the social media imagery data of disasters has taken a new direction.

Researchers Firoj Alam et al. [10] attempted to filter useful actionable information from the images of an ongoing disaster posted on social media by combining human intelligence with the machine. Ferda Ofli et al. [11] fine-tuned the convolution neural network to perform the experiments on four disaster images for damage assessment under event-specific and cross-event-specific settings. Brett Robertson et al. [12] stated that the scale of social media data relevant to studying disasters is only likely to grow. Their study was focused on filtering the noise on social media images and identification of urgent messages and authentic calls for help based on imagery data during a disaster. Xukun Li et al. [13] proposed an approach based on class activation map (CAM) and convolution neural network (CNN) and located damaged buildings in the image of a disaster and also estimated the degree of damage. Firoj Alam et al. [14] showed that textual and imagery content on social media provides complementary information that can be very useful to improve situational awareness. They proposed a methodological approach that combines several natural language processing and computer vision techniques in a unified framework to help humanitarian organizations in their relief operations.

The above studies done on disaster imagery data suggest that convolution neural network (CNN) has brought significant advances to the deep learning techniques applied to images. At the same time, there is no doubt that the availability of labeled data is very crucial for the success of CNN models. However, in critical situations like disasters, it is quite challenging to get the labeling of social media data done as it involves human intervention. For emergency responders, getting the situational information during the early hours of a disaster is very critical, and it has been experienced in the past that the information does not reach them in time to take necessary action. In such situations, it is suggested that a CNN model can leverage labeled data from a previous disaster and unlabeled data from the ongoing disaster, even though the two disasters are quite different in nature and belong to different domains by applying much existing domain adaptation (DA) techniques.

Till now, the application of domain adaptation techniques on image classification is very limited as compared to its application on natural language processing. The objective of this study is to fill this gap and propose a semi-supervised domain adaptation technique for images that are based on the most popular DA method introduced by Ganin et al. [15] called unsupervised deep adversarial neural network (DANN). Three models are built: Model 1-simple transfer learning, Model 2-unsupervised domain adaptation and Model 3-semi-supervised domain adaptation. These models are based on source labeled, target unlabeled and a small batch of target labeled datasets of three disasters. A total of six classification networks are built to answer the following research questions:

RQ1: Performance of domain adaptation framework versus simple transfer learning,

RQ2: Performance of domain adaptation on similar disasters versus different disasters,

RQ3: Performance of unsupervised versus semi-supervised domain adaptation.

To the best of our knowledge, semi-supervised domain adaptation based on DANN is used for the first time to classify the disaster images.

The rest of the paper is structured as follows: In Sect. 2, work done in the field of domain adaptation particularly for disaster management is discussed which has motivated us to dig deeper into this area. Section 3 covers the details of transfer learning, domain adaptation and deep adversarial neural network. In Sect. 4, the data source is given and the experimental setup for developing the proposed models. The results and analysis are presented in Sect. 5. Finally, the conclusion and future work are there in Sect. 6.

## 2 Background and Motivation

Domain adaptation is a desirable approach in disaster response as it enables us to take quick action when a disaster occurs. In the past few years, many studies have applied this approach to disaster data.

Li et al. [16] performed experiments to use the data from a prior disaster that has already been labeled and called it “source disaster”. The ongoing disaster was named as “target disaster”. They built a classifier for the target disaster whose data is not labeled as yet by using the data of the source disaster. They developed two models, the first was based on labeled text data from the source disaster using Naive Bayes classifier and the second was based on domain adaptation Naive Bayes classifier that used labeled source data together with unlabeled target data. Their results showed that the first model performed better if the tasks are similar across the source and target disasters and the second model can be preferred for the tasks that are more related to the target disaster. They collected the data from Twitter API for two disasters: Hurricane Sandy and Boston Marathon bombings.

Caragea et al. [17] proposed a different version of the weighted Naive Bayes DA algorithm introduced by Li et al. [16]. They introduced hard labels with self-training, instead of soft labels to identify disaster-related tweets. They used the dataset CrisisLexT6, consisting of labeled tweets of six disasters. These disasters had occurred in Canada, the USA and Australia between Oct 2012 and July 2013. They then compared their classifier with both the supervised Naive Bayes classifiers learned from source only and EM-based classifiers learned using the original approach proposed by Li [16]. Their results confirmed that domain adaptation approaches are always better than supervised learning with source data only.

Firoj Alam et al. [18] developed a method that can use the labeled and unlabeled data of any past event. The main aim was that there should not be any need for the data of the new event. Their framework consisted of three networks—basic task solving network (a CNN), domain adaptation network based on DANN and a semi-supervised network based on graph embedding. They worked on the datasets of Nepal Earthquake which was held in 2015 and Queensland floods held in 2013. The tweets



were collected from Twitter API, and annotations were done by CrowdFlower. The experiments were performed under three scenarios: only with the semi-supervised component, only with a domain adaptation framework and then as a combination of all the components together. Their results demonstrated significant improvement by combining a semi-supervised approach along with domain adversarial training over their respective baselines.

Firoj Alam et al. [19] observed that the researchers have been focusing mainly on text data from social media to extract useful information from it. They handled this limitation and released a large multimodal dataset of seven different disasters that occurred in 2017. The data was collected from Twitter API.

Xukun Li et al. [20] worked on an image dataset of four disasters published by Nguyen et al. [11] and classified the images as damaged or not damaged. This was the first study in which the technique of domain adaptation was applied to the images from an emergent disaster. Since no labeled data for the current disaster is available, the classifier is trained on the labeled data of a disaster that has already happened in the past.

### **3 Transfer Learning, Domain Adaptation and DANN**

Although deep learning models can solve many complex problems, they need a lot of labeled data for the learning process. To label, a whole lot of data for supervised learning requires time as well as effort. ImageNet is a dataset that has more than 15 million high-resolution labeled images classified into 22,000 classes. Many state-of-the-art models are trained on this dataset which has very high accuracy and is publicly available. But when these models are used for different domains or tasks, they do not perform well. In these situations, transfer learning plays an important role and helps us to use these pre-trained models as the starting point to solve new tasks on different domains.

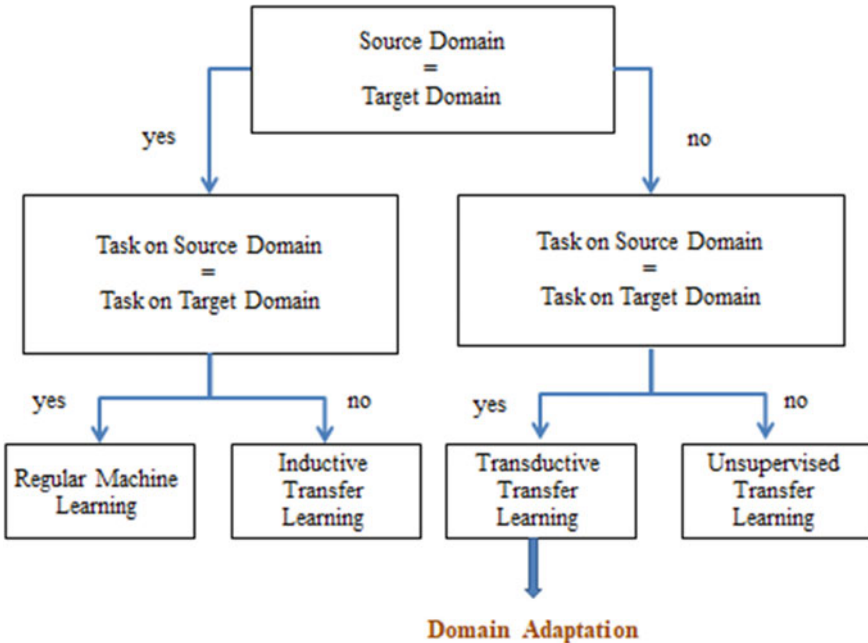
#### ***3.1 Transfer Learning***

“Neural Information Processing Systems (NIPS) 1995” conducted a workshop “*Learning to Learn: Knowledge Consolidation and Transfer in Inductive Systems*” which initiated the interest of researchers in the field of “transfer learning” [21]. Situations where enough data is not available for a model to learn from scratch, transfer learning can play an important role. It allows us to transfer the knowledge in the form of features and weights learned from some previous model (source) to the newer model (target) having insufficient data related to the task. Source and target may belong to different domains having different distributions or are trying to solve different tasks. Transfer learning has a major impact in the field of deep learning especially in image classification as the low-level features of the images can easily

be shared across various tasks [22]. The formal definition of transfer learning as given by Pan et al. [21] is: “transfer learning aims to extract the knowledge from one or more source tasks and applies the knowledge to a target task.” Another reason for the popularity of transfer learning in the area of deep learning is that training a model from scratch in deep learning requires huge resources for large and challenging datasets.

There are different transfer learning strategies and techniques, which can be applied based on the domain, task and availability of data (Fig. 1).

- **Inductive Transfer Learning:** This type of transfer learning deals with the same source and target domain, but the different source and target tasks.
- **Transductive Transfer Learning:** Arnold et al. [23] first used the term transductive transfer learning where the source and target domains are different, but the source and target tasks are the same. Further, there may be two ways in which the domains are different: (a) The feature space of source domain and target domain is different, and (b) the feature space between the domains is the same, but the probability distributions of the input data of two domains are different. In the transductive transfer learning, a lot of labeled data is available for the source domain, but the same is not true for the target domain.
- **Unsupervised Transfer Learning:** Source and target domains, as well as tasks, are different. No labeled data is available for either of the domains.



**Fig. 1** Different types of transfer learning based on whether the source and target domains are the same or not and whether the source and target tasks are the same or not (Adapted from [21])

### 3.2 *Domain Adaptation*

Transfer learning is a general concept where the target task is tried to solve using the knowledge of source task and source domain. As mentioned above in case (b) of transductive transfer learning when the marginal probability distributions between the source and target domains are different, there is an inherent shift in the data distribution of source and target domains that requires tweaks to transfer the learning. This type of transfer learning is called domain adaptation. Thus, in domain adaptation, the model is trained on the source distribution having labeled data and is used to solve the task for a related but different target distribution after bringing the source and target distributions closer. Domain adaptation technique can be further categorized into three cases:

- (1) Supervised DA: Labeled data for the target domain is present, but its size is small which is not sufficient for the learning task.
- (2) Semi-supervised DA: Abundant of unlabeled data and a small amount of labeled data in the target domain is available for the training. Many methods are available which can make use of this data for learning.
- (3) Unsupervised DA: Only sufficient unlabeled data of the target domain is available.

Wang et al. [24] did a survey and presented various domain adaptation techniques to handle the shifts between source and target distributions.

### 3.3 *DANN*

Ganin et al. [15] proposed an unsupervised method for domain adaptation that uses the labeled data of source domain and unlabeled data of target domain to learn a classification model on the source domain and apply this model on the target domain. This method is based on the fact that the probability distributions of the source domain and target domain are not the same and apply some domain transformations to bring them closer. The model is trained on the transformed source domain, which certainly gives good results on the target domain as both the domains are now closer to each other. The classifier predictions are based on the features that are not able to discriminate between the source domain and the target domain. The two jobs, feature extraction and domain adaptation, are combined in a single training process.

The DANN architecture includes:

- (a) Feature Extractor: Transformations are performed on the source and target domains to bring them closer, and simultaneously, features are learned on these transformed domains.

- (b) Label Classifier: A neural network to learn the classification on the labeled transformed source domain.
- (c) Domain Classifier: Train a network that can predict if the output from (a) belongs to the source domain or the target domain.

## 4 Experimental Setup

Our experiments have been performed on the disaster images of Hurricane Harvey (August 2017) as the source disaster. Images of Hurricane Irma (September 2017) and California wildfires (October 2017) are used as the target images. Table 1 shows the total number of images for each disaster with the number of images in each category, informative versus non-informative. The objective of choosing these datasets is as follows:

- (a) Due to the chronological order of the occurrence of these three disasters.
- (b) To see the performance of proposed models on similar (Hurricane Harvey and Hurricane Irma) and different (Hurricane Harvey and California wildfires) disasters.

The datasets of these disasters are downloaded from publicly available “Artificial Intelligence for Disaster Response” (AIDR) system [25]. AIDR is an open-source platform that collects messages from Twitter at the onset of any disaster and classifies them. It is designed and developed at the Qatar Computing Research Institute (QCRI). Many emergency response departments including United Nations use the resources that are available at AIDR. A large number of resources can directly be downloaded from (<http://aidr.qcri.org>) [25].

The following three classification models are proposed to build:

**Table 1** The total number of images in three disaster datasets with the number of images in each category: informative versus non-informative

Class	Hurricane Harvey	Hurricane Irma	California wildfires
Informative	2449	2203	399
Non-informative	1947	2262	193
Total	4396	4465	592

- Model 1:** A network that is trained on the:
  - Labeled data of a previous disaster (Source)
- Model 2:** A network that is trained on the:
  - Labeled data of previous disaster (Source)
  - Unlabeled data of the current disaster (Target)
- Model 3:** A network that is trained on the:
  - Labeled data of some previous disaster (Source)
  - Unlabeled data of the current disaster (Target)
  - A small batch of labeled data of the current disaster (Target)

Simple Transfer Learning

Domain Adaptation (Unsupervised).

Domain Adaptation (Semi-supervised)

The experiments are performed according to the following steps:

**STEP-0: Off-the-shelf pre-trained model VGG-19 as a feature extractor for Model 1, Model 2 and Model 3**

Pre-trained VGG-19 is used to initialize the parameters of the three proposed network models for all the layers except the last fully connected layer which has a different dimension.

VGG-19 is based on the database ImageNet consisting of around more than a million images which are categorized into 1000 different categories [26]. The model has 16 convolution layers followed by ReLU nonlinear activation and five Maxpooling layers. Out of the last three layers, two are fully connected layers with dimension 4096. The last layer is the softmax layer having a dimension of 1000 units. The model is being used quite frequently as a feature extractor for some other network for new images. In the last fully connected layer of this pre-trained network, we changed 1000 to 2, since we had only two classes namely ‘informative’ versus ‘non-informative’ (Fig. 2).

**STEP-1: Simple transfer learning to build Model 1**

Fine-tune the parameters of VGG-19 of Step-0 based on the source dataset of Hurricane Harvey. Source labeled images (Hurricane Harvey) are used as the training set to learn a supervised classifier. The resultant classifier is tested on two target disasters (Hurricane Irma and California wildfires). Model 1 classifier serves as the baseline or lower bound for the other two models that are based on adversarial domain adaptation (Fig. 3).

**STEP-2: DANN (Unsupervised) to build Model 2**

Fine-tune all the parameters of VGG-19 of Step-0 based on the source labeled images (Hurricane Harvey) and on target unlabeled images (Hurricane Irma/California wildfires) together to train a deep domain adaptation classifier for the target. The technique used for domain adaptation is DANN, which is already explained in Sect. 3.3. The resultant classifier is used to test the target images (Hurricane Irma/California wildfires) and the accuracy is noted (Fig. 4).

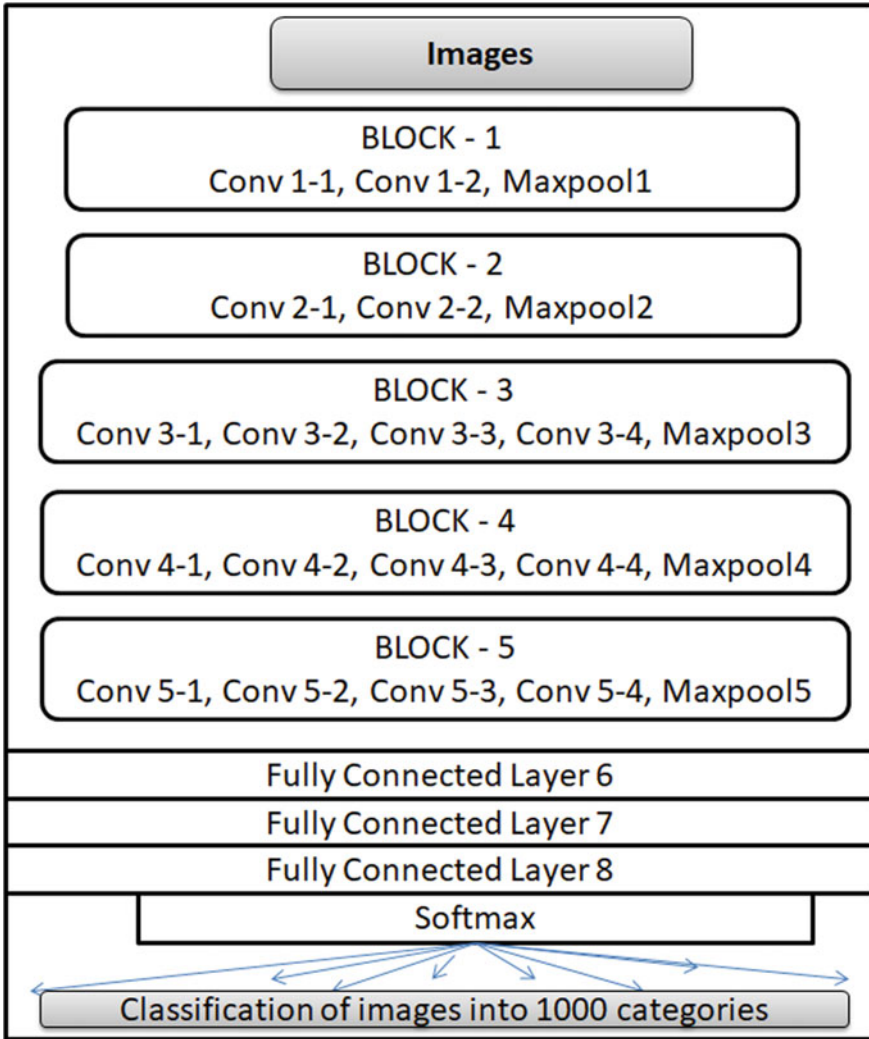


Fig. 2 VGG-19 architecture [26]

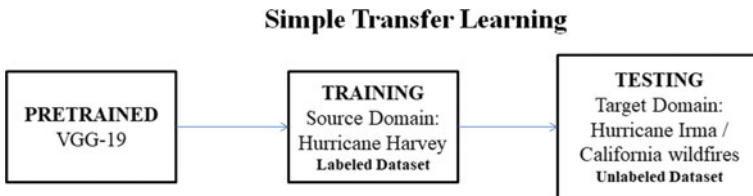
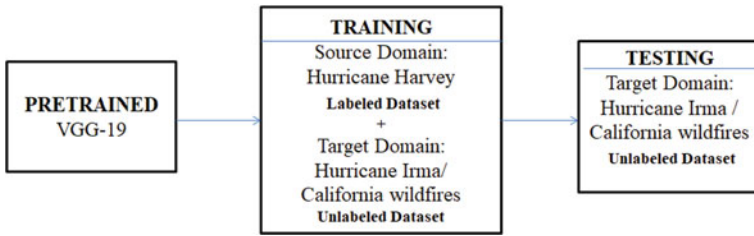


Fig. 3 Model 1-simple transfer learning-source: Hurricane Harvey (labeled dataset), target: Hurricane Irma/California wildfires (unlabeled dataset)

### Unsupervised Domain Adaptation



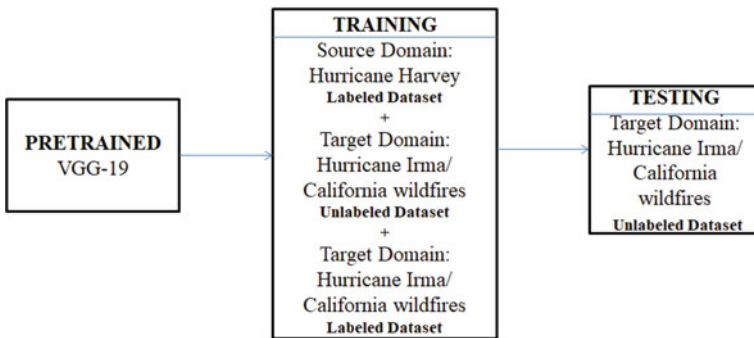
**Fig. 4** Model 2-unsupervised domain adaptation source: Hurricane Harvey (labeled dataset) + Hurricane Irma/California wildfires (unlabeled dataset), target: Hurricane Irma/California wildfires (unlabeled dataset)

### STEP-3: DANN (Semi-supervised) to build Model 3

Fine-tune all the parameters of VGG-19 of Step-0 based on the source labeled images (Hurricane Harvey), target unlabeled images (Hurricane Irma/California wildfires) and a small batch of target labeled images (Hurricane Irma/California wildfires) together to train a deep domain adaptation classifier for the target by using DANN. The resultant classifier is used to test the target images (Hurricane Irma/California wildfires) and the accuracy is noted (Fig. 5).

This work has used PyTorch (<https://pytorch.org/>) as a deep learning framework run on Google Colab (<https://colab.research.google.com/>) that provides a single 12 GB NVIDIA Tesla K80 GPU. Optimizer used in stochastic gradient descent (SGD), learning rate =  $1.0e-3$ , momentum = 0.9 and weight decay =  $1.0e-3$ . The decay learning rate also a factor of 0.1 after 50 and 75 epochs is also used.

### Semi-supervised Domain Adaptation



**Fig. 5** Model 3-semi-supervised domain adaptation source: Hurricane Harvey (labeled dataset) + Hurricane Irma/California wildfires (unlabeled dataset) + Hurricane Irma/California wildfires (a small batch of the labeled dataset), target: Hurricane Irma/California wildfires (unlabeled dataset)

## 5 Results and Discussion

Three models have been built based on simple transfer learning, unsupervised domain adaptation and semi-supervised domain adaptation. Each model uses Hurricane Harvey as the labeled source dataset. Within each model, two networks are built for two target datasets: Hurricane Irma and California wildfires. The accuracy, precision, recall and F1-score of all six networks are tabulated in Table 2.

Based on the quantitative results, the research questions raised in Sect. 1 are re-stated and answer each of them in turn:

RQ1: Performance of domain adaptation framework versus simple transfer learning.

- There is up to 9.5% absolute improvement in F1-score when the domain adaptation framework is applied instead of simple transfer learning.

RQ2: Performance of domain adaptation on similar disasters versus different disasters.

**Table 2** Experiment results: Model-1 uses source labeled data, Model-2 uses source labeled and target unlabeled data and Model-3 uses source labeled, target unlabeled and a small batch of target labeled data

Method applied	Source disaster	Target disaster	Accuracy	Precision	Recall	F1-score	
Model-1	Simple transfer learning Source: Labeled Target: Unlabeled	Hurricane harvey	Hurricane Irma	76.11	76.23	76.07	76.06
			California wildfires	68.55	70.1	71.14	68.4
Model-2	Domain adaptation (Unsupervised DANN) Source: Labeled Target: Unlabeled	Hurricane harvey	Hurricane Irma	76.78	76.86	76.82	76.78
			California wildfires	79.24	77.9	78.1	77.98
Model-3	Domain adaptation (Semi-supervised DANN) Source: Labeled Target: Unlabeled + Labeled	Hurricane harvey	Hurricane Irma	77.901	77.88	77.97	77.9
			California wildfires	79.87	79.11	77.27	77.9

For each model, source dataset: Hurricane Harvey and target datasets: Hurricane Irma and California wildfires, this results in a total of six networks. Quantitative results show that Model-3 performs better than Model-1 and Model-2



- There is up to 1.84% absolute improvement in F1-score when the domain adaptation framework is applied to similar disasters (Hurricane–Hurricane) instead of simple transfer learning.
- There is up to 9.5% absolute improvement in F1-score when domain adaptation framework is applied to different disasters (Hurricane–wildfires) instead of simple transfer learning.

RQ3: Performance of unsupervised versus semi-supervised domain adaptation.

- There is up to 1.12% absolute improvement in F1-score when the proposed semi-supervised domain adaptation framework is applied instead of unsupervised domain adaptation.

## 6 Conclusion

Research studies have shown the significance of the social media data generated at the time of the disaster and its contribution in situation awareness and other relief activities conducted by emergency responders. At the same time, researchers have also realized the challenges in handling this high-velocity data stream due to the lack of availability of tools for efficient analysis. Most of the existing methods used for classification need a large amount of annotated data specific to the particular event for training the model to achieve good results. In this study, the possibility of overcoming this limitation is tried to explore by presenting three models: simple transfer learning from the source domain to target domain (Model 1), unsupervised DA approach (Model 2) and semi-supervised DA approach (Model 3). The results obtained from these models indicate that the performance of domain adaptation models (Model 2, Model 3) is better than the simple transfer learning model (Model 1) and also that the performance of the domain adaptation model improves if the two disasters are different. Among the unsupervised and semi-supervised DA models, semi-supervised models (Model 3) perform better than unsupervised DA (Model 2). Thus, if a small batch of labeled target data is added to the labeled source data and unlabeled target data at the time of learning, the accuracy of the classification model can be definitely improved.

Overall, our experiments suggest that source data from a prior disaster can be used to build classification models for the ongoing target disaster, especially if the tasks to be performed are the same across the two disasters. Furthermore, using source labeled data together with target unlabeled and a small batch of target labeled data in a semi-supervised domain adaptation framework has the potential to produce better classifiers.

As part of future work, other domain adaptation approaches are planned to study that may be applied to images of various disasters for efficient disaster response.

## References

1. Vieweg S, Hughes AL, Starbird K, Palen L (2010) Microblogging during two natural hazards events. In: Proceedings of the 28th international conference on human factors in computing systems, CHI 2010, p 1079. <https://doi.org/10.1145/1753326.1753486>
2. Imran, M, Elbassuoni, S, Castillo, C, Diaz, F, Meier, P (2013) Extracting information nuggets from disaster—related messages in social media. In: ISCRAM 2013 10th international conference on information systems for crisis response and management, pp 1–10
3. Rudra K, Ghosh S, Ganguly N, Goyal P, Ghosh S (2015) Extracting situational information from microblogs during disaster events. In: 7th international conference on communication systems and networks (COMSNETS), pp 583–592. <https://doi.org/10.1145/2806416.2806485>
4. Pandey MN, Natarajan S (2016) How social media can contribute during disaster events? In: 2016 international conference on advances in computing, communications and informatics, pp 1352–1356. <https://doi.org/10.1109/ICACCI.2016.7732236>
5. Nguyen DT, Al Mannai KA, Joty S, Sajjad H, Imran M, Mitra P (2017) Robust classification of crisis-related data on social networks using convolutional neural networks. In: Proceedings of the 11th international conference on web and social media, ICWSM 2017. AAAI Press, pp 632–635
6. Basu M, Shandilya A, Ghosh K, Ghosh S (2018) Automatic matching of resource needs and availabilities in microblogs for post-disaster relief. Association for Computing Machinery (ACM), pp 25–26. <https://doi.org/10.1145/3184558.3186911>
7. Purohit H, Castillo C, Imran M, Pandev R (2018) Social-EOC: serviceability model to rank social media requests for emergency operation centers. Proceedings of 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 119–126. <https://doi.org/10.1109/ASONAM.2018.8508709>
8. Canon MJ, Satuito A, Sy C (2019) Determining disaster risk management priorities through a neural network-based text classifier. In: Proceedings of 2018 international symposium on computer, consumer and control (IS3C 2018), pp 237–241. <https://doi.org/10.1109/IS3C.2018.00067>
9. Khattar A, Quadri SMK (2020) Emerging role of artificial intelligence for disaster management based on microblogged communication. In: International conference on innovative computing and communication (ICICC 2020). <https://dx.doi.org/10.2139/ssrn.3562973>
10. Alam F, Offi F, Imran M (2018) Processing social media images by combining human and machine computing during crises. *Int J Hum Comput Interact* 34:311–327. <https://doi.org/10.1080/10447318.2018.1427831>
11. Nguyen DT, Offi F, Imran M, Mitra P (2017) Damage assessment from social media imagery data during disasters. In: Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2017. Association for Computing Machinery, Inc, pp 569–576. <https://doi.org/10.1145/3110025.3110109>
12. Robertson BW, Johnson M, Murthy D, Smith WR, Stephens KK (2019) Using a combination of human insights and ‘deep learning’ for real-time disaster communication. *Prog Disaster Sci* 2:100030. <https://doi.org/10.1016/j.pdisas.2019.100030>
13. Li X, Caragea D, Zhang H, Imran M (2019) Localizing and quantifying infrastructure damage using class activation mapping approaches. *Soc Netw Anal Min* 9:1–15. <https://doi.org/10.1007/s13278-019-0588-4>
14. Alam F, Offi F, Imran M (2019) Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria. *Behav Inf Technol* 1–31. <https://doi.org/10.1080/0144929X.2019.1610908>
15. Ganin Y, Lempitsky V (2015) Unsupervised domain adaptation by backpropagation. In: 32nd international conference on machine learning (ICML 2015), vol 2, pp 1180–1189
16. Li H, Guevara N, Herndon N, Caragea D, Neppalli K, Caragea C, Squicciarini A, Tapia AH (2015) Twitter mining for disaster response: a domain adaptation approach. In: ISCRAM 2015 conference proceedings—12th international conference on information systems for crisis response and management, January 2015

17. Li H, Caragea D, Caragea C, Herndon N (2018) Disaster response aided by tweet classification with a domain adaptation approach. *J Contingencies Cris Manag* 26:16–27. <https://doi.org/10.1111/1468-5973.12194>
18. Alam F, Joty S, Imran M (2018) Domain adaptation with adversarial training and graph embeddings. In: *ACL 2018—56th annual meeting of the association for computational linguistics proceedings conferences*, Long Pap. 1, pp 1077–1087. <https://doi.org/10.18653/v1/p18-1099>
19. Alam F, Ofli F, Imran M CrisisMMD: multimodal twitter datasets from natural disasters
20. Li X, Caragea C, Caragea D, Imran M, Ofli F (2019) Identifying disaster damage images using a domain adaptation approach. In: *Proceedings of International ISCRAM Conference*, May 2019, pp 633–645
21. Pan SJ, Yang Q (2010) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22:1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
22. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? *Adv Neural Inf Process Syst* 4:3320–3328
23. Arnold A, Nallapati R, Cohen WW (2007) A comparative study of methods for transductive transfer learning. In: *Proceedings of IEEE international conference on data mining, ICDM*, pp 77–82. <https://doi.org/10.1109/ICDMW.2007.109>
24. Wang M, Deng W (2018) Deep visual domain adaptation: a survey. *Neurocomputing* 312:135–153. <https://doi.org/10.1016/j.neucom.2018.05.083>
25. Imran M, Castillo C, Lucas J, Meier P, Vieweg S (2014) AIDR: Artificial intelligence for disaster response. In: *WWW 2014 Companion—proceedings of the 23rd international conference on world wide web*. Association for Computing Machinery, Inc, pp 159–162. <https://doi.org/10.1145/2567948.2577034>
26. Simonyan, K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: *3rd international conference on learning representations, iclr 2015—conference track proceedings*, pp 1–14

# Automated Intelligent IoT-Based Traffic Lights in Transport Management System



Akhilesh Kumar Singh and Manish Raj

**Abstract** Transport plays a crucial and vital role in the living standard of a nation. The transportation regulatory authorities across the world are continuously facing issues in traffic congestion, severe increase in pollution level and rise in the impurity level of air quality index. To overcome these problems, there is always a demand of intelligent strategies in transport management through automated software and IoT-enabled mechanisms. Due to cost-effective ration and lack of awareness, these technologies are not fully adopted in transportation industries. The primary aim of this research is to highlight prior researches in transport management and identify the issues. Further, the issues are reduced through the proposed model which improves the performance and reduces congestion so that environmental impact and productivity could be increased.

**Keywords** Automated intelligent transport systems (AITS) · IoT · Roadside unit (RSU) · Route guidance system

## 1 Introduction

Throughout the world, transportation industries are suffering from strategic issues like worst traffic congestion, improper infrastructures, increased pollution, air quality index and increased emissions. Out of these issues, traffic congestion is one of the basic problems in traffic management. Congestion is directly related to the socio-economic and financial growth in transport industries. One of the major reasons for traffic congestion is vast increase in volume of vehicles on the road, lack of updated and modern road infrastructure and tradition non-automated non-smart traffic signals.

---

A. K. Singh (✉) · M. Raj  
GLA University, Mathura, India  
e-mail: [akhileshkr.singh@gla.ac.in](mailto:akhileshkr.singh@gla.ac.in)

M. Raj  
e-mail: [manish.raj@gla.ac.in](mailto:manish.raj@gla.ac.in)

Tons of fuel are wasted on rush hour traffic in major metro cities due to this congestion. Hence, an efficient solution for traffic congestion is required using automated, and intelligence IoT is essential at present for transportation systems. These mechanisms may include real-time traffic info systems, in-car navigation routing, integration of vehicle and infrastructure, integration of vehicle to vehicle, adaptive dynamic traffic signal control, automated meter service, online toll and tax submission, congestion pricing, automated fee-based express lanes, vehicle usage-based mileage fees and vehicle accident and breakdown avoidance technologies.

Major classification of automated intelligent IoT-based transport management system includes as follows:

- Smart geographic information system-based traffic congestion system,
- Smart traffic modeling and traffic light control system,
- Density-based traffic movement system,
- Route guidance system,
- Safety, planning, simulation and management of traffic,
- Social economy and optimization for transport management system,
- IoT-based intelligent transport management system,
- Traveler and passenger information system.

The performance and efficiency of automated intelligent IoT-based transport management system are of vital importance for commercial growth. Hence, the fundamental objectives of automated intelligent IoT [7, 8] based transport management system may include enhanced and improved traffic safety, efficient utilization of transportation services, reducing air pollution level and air quality index thereby increasing energy efficiency and improving and adding smart services to get rid of traffic congestion. This research work analyzes the issues for promotion and development of transport management system by acquiring automated and intelligent IoT-based services.

## 2 Literature Review

Recent innovations in traffic control system have its own advantages as well as shortcomings. Some of the methodologies proposed by different authors are analyzed in this section to obtain the issues as a summary. A framework to deal on vehicle rush on road and minimizing vehicular interference and hurdles is reported in literature [1] to obtain a non-congested travel path for service required conveyance such as emergency vehicles and priority vehicles. The framework proposed by the author includes monitoring of traffic control signals with the help of Raspberry Pi, Node MCU, RFID Tag and Reader for efficient transportation system. Similarly, intelligent transport system is reported in literature [3] which used graph theory to obtain safe and less congested route to the drivers. Vienna development method specification language is used in the proposed design to obtain efficient route description process.

In [4], a model for automated intelligent traffic management system using Internet of Things is proposed with the help of roadside unit (RSU), and friction monitoring environmental sensors, database and mobile interface are used in the proposed design to achieve better performance. Proposed design provides near future traffic signal conditions on traveling route, and the acceleration speed of the motor is monitored and maintained for bypassing signal lights on junction without waiting for the reply. The proposed design has better performance in terms of reduced fuel consumption and improved efficiency.

Real-time vehicular rush query for intelligent transport system was proposed in literature [5] using Internet of Things. The proposed design uses IoT, TOPSIS and PEPA to analyze the congestion and provides better solution to the user to reduce the fuel consumption and increased efficiency.

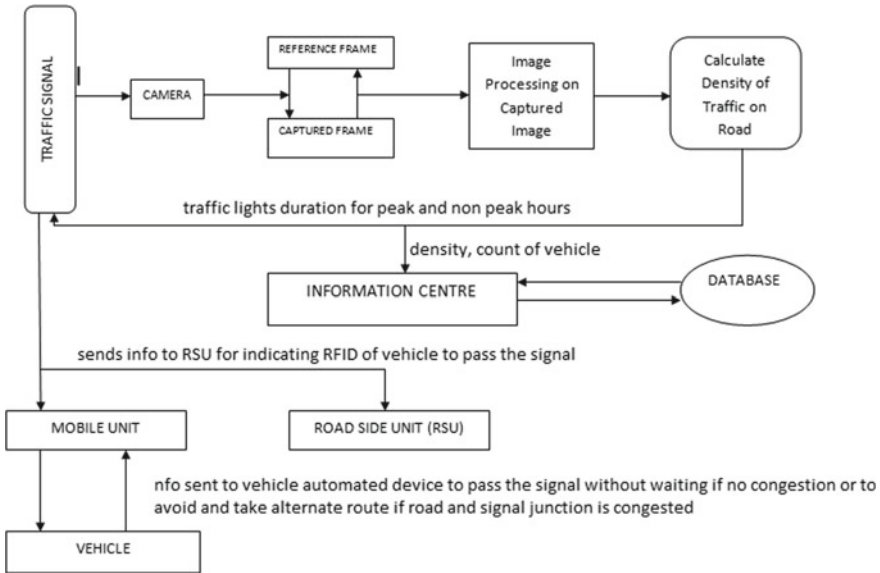
Internet of Things and image processing-based smart parking system was proposed in literature [6] with the help of roadside unit (RSU), thereby reducing the chances of vehicle theft.

Intelligent motor monitoring framework is discussed in literature [9] by implementing IoT with RFID, GPS and cloud computing to control fatalities to obtain a better solution for traffic congestion and fuel performance in traffic management. In [10], secured traffic management system for emergency vehicle using IoT is proposed and implemented for analyzing the necessity of smart systems in traffic management. In [12, 13], author presented a strategic mechanism for dynamic route determination with the help of geographic information system which is proposed and implemented for finding shortest route, alternate route and facility-based route using various heuristic approaches.

Routing is foremost important component of an automated intelligent IoT-based transport management system. It provides infrastructure condition such as current road congestion, current traffic statistics and trip suggestion and recommendation to the driver in order to provide optimal traveling decisions. It can be used as an automated system through centralized traffic control center as centralized approach to evaluate the vehicle performance. The route guidance is mainly categorized into bifurcation, descriptive or prescriptive, static or dynamic, reactive or predictive and system optimal or user optimal [15].

### 3 Proposed Methodology

Out of the several transportation issues mentioned above, the architecture of proposed model for congestion-free transit of traffic is shown in Fig. 1. The model consists of Roadside Module, RFID (RFID Tag used here is highly integrated RFID card reader, and it works on non-contact communication. RFID Reader scans the RFID tag attached to vehicle), Camera, Mobile Unit, Global Positioning System (GPS), Database and Image Processing concepts. The roadside unit is equipped to ease the exchange of information and conversation within the vehicles, transportation servers and other vehicle equipped devices by transiting data over roadside unit according



**Fig. 1** Proposed methodology

to the standards of industries. The roadside unit is fitted in every signal light of all four-way junctions that are equipped with microcontroller and camera installed on it. It also does the quick scan with respect to the transportation particular algorithm of the microcontroller and revert the scanned and asked data to the database server. More scanning of RFID indicates the increasing graph of traffic congestion on the road.

An image processing technique which mainly focuses on the four-way junction captures a reference frame [10], and based on some algorithmic concept, it calculates density of road and traffic junction, thereby signaling traffic lights and mobile unit of in-car automated device to whether pass from signal or had to take an alternate route also based on the percentage of density, the passing light, i.e., green light is signaled and controlled.

The actual traffic will also be monitored by the help of Global Positioning System (GPS). With the help of GPS, real-time information can be calculated, and drivers can be informed whether to take acceleration to jump the traffic signal or to slow down to prevent signals from overwhelming due to congestion.

So the proposed framework not only provides the detailed info about traffic rush to traffic signals, but also informs the drivers with the help of mobile unit installed within car to whether or not cross traffic junction by accelerating or to take an alternate route.

## 4 Conclusion

Throughout the world, transportation industries are facing various issues while implementing traffic management systems. Thus, it is an era for the nation to seriously look toward an automated intelligent IoT-based transport management system which will help to minimize the traffic congestion and other issues.

A new framework for minimizing traffic congestion is proposed and discussed in this paper. The effective and proper implementation of the proposed framework allows congestion-free transit of traffic that can bring a drastic change in the field of transport management system. In this paper, recent technologies of IoT and image processing are integrated so that fuel consumption and pollution can be minimized. The proposed framework furthermore informs the drivers to cross a particular traffic junction and its details. Further, this research could be improved by considering fuel management through hybrid systems.

## References

1. Singh AK, Raj M, Sharma V (2020) Architecture, issues and challenges in monitoring based on IoT for smarter environment. In: 2020 fourth international conference on computing methodologies and communication (ICCMC), Erode, India, pp 142–146. <https://doi.org/10.1109/iccm48092.2020.iccmc-00029>
2. Intelligent transport systems: synthesis report on ITS including issues and challenges in India Department of Civil Engineering, IIT Madras
3. Western JL, Ran B (2000) Information technology in transportation: key issues and a look forward. Western JL-Wisconsin Department of Transportation. Ran B-University of Wisconsin, Madison
4. Brian MAK, Hong KLO (2005) passenger route guidance system for multimodal transit networks. The Hong Kong University of Science and Technology, Hongkong
5. Herbert W, Mili F (2008) Route guidance: state of art vs. state of practice. In: IEEE intelligent vehicles symposium, Eindhoven University of Technology
6. Edwards A, Mackaness W (2003) Intelligent road network simplification in urban areas. The University of Edinburgh
7. Ertl G (1996) Optimierung und Kontrolle, shortest path calculations in large road networks. Project in Discrete Optimization, Karl-Franzens-Universität Graz
8. Singh SP, Kumar V, Singh AK (2019) A Survey on Internet of Things (IoT): layer specific vs. domain specific architecture. In: Proceedings of the 2nd international conference on computer networks and inventive communication technologies (ICCNCT2019) at RVS Technical Campus Coimbatore, India
9. Hart PE, Nilson NJ (1968) A formal basis of the heuristic determination of minimum cost paths. *IEEE Trans Syst Sci Cybern* 4(2):100–107
10. Jagadish GR, Srikanthan T (2008) Route computation in large road networks: a hierarchical approach. Nanyang Technology University Singapore
11. Pearsons J (1998) Heuristic search in route finding. Master's Thesis, University of Auckland, New Zealand
12. Samet H, Sankarnarayanan J (2008) Scalable network distance browsing in spatial databases. Institute for Advanced Computer Studies, University of Maryland, Maryland
13. Timpf S, Volta GS (1992) A conceptual model of wayfinding using multiple levels of abstraction. In: Theories and methods of spatio-temporal reasoning in geographic space. Springer, Berlin, pp 348–367



14. Sharma V, Saxena HK, Singh AK (2020) Docker for multi-containers web application. In: 2020 2nd international conference on innovative mechanisms for industry applications (ICIMIA), Bangalore, India, pp 589–592. <https://doi.org/10.1109/icimia48430.2020.9074925>
15. Deepak Mangal AKSJPS (2020) A comparative study of functional and non functional parameter on advanced iot deployment tools. *Int J Adv Sci Technol* 29(1):1582–1591. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/3726>

# A Document Clustering Approach Using Shared Nearest Neighbour Affinity, TF-IDF and Angular Similarity



Mausumi Goswami

**Abstract** Quantum of data is increasing in an exponential order. Clustering is a major task in many text mining applications. Organizing text documents automatically, extracting topics from documents, retrieval of information and information filtering are considered as the applications of clustering. This task reveals identical patterns from a collection of documents. Understanding of the documents, representation of them and categorization of documents require various techniques. Text clustering process requires both natural language processing and machine learning techniques. An unsupervised spatial pattern identification approach is proposed for text data. A new algorithm for finding coherent patterns from a huge collection of text data is proposed, which is based on the shared nearest neighbour. The implementation followed by validation confirms that the proposed algorithm can cluster the text data for the identification of coherent patterns. The results are visualized using a graph. The results show the methodology works well for different text datasets.

**Keywords** Document classification · Density-based · Clustering · Unstructured data email documents · Similarity · DBSCAN · SNN

## 1 Introduction

Clustering is a major task in many text mining applications. A new algorithm is proposed which is based on a shared nearest neighbour, and this approach of clustering does not require the awareness of the number of clusters in advance. It is capable to find the natural number of clusters. The space complexity and computational complexity are improved by using sparse algorithm. In the literature, many density-based clustering techniques [1–9] are applied in document categorization to find coherent document. There are many clustering algorithms which identify

---

M. Goswami (✉)  
CHRIST (Deemed to be University), Bengaluru, India  
e-mail: [mausumi.goswami@christuniversity.in](mailto:mausumi.goswami@christuniversity.in)

coherent documents based on density-based approach. This work is focused on investigating the limitations of the existing algorithm. This work emphasizes a density-based (DB) technique to cluster documents. The comparison of the technique with other clustering technique has shown satisfactory results.

In straightforward terms, clustering [10–15] implies gathering a set of similar [16] items or objects in a cluster or group with the end goal that the items in a cluster are more similar to each other than the items in some other cluster. It is also called unsupervised learning [16–30]. The main applications are in search engines, medicine, biology and marketing. One among the various approaches of clustering is K-means algorithm which falls under partition category. Another approach is hierarchical. This algorithm gathers similar items into groups or clusters. The nonlinear complexity is given as  $O(n)^2$ . Finally, there is a set of clusters in which each cluster is distinct from the other cluster, but the items in a cluster are similar to each other.

The mathematical definition for the representation of clustering algorithms is: Given a set of parameters  $Y = \{Y_1, Y_2, Y_3, \dots, Y_i, \dots, Y_m\}$ , where  $Y_j = \{y_{i1}, y_{i2}, \dots, y_{id}\}^P \in R^T$  where  $y_{mn}$  is a characteristic feature. A crisp or non-soft attempt of grouping objects attempts to discover m-partitions of  $Y$ ,  $A = \{A_1, A_2, \dots, A_m\} \forall m \leq k$  such that  $A_j \neq \emptyset$ , where  $j = 1, 2, \dots, n$ ;  $\bigcup_{j=1}^m A_j = Y$ ;  $A_i \cap A_j = \phi$ ;  $i, j = 1, \dots, m$  and  $i \neq j$ .

One more approach to grouping is oriented on density or spread of objects and interrelation among them. Density is measured as a function. DBSCAN is considered as one of the most significant algorithms in this area. The paper follows the flow as mentioned below: Sect. 2 emphasizes on the related work done. Section 3 explores a comparative analysis of nonlinear data clustering algorithm. Section 4 puts emphasis on the steps of the algorithm. Section 5 discusses experiments and results.

## 2 Related Work

In this section, an attempt towards the study of different density-based clustering algorithms and its variations is done. This work is also extended to propose a new technique for clustering. E. Bici proposed a method to collect and link regions based on their spread. An upper limit is set by the cluster centre. Birant proposed ST-DBSCAN which is based on DBSCAN and uses two input parameters Eps1 and Eps2. In this work, both temporal and spatial domains are considered. Density-differentiated spatial clustering (DDSC) proposed by Borah and Bhattacharya is an algorithm that is an extension of DBSCAN. DBSCAN cannot find coherent groups of different density in the database containing noise. DDSC can detect clusters that differ in densities.  $O(n * \log n)$  is the computational complexity. Grid-based DBSCAN (GRPDBSCAN) proposed by Darong and Peng is an algorithm mainly used for gigantic data. Various density-based clustering algorithms over nonlinear dataset are as follows:

- **DBSCAN:** It is considered good for the arbitrary shape of group identification. Two parameters  $\epsilon$  and MinPoints are essential.  $\epsilon$  stand for the minimum inter-object distance. The neighbourhood is defined using  $\epsilon$ . If the value of the  $\epsilon$  is small, then many points would not be clustered. On the contrary, if the value is large, then a large number of points will be clustered into the same cluster. The value of  $\epsilon$  should be chosen based on the distance in the dataset. Here, MinPoint defines a dense region of interest. If the parameter is set to 4, then at least 4 points are required for the formation of the dense region. At the beginning, an arbitrary point is selected as a beginning point. Next, it identifies its  $\epsilon$  neighbourhood. If Minpoints is greater than the count of points in the  $\epsilon$  neighbourhood, then the point is considered as noise point; otherwise, the point would have been considered as a dense point. Repeat the step till no more expansion possible. If it is stabilized, then the partition with that cluster is concluded and no new points are added in successive iterations. The clustering process ends when each and every point is assigned to either of the two categories. The first category is called a cluster, and the second category is called noise.
- **DENCLUE:** It is based on the calculation of density defined by using kernels. The algorithm has two operating stages. Pre-clustering step emphasizes the map construction.
- **DBCLASD:** Even in the absence of input argument on the number of clusters, arbitrary shaped clusters can be identified.
- **OPTICS:** It is an indirect method that does not produce any conventional dataset.
- **BRIDGE:** It is a hybrid method. It is faster and cost-effective compared to many other methods.
- **CUBIN:** This clustering technique is used to discover flexible extent (size is varying) and “non-spherical” clusters. It has linear complexity. Erosion technique is applied first, and next nearest neighbouring strategy is applied. Lastly, clusters are formed.
- **ST-DBSCAN:** It is suitable for applications involving spatial and temporal dataset such as weather forecasting and medical imaging.
- **VDBSCAN:** It stands for “varied density-based spatial clustering of application with noise”.
- **DVBSCAN:** Few parameters used in this algorithm are threshold values, radius and minimum objects. The algorithm calculates growing cluster density mean and variance for any object.

### **3 Comparative Analysis of Nonlinear Data Clustering Algorithm**

In this section, a comparative analysis of different nonlinear clustering approaches is demonstrated along with their pros and cons.

Algorithm	Time complexity	Pros and cons
DBSCAN	$O(n^2)$ , $O(n \log n)$ for spatial indexed data	<ul style="list-style-type: none"> <li>• Any shape of a coherent group can be identified</li> <li>• Natural number of clusters may be computed</li> <li>• To determine the parameter values is difficult</li> <li>• In the case of a variation in density, noise points can be detected</li> </ul>
DVBSCAN	$O(n^2)$	<ul style="list-style-type: none"> <li>• Handles local density within the cluster</li> <li>• Outperforms DBSCAN in case of local density</li> </ul>
OPTICS	$O(n^2)$ , $O(n \log n)$ for spatial indexed data	<ul style="list-style-type: none"> <li>• Density threshold value can be determined without user's intervention</li> <li>• Less sensitivity to data with error</li> </ul>
ST-DBSCAN	$O(n^2)$	<ul style="list-style-type: none"> <li>• Clustering of temporal-spatial data can be done on the basis of all the attributes</li> <li>• Parameters are not generated automatically</li> </ul>
DBCLASD	$O(n^2)$	<ul style="list-style-type: none"> <li>• Does not require input parameter</li> <li>• Time complexity is too high or a set of objects</li> </ul>
DENCLUE	$O(n^2)$	<ul style="list-style-type: none"> <li>• Exclusion of density sensitivity</li> <li>• Possibility to have arbitrary cluster shape</li> <li>• Less sensitivity to outliers</li> </ul>
VDBSCAN	$O(n^2)$	<ul style="list-style-type: none"> <li>• Clusters of varying densities can be detected</li> <li>• Slow due to nonlinear complexity of time</li> </ul>
CUDA-DClust	$O(n * \log n)$	<ul style="list-style-type: none"> <li>• Outperforms DBSCAN</li> <li>• Computation of extent of remoteness may be ignored</li> </ul>
FDBSCAN	$O(n + nd)$	<ul style="list-style-type: none"> <li>• Better performance than DBSCAN</li> <li>• Better runtime complexity and computational time than DBSCAN</li> <li>• Higher object loss than DBSCAN</li> </ul>
LSDBC	$O(n \log n)$	<ul style="list-style-type: none"> <li>• Identifies cluster of arbitrary shape on noisy background</li> <li>• Can summarize and segment images into meaningful regions</li> <li>• Time complexity is high</li> </ul>

## 4 Algorithm Sparse and Proposed Methodology of Density-Based Clustering for Documents

### Algorithm: Sparsification

1. Read the text documents.
2. Represent the text using term frequency and inverse document frequency after performing preprocessing techniques like lemmatization, stop word processing, etc.
3. Compute similarity using angular distance between each pair of document vectors.
4. Focus on the upper triangular matrix and retain the similarity values.
5. Use a threshold value (user-defined threshold  $\alpha$ ) on extent of similarity and store the results in a sparse matrix form.
6. This result is used as input for the density-based clustering algorithm.

The steps followed to identify similar patterns on real-life dataset are described below:

- Step 1: Read the unstructured data from input files to conduct preprocessing activities.
- Step 2: Use TF-IDF to construct the term document matrix.
- Step 3: Apply the concept of angular similarity (31) to find the coherence of each document with all other documents. Apply sparsification algorithm.
- Step 4: Find the list of first-level neighbours for each document based on user-defined threshold  $\alpha$ .
- Step 5: The list of neighbours exhibits similar traits. The first level of neighbours which come in direct contact of the seed document forms the initial cluster. For each starting document (seed document), the direct contact documents will form initial clusters.
- Step 6: Compute shared neighbours for each pair of documents based on a user-defined threshold called crisp link threshold. This is the second level of neighbourhood to compute affinity groups. Identify the initial cluster and leader/representative or core document. The crisp link threshold determines affinity groups. Computation of the link is defined below:  

$$\text{link}(d_i, d_j) = \{|\text{affinity}(d_i, d_j)| \geq \beta\}$$
, where  $\beta$  is a user-defined parameter, i.e., two documents will be connected by a link if they have affinity or number of shared neighbours more than user-defined parameter  $\beta$ .  $\beta$  is called crisp link threshold.
- Step 7: Refine the groups based on common members. This step adds members to the cluster in an iterative fashion.
- Step 8: Collection of linked interconnected documents in the initial cluster with the linked neighbours based on affinity gives clusters.

- Step 9: Compute candidate documents. These are the documents which are shared neighbours to  $d_i$  and  $d_j$  but got cut off because of high. (Candidate documents are potential candidates to create an overlapping or intersecting cluster.)
- Step 10: Visualization of results using a graph. A graph may be constructed to visualize the clusters.
- Step 11: Comparison of the results with other existing algorithms of clustering.

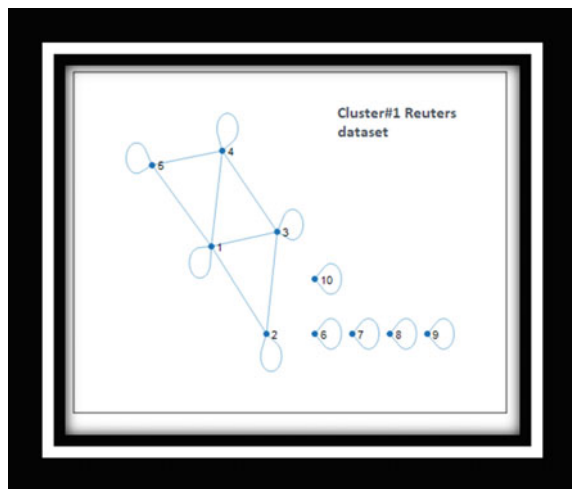
## 5 Empirical Results

Python 2.7 is used for tokenizing, stemming and stop word processing of text data. The computation of the tasks which deals natural language processing is performed on various datasets. MATLAB is used for generation of the graphs from the output obtained after identification of the clusters. To visualize the construction of clusters, a graph is drawn (depicted below). Reuters dataset is used as an input document. The graph is shown in the figure below. This indicates Cluster#1.

Figure 1 shows a closer look at the document IDs present in the cluster identified. Figure 1 also shows the outliers present in the 1000 documents clustered using this algorithm. Documents with IDs 6, 7, 8, 9 are the outliers. The nearest neighbour relation obtained with NN affinity threshold value “0.2” is shown below. The first two columns reflect the nearest neighbour documents. The third column reflects the affinity value. The nearest neighbours are shown in Table 1. The extent of similarity is also shown.

If link threshold is considered as 1, then the list of neighbours generated using DocSNN algorithm is shown below. This threshold value is also called DocSNN threshold. The list of neighbours are shown in the sample neighbour list (Fig. 2).

**Fig. 1** A cluster graph with all 1000 documents from Reuters dataset



**Table 1** Nearest neighbour relation

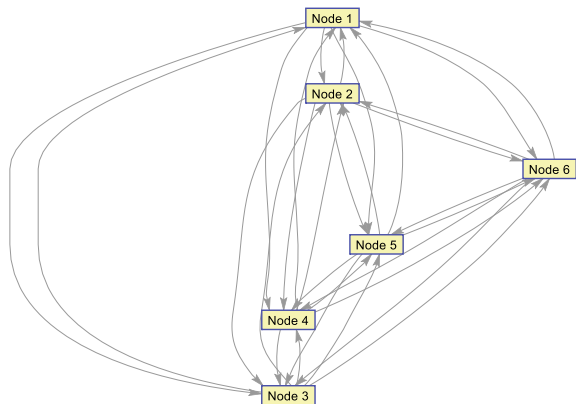
1.0000	1.0000	1.0000
1.0000	2.0000	0.9269
1.0000	4.0000	0.9269
2.0000	2.0000	1.0000
2.0000	4.0000	0.9975
3.0000	3.0000	1.0000
3.0000	5.0000	0.9985
3.0000	6.0000	0.9985
4.0000	4.0000	1.0000
5.0000	5.0000	1.0000
5.0000	6.0000	0.9996

**Fig. 2** Neighbours list

1	0	2	4
2	1	0	4
3	0	0	0
4	1	2	0
5	0	0	0
6	0	0	0

Above relation or matrix indicates that document IDs 1, 2 and 4 are in the same cluster, and cluster ID is 1. Document 3, Documents 5 and Document 6 are outliers. The graph generated using affinity can be visualized using NN affinity graph (Fig. 3). This is an intermediate graph generated before the computation of the final clusters. This graph will be further refined.

**Fig. 3** NN affinity graph generated with documents with ID 1, 2, 3, 4, 5, 6





In Figs. 4 and 5, details of the pattern identified are shown. The experiments have confirmed the correct identifications of patterns using the proposed methodology.

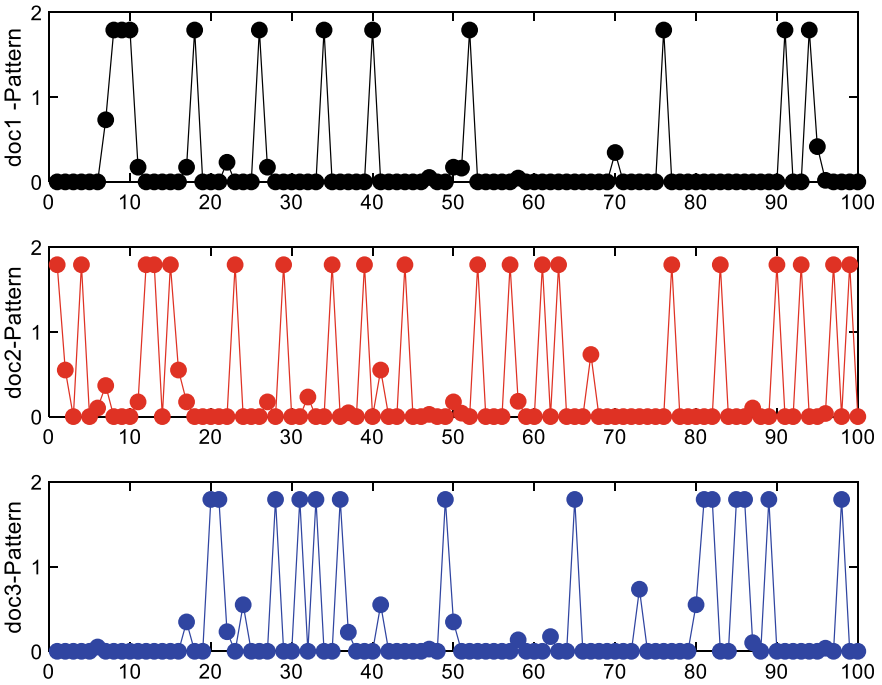


Fig. 4 Documents belonging to the same cluster are plotted

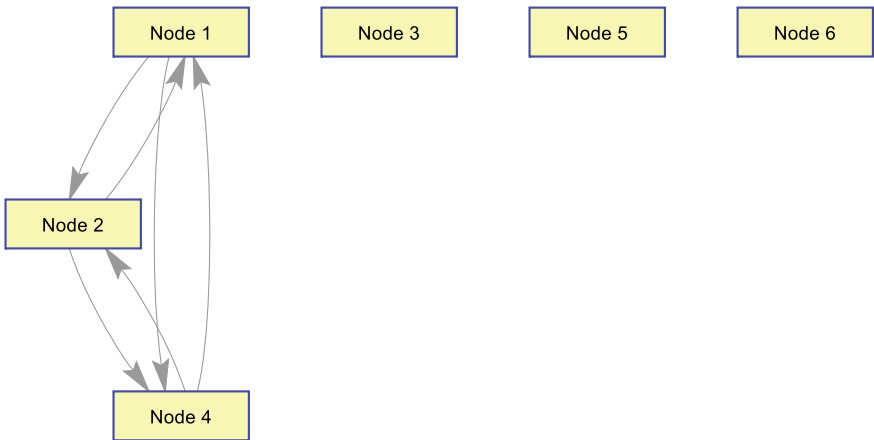


Fig. 5 Identified coherent pattern

The cluster is formed using three different nodes is shown in Fig. 4. Node 1, 2 and 4 are the members of the cluster. Node 3, 5 and 6 are outliers.

## 6 Conclusion

The methodology of clustering documents proposed in this paper is independent of the prerequisite of the number of clusters. This is an advantage to identify the natural number of clusters. This approach also can identify embedded cluster since it finds the commonality between different initial clusters. This approach does not take an additional parameter of a radius to define density. Here, the neighbourhood is defined by angular similarity. Later, the neighbourhood is refined based on a user-defined parameter. This algorithm is implemented on real-life datasets, and the results are found satisfactory. The selection of user-defined parameter may be done empirically. Datasets used are Reuters 21,578 dataset and spam base dataset from UCI machine learning repository. Comparative analysis with other document clustering has shown better space complexity and computational complexity due to the sparsification algorithm. The proposed shared nearest neighbour-based algorithm is able to identify patterns which are coherent in nature. The density-based approach of document clustering emphasized in this paper is able to detect similar clusters as detected by K-means and fuzzy C-means. This algorithm is being tested on different datasets. Next, work will be focused on detailed methodology on different datasets, related definitions and lemmas. More investigations are being performed to evaluate effectiveness.

## References

1. Singh P, Meshram PA (2017, November) Survey of density based clustering algorithms and its variants. In: 2017 international conference on inventive computing and informatics (ICICI), pp 920–926. IEEE
2. Ali T, Asghar S, Sajid NA (2010, June) Critical analysis of DBSCAN variations. In: 2010 international conference on information and emerging technologies, pp 1–6. IEEE
3. Chandrasekar RSV, & Britto GA (2019, June) Comprehensive review on density-based clustering algorithm in data mining. *Int J Res Anal* 6(2):5–9
4. Chauhan R, Batra P, & Chaudhary S (2014) A survey of density based clustering algorithms. *Int J Comput Sci Technol* 5(2):169–171
5. Maitry N, Vaghela D (2014) Survey on different density based algorithms on spatial dataset. *Int J Adv Res Comput Sci Manag Stud* 2(2):2321–7782
6. Böhm C, Noll R, Plant C, Wackersreuther B (2009, November) Density-based clustering using graphics processors. In: Proceedings of the 18th ACM conference on Information and knowledge management. ACM, pp 661–670
7. Goswami M, Sarmah R, Bhattacharyya DK (2011) CNNC: a common nearest neighbour clustering approach for gene expression data. *Int J Comput Vis Robot* 2(2):115–126

8. Goswami M, Purkayastha BS (2019, October) Discovering patterns using feature selection techniques and correlation. In: International conference on innovative data communication technologies and application. Springer, Cham, pp 824–831
9. Brown D, Japa A, & Shi Y (2019, April) An attempt at improving density-based clustering algorithms. In proceedings of the 2019 ACM Southeast conference (pp. 172–175)
10. Chen CL, Tseng FS, Liang T (2011) An integration of fuzzy association rules and WordNet for document clustering. *Knowl Inf Syst* 28(3):687–708
11. Wei CP, Yang CS, Hsiao HW, Cheng TH (2006) Combining preference-and content-based approaches for improving document clustering effectiveness. *Inf Process Manage* 42(2):350–372
12. Mugunthadevi K, Punitha SC, Punithavalli M, Mugunthadevi K (2011) Survey on feature selection in document clustering. *Int J Comput Sci Eng* 3(3):1240–1241
13. Willett P (1988) Recent trends in hierarchic document clustering: a critical review. *Inf Process Manage* 24(5):577–597
14. Luo C, Li Y, Chung SM (2009) Text document clustering based on neighbors. *Data Knowl Eng* 68(11):1271–1288
15. Hatamlou A, Abdullah S, Nezamabadi-Pour H (2012) A combined approach for clustering based on K-means and gravitational search algorithms. *Swarm and Evolutionary Computation* 6:47–52
16. Goswami M, Babu A, Purkayastha BS (2018) A comparative analysis of similarity measures to find coherent documents. *Appl Sci Manag* 8(11):786–797
17. Karol S, Mangat V (2013) Evaluation of text document clustering approach based on particle swarm optimization. *Open Comput Sci* 3(2):69–90
18. Shah N, Mahajan S (2012) Document clustering: a detailed review. *Int J Appl Inf Syst* 4(5):30–38
19. Huang A (2008, April) Similarity measures for text document clustering. In: Proceedings of the sixth New Zealand computer science research student conference (NZCSRSC2008), vol 4, Christchurch, New Zealand, pp 9–56)
20. Baghel R, Dhir R (2010) A frequent concepts based document clustering algorithm. *Int J Comput Appl* 4(5):6–12
21. Abualigah LM, Khader AT, Al-Betar MA, Awadallah MA (2016, May) A krill herd algorithm for efficient text documents clustering. In: 2016 IEEE symposium on computer applications & industrial electronics (ISCAIE), pp 67–72. IEEE
22. Patil LH, Atique M (2013, February) A novel approach for feature selection method TF-IDF in document clustering. In: 2013 3rd IEEE international advance computing conference (IACC), pp 858–862. IEEE
23. Chen CL, Tseng FS, Liang T (2010) An integration of WordNet and fuzzy association rule mining for multi-label document clustering. *Data Knowl Eng* 69(11):1208–1226
24. Andrews NO, Fox EA (2007) Recent developments in document clustering. Department of Computer Science, Virginia Polytechnic Institute & State University
25. Gil-García R, Pons-Porrata A (2010) Dynamic hierarchical algorithms for document clustering. *Pattern Recogn Lett* 31(6):469–477
26. Steinbach M, Karypis G, Kumar V (2000) A comparison of document clustering techniques, KDD workshop on text mining
27. Cui X, Potok TE (2005) Document clustering analysis based on hybrid PSO + K-means algorithm. *J Comput Sci (Spec issue)* 27:33
28. Beil F, Ester M, Xu X (2002, July) Frequent term-based text clustering. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 436–442. ACM
29. Fung BC, Wang K, Ester M (2003, May) Hierarchical document clustering using frequent itemsets. In: Proceedings of the 2003 SIAM international conference on data mining. Society for industrial and applied mathematics, pp 59–70
30. Chen CL, Tseng FS, Liang T (2010) Mining fuzzy frequent itemsets for hierarchical document clustering. *Inf Process Manage* 46(2):193–211

# A Privacy Preserving Hybrid Neural-Crypto Computing-Based Image Steganography for Medical Images



Tejas Jambhale and M. Sudha

**Abstract** With advancements in X-ray technology, there is an increase in the number of digital images used in the diagnosis of a patient. Whether it be a simple X-ray, MRI, CT scan or even a photo taken from a camera, the rise in the use of digital images has increased sharply. Though this has eased the entire process, it has brought the threat of cyber-attacks and breaches. The proposed method bridges this existing gap by incorporating suitable security mechanisms to preserve the privacy and confidentiality of medical diagnostic information of an individual. The approach utilizes neural networks to perform image steganography and combines it with a cryptographic algorithm (RSA) to secure medical images. The proposed method uses a two-level security providing a lower loss of 0.002188 on medical images improving upon existing image steganography techniques.

**Keywords** Image steganography · Neural networks · Deep learning · Cryptography · Security · Encryption · Medical images

## 1 Introduction

Whenever the security of images is concerned, steganography has always been in the conversation as it has multiple benefits and if combined with cryptographic techniques it can be used as a technique to secure medical data against security breaches. With an increase in datasets, neural network models have increasingly become better. Models currently are being used to solve problems in medical image analysis like detection of tumours in X-rays scans and segmentation of tissues in magnetic resonance imaging (MRI).

---

T. Jambhale · M. Sudha (✉)  
Vellore Institute of Technology, Vellore, India  
e-mail: [msudha@vit.ac.in](mailto:msudha@vit.ac.in)

T. Jambhale  
e-mail: [Tejas.jambhale98@gmail.com](mailto:Tejas.jambhale98@gmail.com)

These neural networks can also be modified and applied for image steganography. This area is still largely research-based. In medical images, image steganography can be used to maintain integrity, security and confidentiality. The images happen to have special requirements regarding medical images. Perceptual differences can lead to wrong analysis leading to an incorrect diagnosis. With the rise of CT scan and MRI [1], there is a rise in the importance of securing these images from breaches. Only the doctor, patient or any other authorized person should be able to access this data. Thus, in such a way, steganography has become an urgent requirement in the medical field.

Abnormalities in medical images like MRI have to be preserved during any type of manipulation. These images have important quantitative information with no specific canonical orientation. Algorithms must take such differences into account before they begin any manipulation steps.

Medical data contains quantitative information that must be preserved including intensity of neighbourhood pixel, the scale of abnormalities and their location in a scan. Change in any of these values can lead to wrong diagnosis leading to failure of the complete system. Contrary to other natural images, the orientation can often be ignored but not always the case. In such a way changing architectures that were developed for other natural images and taking these changes into account, the performance of the algorithm can greatly be improved and released to market-wide use.

## ***1.1 Protection Methods***

To protect patient identity, patient details should be protected by either encrypting them with cryptography or steganography or removed altogether. Even if the patient ID is protected, it is often desirable to encrypt other data like test details, scans and patient reports. This again can be completed using steganography, cryptography or a union of both. Security relies on either algorithm used for encryption and/or key management between concerned people.

Hence, the second method too must be incorporated by encrypting the report with embedded data for stronger security. The decryption of this encrypted data by the attacker is the only way security may be breached. A mixture of both cryptographic and steganography methods together may lead to the best results in protecting medical information in digital form. Cryptography used to encrypt the reports and which is then embedded through steganography.

Cryptography concentrates on encrypting the message and keeping its content secured while steganography is used for keeping the existence of a message a secret along with encrypting the message itself [2]. Once the presence of an embedded message is disclosed then the strength of steganography reduces, therefore combining steganography with cryptography is the best option.

Steganography is used for securing communication along with information hiding [3–5]. Only the concerned parties can detect the message sent through the channel. Since the message sent is invisible (embedded with cover image), the secret message

remains undetected. After the receiver extracts the secret image, the remaining cover image is considered irrelevant.

## ***1.2 Why Medical Data Needs to Be Secured***

Security is an issue that affects not only patients but also doctors, hospitals and concerned parties [6]. The medical industry like other industries is now especially required to protect patients against cyber-attacks providing different security features. An ongoing occurrence is a rise in the number of attacks at hospitals and medical research centres mainly due to their lack of security measures [7].

Patients are slowly becoming aware of the importance of privacy and have started worrying about their health information being accessed by unauthorized people. Historically, protection of medical data was never considered vital, but the increasing use of digital images in all tests and procedures means the requirement of some form of security has risen. A healthcare environment built on privacy, authenticity and security is a must.

## ***1.3 Problems with Existing Methods***

Currently, there exist multiple techniques to perform image steganography [6, 8–11] each with varying results and interpretations. Though these techniques are successful in completing the task, each technique faces some challenges making them a poor choice for real-life applications.

First is the most common and simple least significant bit (LSB) technique [6]. Since an image consists of 8 bits, LSB takes advantage of the fact that relevant data is stored in the first 4 MSB, and as moved towards the LSB, the percentage of relevant data decreases. The image to embed can then be encoded in these redundant bits. This method was considered one of the easiest methods for steganography but was found to have low imperceptibility. Another significant drawback was that it was not robust and was susceptible to image modification techniques [12, 13].

Edge-based image steganography [14] is an extension of the LSB technique where the secret image again is stored in the LSB of the cover images but only in the 3-bit LSB of the edges detected using edge detection. This algorithm faces similar issues as the original LSB technique.

Another existing method for image steganography specifically is discrete cosine transform [15] based which takes advantage of the frequency domain of images. This technique is more robust than spatial domain techniques such as LSB and used in many different applications like image compression. This technique involves the frequency transformations on the image [16].

Finally, it comes deep steganography which uses neural networks in the process of image steganography. Convolutional neural networks learn structures corresponding

to the logical features rather than spatial or frequency domains. CNNs can solve problems that come with other methods. CNNs can understand patterns, mark redundant areas in the cover image and take advantage of these areas, thus increasing both capacity and accuracy. This is done dynamically and hence is difficult for attackers to understand how encryption has been done, and only a person with access to the structure of the neural network would be able to decrypt. The proposed algorithm utilizes deep steganography to secure images and solve the problems faced by other techniques.

Some of the above methods have been tried in the medical industry with varying degrees of success, but a suitable solution is yet to be implemented, and so there is still a need for some form of security in the industry.

## 2 Parameters

The balance between distortion and the amount of encoded information in the image must be maintained.

$$L(c, c', s, s') = ||c - c'| + \beta ||s - s' || \quad (1)$$

where  $c$  is the input cover, and  $c'$  is a container image.  $s$  and  $s'$  are the secret image and secret cover image. This is the loss function given in [17] and has been widely accepted as a good metric in deciding a suitable model. This essentially gives the loss  $L$  with  $\beta$  parameter providing a method to decide the weightage given to the reconstruction of the secret image. Reducing this loss is considered the primary aim of any steganography model. The differentiation of this function helps in training the model in the neural network.

“The covered image should look very close to the cover image, and when revealed, the revealed image should look very close to the secret image” [17].

The following parameters must be taken into consideration while choosing an image steganography technique

- Capacity-Quality of the data should be preserved in the cover image. U-net gives higher payload capacity as compared to other techniques. In the medical field, this has high importance as image size varies from different use cases, and the quality has to be maintained without much distortion. Many techniques have been tried to maximize this parameter [18].
- Imperceptibility-Information hidden is undetected and undistinguishable by the human eye.
- Speed-Speed is also another factor as a slow system will be of no use in many high-frequency environments. Encryption time should be as less as possible.
- Security-It is the resistance of the algorithm to attacks even after the existence of secret has been revealed. Knowledge of the algorithm should not be enough to decrypt the image [13, 19]. It is the resistance to steganalysis attacks.

- Robustness-Should be resistant to image manipulation activities like cropping, compression, rotation and filtering in the encrypted image.
- Complexity-Complexity of the embedding algorithm should be as less as possible but not at the expense of other metrics.

### 3 Deep Steganography Approaches

The first type is the network specified by [17]. It consists of three stages: a preparation step, a hiding network and revealing network. The preparation step is important to prepare the data to be fed into the subsequent stages. The Hiding Net then concatenates the input image data with a cover image to get a container image. Lastly, the revealing network extracts the embedded data from the container image. The paper suggests noise should be added before the revealing network to improve stability.

Another model is two-part U-net similar to the one mentioned in [20, 21] which does not require a preparation phase and is seen to give more accurate results as compared to the previous network. This paper proposes using this architecture for implementation. The U-net takes its structure from fully convolutional network (FCN) and built such that it gives better segmentation results in multiple applications including medical imaging. U-net has symmetric structure, and concatenation is applied on skip connections of down-sampling and up-sampling path. Skip connections help extract local data for the global data during up-sampling. As it is symmetric, more feature maps in the up-sampling part allow the transfer of information. The symmetric shape makes U-net unique, differentiating it from other FCN variants. The input to the hiding network is  $256 * 256$  images and has seven down-sampling layers with a filter size of 4 and stride equal to 2, each having LeakyReLU activation function and batch normalization. The up-sampling phase is similar, thus maintaining symmetry. In the down-sampling, the feature channels are doubled in each stage whereas in up-sampling they are halved. Finally, the container image is given by a sigmoid function. The Reveal Net consists of six convolutional layers with a filter size of 3 again having ReLU activation functions and batch normalization. Here, the sigmoid function gives us the extracted secret image.

In medical imaging, the biggest advantage of using U-net is that it consists of no preparation phase. Therefore, it leads to multiple advantages and becomes suitable to handle medical images [22]. One being is that large computation time is saved by skipping the preparation phase required to process the image and speed being of utmost importance is improved upon by using U-net. Secondly and more importantly, since medical images are taken from different devices like x-ray, scanner, MRI or even a normal camera image size varies drastically, therefore, using conventional neural networks would require different cover images corresponding to the input secret image. By using U-net, the same cover image may be used as U-net handle different image sizes automatically, thus improving the process drastically. Hence, our algorithm uses U-net to implement image steganography.



### 4 Proposed Algorithm

See Fig. 1.

#### 4.1 Proposed Steps for the System

1. Encrypt the secret image using a standard cryptographic algorithm (here, RSA has been used).
2. Then using the U-net architecture embed this encrypted in a cover image to get the container image.

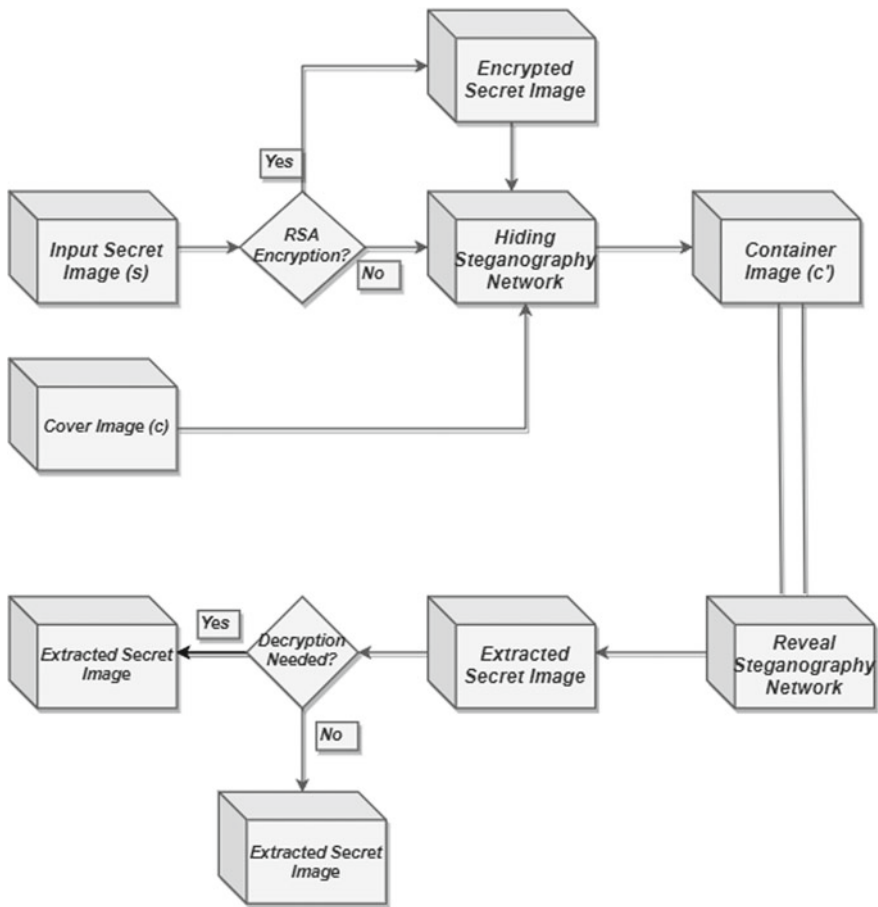


Fig. 1 Hybrid neuro-crypto computing-based image steganography model

Fig. 2 Data hiding

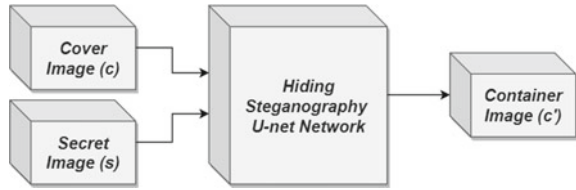
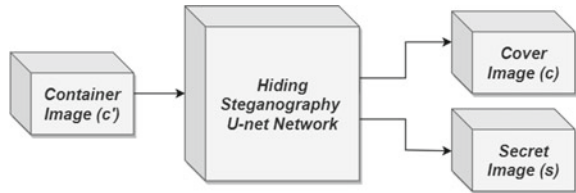


Fig. 3 Data reveal



3. Send this container image through the transmission channel.
4. Reveal the embedded image from container to reveal the secret image using U-net neural network at the receiver end.
5. Decrypt this extracted image for the cryptographic algorithm used initially using a key known only to this receiver.

In other data hiding methods, during the embedding of secret data, the cover image witnesses some distortion. There exist some irreversible disturbances caused to the cover media after extraction of the hidden message [23]. In scenarios like medical images, military, images used as evidence in judiciary and law enforcement, along with imperceptibility, the reversibility of the cover media without distortion is desirable.

A method is proposed to improve the method of steganography used in medical images. Current techniques as seen in [24–26] include other types of image steganography rather than neural networks, commonly referred to as deep steganography. But rather than directly implementing deep steganography on medical images using cover images, encrypting the image is proposed initially using a suitable cryptography algorithm (RSA, AES) depending upon the application. Here, RSA is proposed as public-key cryptography which is better suited to the medical industry where each patient will have his own private key as compared to symmetric techniques like AES. These asymmetric properties provided by the RSA algorithm will improve security by providing a double layer of protection by providing a fall-back option in case one technique fails.

The secret image is initially encrypted using a suitable cryptography algorithm. This image is then embedded into a cover image using steganography through use of hiding network neural networks. This ‘container’ image is sent through a channel to the designated receiver. This embedded data can then be extracted by passing this container through a reveal neural network (Figs. 2 and 3). Finally, the extracted data can be decrypted using a key known only to receiver depending upon the algorithm initially used. The networks are trained on ImageNet dataset which is a vast collection

of images on a range of domains. 60,000 images of the dataset were utilized for training this implementation.

#### ***4.2 Pseudo Code for Encrypting Secret Image Using RSA***

1. Read image and store as stream of pixels
2. Select 2 primes  $p$  and  $q$  randomly
3. Calculate  $n = p * q$  and  $\phi(n)$
4. Select  $e$
5.  $\text{mod2} = \phi(n)$
6. Define function  $\text{pow}(a, b, m)$ 
  - a. Set  $\text{ans} = 1$
  - b. While( $b > 0$ )
    - i. If  $b$  is odd:  $\text{ans} = \text{ans} * x \text{ mod } m$
    - ii.  $y = y//2$
    - iii.  $x = x^2 \text{ mod } m$
  - c. return  $\text{ans}$
7. Calculate  $d = \text{pow}(e, \phi(\phi(n))-1, \text{mod2})$
8. For all  $r, g, b$  pixels of image calculate
  - a.  $R = \text{pow}(r + 10, e, n)$
  - b.  $G = \text{pow}(g + 10, e, n)$
  - c.  $B = \text{pow}(b + 10, e, n)$
9. Set these  $R, G, B$  values as new pixels of image
10. Save the encrypted image

#### ***4.3 Pseudo Code Decrypted Image Using RSA–***

1. Read the encrypted image and convert into stream of pixels
2. For all  $R, G, B$  values
  - a.  $r = \text{pow}(R, d, n)-10$
  - b.  $g = \text{pow}(G, d, n)-10$
  - c.  $b = \text{pow}(B, d, n)-10$
3. Set these  $r, g, b$  values as new pixels
4. Save the decrypted image

### 4.4 Pseudo Code for Steganography Using U-Net

1. Read input secret and cover image
2. Resize image to 256 \* 256
3. Create hiding U-net model with suitable parameters
4. Create reveal network model
5. Set suitable optimiser and scheduler
6. Train model on dataset (ImageNet)
7. Test on input image minimizing MSE
  - a. Insert input images into Hiding Net to get container
  - b. Record loss B between cover and container
  - c. Save the container image
  - d. To extract insert container image into Reveal Net to get revealed secret and cover image
  - e. Record loss A between secret and revealed secret image
  - f. Calculate loss using above loss =  $A + \beta * B$
  - g. Save the revealed secret and cover image

## 5 Implementation

See Fig. 4.

Here, the secret image which is a scan is embedded in a picture which is the cover image to give container image. There is minimal loss as seen in Table 1 during the entire process and loss is as low as 0.000484. This loss is lower than a simple CNN implementation.

As seen above (Figs. 5 and 6), though the loss is very low, a simple difference between the images can slowly reveal the presence of a secret image. If an attacker

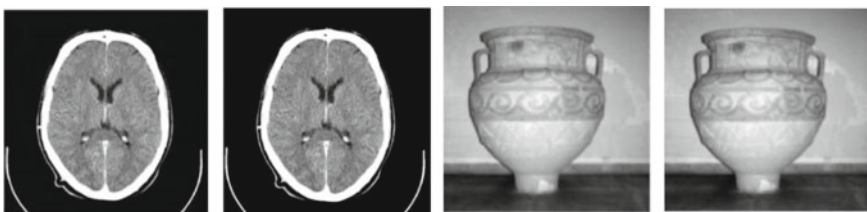
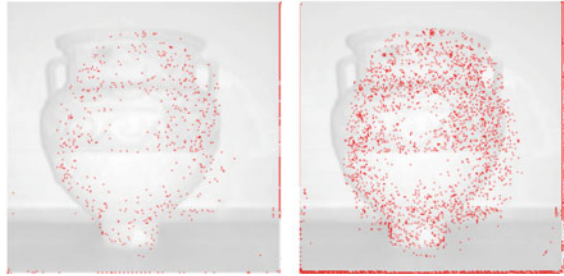


Fig. 4 Secret image, extracted image, cover image and container image. Source [27]

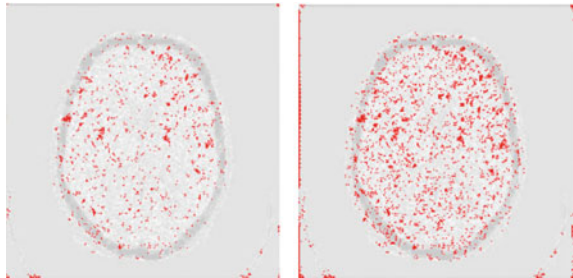
Table 1 Loss value (mean squared error) of brain radiology medical image without applying RSA

$A = \text{Loss for cover image}$	$B = \text{Loss for secret image}$	$\beta$	$\text{Loss} = A + \beta * B$	Time
0.000203	0.000375	0.75	0.000484	14.4 s

**Fig. 5** A difference of original cover and container after extraction for low and high threshold



**Fig. 6** A difference of secret and extracted secret image for low and high threshold



manages to somehow get hold of the original cover image, then the security of the entire system is compromised. This is not desired and lessens the security of the system. Therefore, to mitigate the consequences of this, RSA encryption is also incorporated. This helps with the fact that now even if the presence of stego image is leaked, RSA encryption ensures that the data is still secure and the only receiver with a secret key can retrieve this data (Table 2).

Here, secret image ‘a’ is first taken then encrypted with the RSA algorithm. This encrypted algorithm is then embedded in cover image ‘c’, and this container when arrives at the receiver is used for extraction of the embedded image. After extracting, the image needs to be decrypted. Only then will actual data be revealed. Figure 7 represents the cycle of the entire process.

The advantage of doing the extra step of encrypting is that even if the presence of stego image is revealed, actual patient data remains secure, and as seen above after finding the difference of two images, the attacker is able to only get an idea of general shape, rather than the actual content. From Table 2, the loss is also similar. Therefore, encrypting before embedding strengthens the security of the process. As seen, even if the attacker somehow manages to extract the encrypted image, the secret image

**Table 2** Loss value (mean squared error) of brain radiology medical image applying RSA

$A = \text{Loss for cover image}$	$B = \text{Loss for secret image}$	$\beta$	$\text{Loss} = A + \beta * B$	Time
0.000297	0.00172	0.75	0.001587	17 s

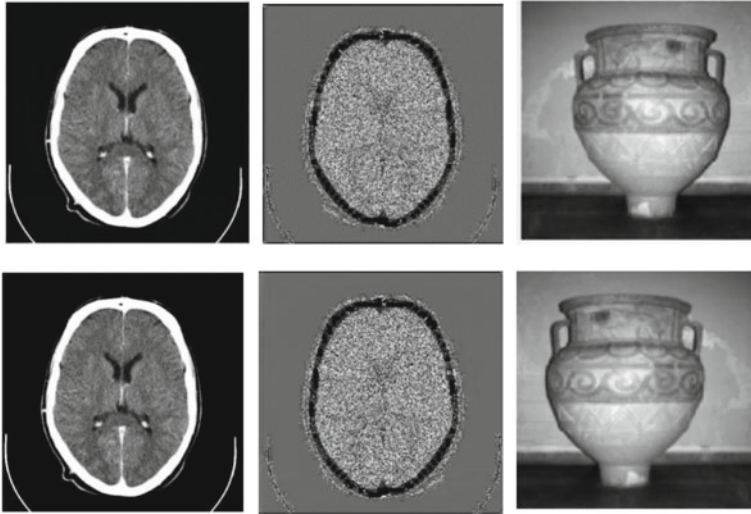


Fig. 7 Entire procedure of steganography

will remain protected as the actual relevant data of the image stays hidden under RSA encryption.

Without initially doing RSA encryption, the overall loss is comparatively less as seen in Table 3, and if the secret image is first encrypted using RSA algorithm, then loss is minutely more Table 4. But the advantage of encrypting is that even



Fig. 8 Without use RSA encryption. Source [28]

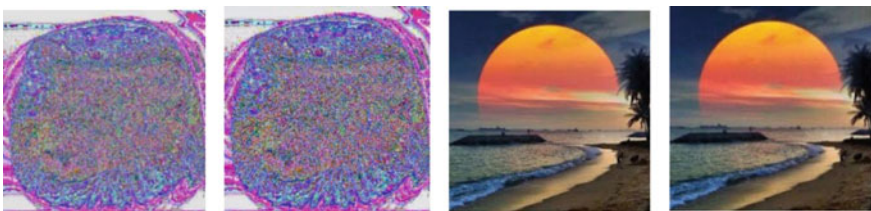


Fig. 9 Same image using RSA encryption before steganography

**Table 3** Loss (mean squared error) of corneal scan image using deep steganography

$A = \text{Loss for cover image}$	$B = \text{Loss for secret image}$	$\beta$	$\text{Loss} = A + \beta * B$	Time
0.000229	0.000679	0.75	0.000738	30 s

**Table 4** Loss (mean squared error) of corneal scan image using deep steganography and RSA

$A = \text{Loss for cover image}$	$B = \text{Loss for secret image}$	$\beta$	$\text{Loss} = A + \beta * B$	Time
0.000329	0.002479	0.75	0.002188	19 s

if the presence of stego image is revealed by subtracting the cover and container images (may not always work), the RSA makes sure that without access to the patients' private key, no attacker will get access to the embedded actual image, thus making the secret image imperceptible and secure (Figs. 8 and 9). Although loss is comparatively more, it still is low enough to provide desirable results without any loss in accuracy or imperceptibility.

When tested on images with high resolution, the images were standardized to the baseline  $256 * 256$ . Loss obtained was still found to be low ( $<0.0025$ ), and time taken was below 45 s. Though this is ideal, there was one limitation. Converting an image to a baseline of  $256 * 256$  can lead to some loss of important features. Hence, testing directly may be done without conversion. To do this, the model must be retrained for the new parameters. The new, higher resolution input means model takes a drastically longer time to train, and a smaller batch size must be taken while training. Though this is not ideal, the proposed algorithm still manages to give adequate results with increased execution time.

As discussed previously, the standard parameters for evaluation in image steganography are security, capacity and imperceptibility along with complexity. But these parameters often conflict with one another, and it remains important to strike a balance between these parameters. Often imperceptibility and capacity are given high value, but other factors like complexity and robustness play their part. The proposed model takes more time for encryption due to the time required to take the input image, encrypt using RSA and embed it into cover image and extraction at the user end. For an image of dimensions,  $256 * 256$  time taken may range from 12 to 45 s for the entire process. Different image resolutions are all transformed into the above dimensions. Currently, systems used in the medical field do not incorporate the required security and are unable to handle the complexity and variety of images available. Different types and sizes of images are often encountered, and the technique should not only be able to incorporate them but also make them resistant against foreign attack.

## 6 Conclusion

The proposed algorithm performs better than existing techniques giving lower loss and higher PSNR with the added benefit of decryption only being done by the receiver. This algorithm also uses RSA to surmount the drawbacks of image steganography used on its own. Therefore, this union of deep steganography along with RSA encryption provides an efficient mechanism to secure medical images. Image steganography using deep networks is yet a growing field with ongoing research. Currently, it can be seen that the proposed algorithm gives much better results than established techniques like LSB or DCT but at the cost of complexity and execution time. Medical images are often of varying sizes, colour depths and taken from different sources; hence, the proposed systems higher payload makes it a suitable technique. All attempts to attack deep steganography images have proved to be difficult. This combined with the strength of RSA makes the proposed system harder to crack and resistant to even the strongest attacks. Also, as seen above, the secret image extracted is hardly different from the original, and the same can be said of the cover image after embedding. In general, the improvement in payload capacity, resistance and imperceptibility makes it a suitable technique for medical applications. The union of cryptography and steganography makes it even more secure and maintains authenticity, integrity and makes it resistant against security breaches.

## References

1. Prabakaran G, Bhavani R, Rajeswari PS (2013, March) Multi secure and robustness for medical image-based steganography scheme. In: 2013 international conference on circuits, power and computing technologies (ICCPCT), pp 1188–1193. IEEE
2. Jain M, Choudhary RC, Kumar A (2016) Secure medical image steganography with RSA cryptography using decision tree. In: 2016 2nd international conference on contemporary computing and informatics (IC3I). <https://doi.org/10.1109/ic3i.2016.7917977>
3. Morkel T, Eloff JH, Olivier M (2005, June) An overview of image steganography. In: ISSA, pp 1–11
4. Santhi G, Adithya B (2017) A survey on medical image protection using various steganography techniques. *Adv Nat Appl Sci* 11(12):89–95
5. Johnson NF, Jajodia S (1998) Exploring steganography: seeing the unseen. *Computer* 31(2):26–34
6. Neeta D, Snehal K, Jacobs D (2006, December) Implementation of LSB steganography and its evaluation for various bits. In: 2006 1st international conference on digital information management, pp 173–178. IEEE
7. Mahmood A, Hamed T, Obimbo C, Dony R (2013) Improving the security of the medical images. *Int J Adv Comput Sci Appl* 4(9)
8. Hayes J, Danezis G (2017) Generating steganographic images via adversarial training. In: *Advances in neural information processing systems*, pp 1954–1963
9. Volkhonskiy D, Borisenko B, Burnaev E (2016) Generative adversarial networks for image steganography. In *ICLR 2016 Open Review*
10. Al-Hazaimeh OMA (2012) Hiding data in images using new random technique. *Int J Comput Sci Issues (IJCSI)* 9(4):49



11. Nissar A, Mir AH (2010) Classification of steganalysis techniques: a study. *Digit Signal Proc* 20(6):1758–1770. <https://doi.org/10.1016/j.dsp.2010.02.003>
12. Boehm B (2014) Stegexpose-a tool for detecting LSB steganography. arXiv preprint [arXiv:1410.6656](https://arxiv.org/abs/1410.6656)
13. Olson E, Carter L, Liu Q (2017) A Comparison study using stegexpose for steganalysis. *Int J Knowl Eng* 3(1)
14. Islam S, Modi MR, Gupta P (2014) Edge-based image steganography. *EURASIP J Inf Secur* 1:8
15. Walia E, Jain P, Navdeep N (2010) An analysis of LSB & DCT based steganography. *Global J Comput Sci Technol*
16. Khalil MI (2017) Medical image steganography: study of medical image quality degradation when embedding data in the frequency domain. *Int J Comput Netw Inf Secur* 9(2):22
17. Baluja S (2017) Hiding images in plain sight: deep steganography. *Adv Neural Inf Proc Syst*
18. Wu P, Yang Y, Li X (2018) Image-into-image steganography using deep convolutional network. In: *Pacific Rim conference on multimedia*. Springer, Cham, pp 792–802
19. Pibre L, Pasquet J, Ienco D, Chaumont M (2016) Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover sourcemismatch. *Electron Imaging* 8:1–11
20. Duan X, Jia K, Li B, Guo D, Zhang E, Qin C(2019) Reversible image steganography scheme based on a U-Net structure. *IEEE Access*. 10.1109/access.2019.2891247
21. Ronneberger O, Fischer P, Brox T (2015, October) U-net: Convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, Cham, pp 234–241
22. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA (2018) H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging* 37(12):2663–2674
23. Zhu J, Kaplan R, Johnson J, Fei-Fei L (2018) Hidden: hiding data with deep networks. In: *Proceedings of the European conference on computer vision (ECCV)*, pp 657–672
24. Elmasry W (2018) New LSB-based colour image steganography method to enhance the efficiency in payload capacity, security and integrity check. *Sādhanā* 43(5):68
25. Sumathi CP, Santanam T, Umamaheswari G (2014) A study of various steganographic techniques used for information hiding. arXiv preprint [arXiv:1401.5561](https://arxiv.org/abs/1401.5561)
26. Tiwari A, Yadav SR, Mittal NK (2014) A review on different image steganography techniques. *Int J Eng Innovative Technol (IJEIT)* 3(7):121–124
27. Seo JD (2018) Encrypting different medical images using deep neural network medium. <https://towardsdatascience.com/>
28. Mrinali Patel Gupta MD, Weill Cornell Medical College, New York. [www.visionaware.org/info/your-eye-condition/guide-to-eye-conditions/albinism-6165/125](http://www.visionaware.org/info/your-eye-condition/guide-to-eye-conditions/albinism-6165/125)

# Low-Power Reconfigurable FFT/IFFT Processor



V. Sarada and E. Chitra

**Abstract** Fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) are being applied in various fields of digital signal processing (DSP) applications because of the advanced technology of VLSI. In recent communication system, orthogonal frequency division multiplexing (OFDM) is the most important FFT and IFFT application. FFT/IFFT processors in these communication systems consume high power, rendering the system inefficient. OFDM-based UWB systems IEEE 802.15.4a employ 64 point and IEEE 802.11a-based systems employ 128/64 point and, therefore, we designed 128/64 point FFT/IFFT processor. The radix-2<sup>5</sup> algorithm that we used reduces the number of non-trivial multiplications. By implementing clock gating technique and a low-power multiplier, we propose to design a low power, reconfigurable, variable-length FFT/IFFT processor. The simulation of the design was done using ModelSim and power consumption has been analyzed with 180 nm CMOS technology using Cadence Encounter tool. The power consumed by the processor without clock gating is 58.27 mW and with clock gating is 2.13 mW.

**Keywords** Radix-2<sup>5</sup> FFT algorithm · Reconfigurable · Fast Fourier transform (FFT)/Inverse fast Fourier transform (IFFT)

---

V. Sarada · E. Chitra (✉)

Department of Electronics and Communication, SRM Institute of Engineering and Technology, Chennai, India

e-mail: [chitrae@srmist.edu.in](mailto:chitrae@srmist.edu.in)

V. Sarada

e-mail: [saradav@srmist.edu.in](mailto:saradav@srmist.edu.in)

**Table 1** FFT size for OFDM applications

System	FFT size
WLAN (IEEE.802.11a)	64 point
UWB System (802.15.4a)	128/64 point

## 1 Introduction

In the area of signal analysis, frequency spectrum estimate and OFDM-based communication systems [1], FFT have already been applied widely. Based on the standards and system parameters of OFDM, various length or size of FFT is needed. In Table 1, FFT size required in various communication standards is shown. Nowadays, low-power consumption is more important in portable application devices. Hence, it is desirable to design a power scalable, domain specific, area efficient, and high speed variable-length FFT/IFFT processor.

The three kinds of variable-length FFT architecture are general-purpose DSP, memory-based [2], and pipeline based. The most flexible one is general-purpose DSP, but it is neither area nor power efficient architecture. Memory-based is most area efficient but need many computation cycles. In pipelined architecture, every stage is processed once the data is available. It is suitable for real-time application of FFT processor because of simplicity, modularity, and high-throughput. The major classification of pipeline-based FFT processor is based on the process and the method of data storage. It is mostly used in real-time-based design.

In this paper, FFT/IFFT processor design is proposed by using single-path delay feedback pipelined-based architecture and Radix-2<sup>5</sup> algorithm. This algorithm is used because of its less hardware cost, area efficient, and low power in higher size FFT [3]. Earlier different variable-length design was proposed [4]. Variable-length FFT/IFFT is achieved [5] by bypassing higher stage in signal flow graph. Clock gating technique is also applied in digit slicing multiplier design for twiddle factor multiplication to reduce power.

The next section explains the algorithm employed in this paper. Section 3 gives an insight to the proposed processor design and sub-sections elucidate the approach to achieve low power and reconfiguration. Section 4 lists the parameters analyzed along with the results. Section 5 concludes the paper and gives the scope for future work.

## 2 FFT Algorithm

The DFT with N point for x (n) is formulated [6] as

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk}, K = 0, 1, \dots, N - 1 \quad (1)$$

where the twiddle factor is defined as

$$w_N^{nK} = e^{-j2\pi nk/N} = \cos\left(\frac{2\pi nk}{N}\right) - j \sin\left(\frac{2\pi nk}{N}\right)$$

The inverse discrete Fourier transform (IDFT) with  $N$  point is given by below expression

$$X(n) = \frac{1}{N} \sum_{k=0}^{N-1} x(k)W_N^{-nk}, n = 0, 1, \dots, N - 1 \tag{2}$$

Here, the time index is denoted by  $n$  and frequency index denoted by  $k$ ,  $X(k)$  is frequency domain signal  $X(n)$  is time domain signal,  $N$  is Length of FFT/IFFT signal.

In Radix-2<sup>5</sup> [7] algorithm a 6-dimensional linear index map is applied in Eq. 1 as shown below.

$$n = \left\langle \frac{N}{2}n_1 + \frac{N}{4}n_2 + \frac{N}{8}n_3 + \frac{N}{16}n_4 + \frac{N}{32}n_5 + n_6 \right\rangle N$$

$$\text{where } n_1, n_2, n_3, n_4, n_5 = 0, 1n_6 = 0 \frac{N}{32} - 1$$

$$k = \langle k_1 + 2k_2 + 4k_3 + 8k_4 + 16k_5 + 32k_6 \rangle N$$

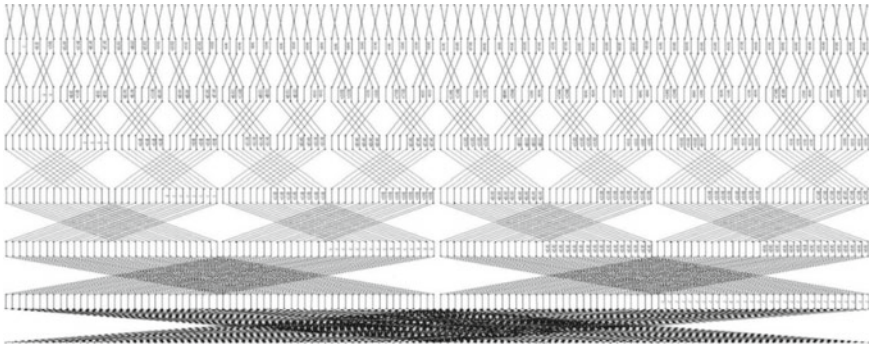
$$\text{where } k_1, k_2, k_3, k_4, \text{ and } k_5 = 0, 1k_6 = 0\left(\frac{N}{32} - 1\right)$$

$$X(k_1 + 2k_2 + 4k_3 + 8k_4 + 16k_5 + 32k_6)$$

$$= \sum_{n_6=0}^{\frac{N}{32}-1} \sum_{n_5=0}^1 \sum_{n_4=0}^1 \sum_{n_3=0}^1 \sum_{n_2=0}^1 \sum_{n_1=0}^1 x\left(\frac{N}{2}n_1 + \frac{N}{4}n_2 + \frac{N}{8}n_3 + \frac{N}{16}n_4 + \frac{N}{32}n_5 + n_6\right) W_N^{nk}$$

$$\sum_{n_6=0}^{\frac{N}{32}-1} J_{\frac{N}{32}}(n_6, k_1, k_2, k_3, k_4, k_5) \cdot W_N^{n_6(k_1+2k_2+4k_3+8k_4+16k_5)} \cdot W_{N/32}^{n_6k_6} \tag{3}$$

Using the above indexing, we calculated the twiddle factors for every stage for 128 point and 64 point input. The signal flow graph for 128 point FFT is as shown in Fig. 1. As the power of the radix increases the multiplication complexity decreases due to the elimination of non-trivial multiplications which we studied in [3].



**Fig. 1** Signal flow graph for 128 point FFT along with twiddle factors for each stage

### 3 Proposed Processor Design

The most popular FFT processor architecture is pipeline based because it is designed by emphasizing the regularity of data path and improved speed performance factor. The types of pipeline architecture are classified [1] as given below

- Single-path delay feedback (SDF) architecture
- Multiple-path delay commutator (MDC) architecture
- Single-path delay commutator (SDC) architecture.

In the last decades, many FFT algorithms and architecture were proposed to reduce the complexity in computation. MDC [8, 9] architecture will require much more memory, number of adders and multipliers than other architecture though it provides highest throughput.

Our proposed processor is of SDF architecture because of its high regularity and high utilization of components. The proposed architecture will be able to perform both FFT and IFFT for 64 as well as 128 point with the help of two external signals ‘a’ and ‘b’. Where ‘a’ differentiates between FFT and IFFT and ‘b’ differentiates between 64 and 128 point. The combinations are as shown in Table 2.

As seen in Table 2, when ‘a’ is logical 0, then FFT operation is performed and when it is logical 1, IFFT operation is done. When ‘b’ is logical 0, then 128 point is considered and when it is logical 1, then 64 point is taken. The below diagram in Fig. 2 is a schematic of the design we propose using SDF architecture.

**Table 2** Operation modes in the proposed design

Operation mode selection (ab)	Operation performed
00	128 point FFT
01	64 point FFT
10	128 point IFFT
11	64 point IFFT

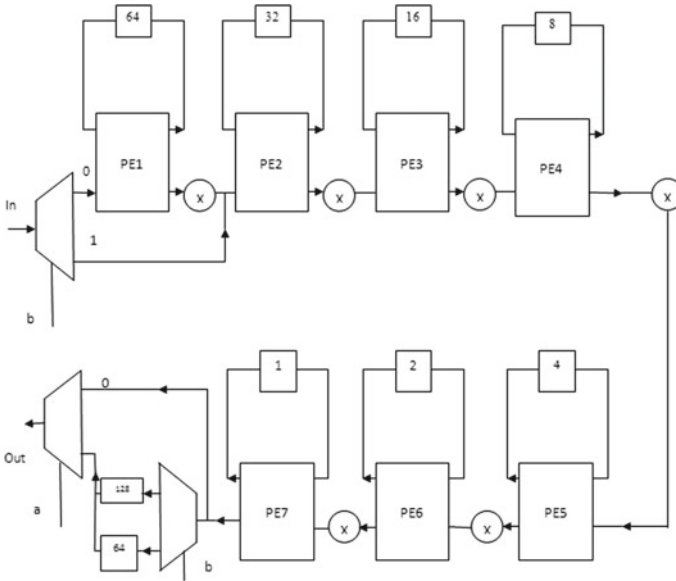


Fig. 2 128/64 point FFT/IFFT processor block diagram

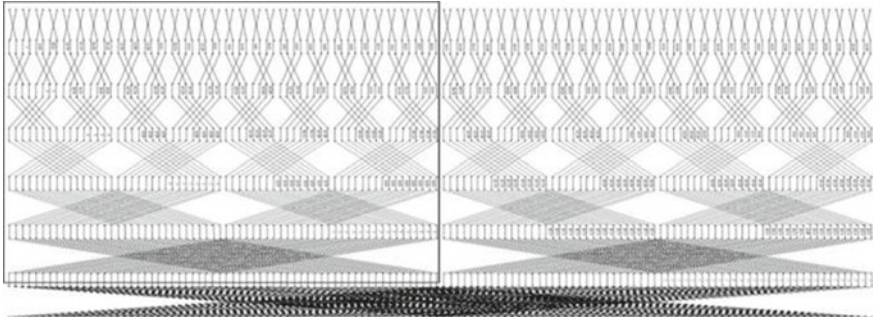
As shown in Fig. 2, the length is selected using the selection line ‘b’. Programming element (PE) consists of butterfly unit and switch to redirect the data into next stage for processing. PE switch performs two types of operation. One of the operation input data is pushed into last location of the buffer and the data is popped from the first location to output port. In the second operation, one of the outputs from butterfly unit is passed directly to the next stage and the other part of butterfly unit is returned back to the buffer.

The buffer in the first stage has a stack height of  $N/2$  and in our case since  $N = 128$ , the buffer has a height starting from 64. After the last butterfly stage, the output is selected accordingly based on ‘a’ and ‘b’.

The next sub-sections discuss the approach to proposed reconfiguration and low-power consumption design.

### 3.1 FFT and IFFT

Most of the communication systems require performing IFFT along with the FFT operation. In order to use the same architecture for IFFT, we first conjugate the input, then perform the usual operations in each stage, divide the last stage output by  $N$ , and then conjugate the output [10]. Another method to find IFFT is to conjugate the twiddle factors in each stage. As shown in the block diagram, the selection line ‘a’ is used to choose between FFT and IFFT. If ‘a’ is logical zero, then the processor will



**Fig. 3** 64 point signal flow graph (Highlighted)

perform FFT operation for the serial input; otherwise, the processor will perform IFFT.

### 3.1.1 128 Point and 64 Point FFT/IFFT

By extending the processor to variable length, processor generic design is made to certain extent for systems employing these lengths. The radix  $2^5$  algorithm evinces the advantage of having the same set of twiddle factors within the stages for any given length of the input. In the signal flow graph, if we eliminate the first stage and consider it from the second stage [11, 12] alone with 64 inputs, then it resembles exactly to the 64 point signal flow graph. We exploit this to extend our processor to lower lengths. The below graph in Fig. 3 elucidates the graph in a simpler manner. The highlighted portion is the 64 point FFT signal flow graph using radix  $2^5$  algorithms. This is exploited to integrate variable length in our processor.

### 3.1.2 Low-Power Consumption

The two main techniques that led to the low-power consumption of the proposed FFT processor are digit dlicing multiplier [13] and clock gating technique. The power consumption and computation complexity are reduced in digit slicing multiplier. The concept used in digit slice multiplier is one of the multiplication parameter in binary number which is divided into block of short binary number with short word length and perform multiplication in parallel. Based on the word length, the fundamental sliced algorithm is applied to the binary number.

Clock gating [14] is a method used to reduce power dissipation by blocking unwanted clock signal in register and associated logic. The clock is responsible for the power dissipation significantly in synchronous digital circuit. By disabling the unwanted switching on the parts of clock signals, the dynamic power dissipation is

reduced. Clock gating logic can be added into a design in many ways coded into the RTL code at system level and gate level.

1. By adding enable condition in RTL code, the synthesizer tools automatically generates clock gating logic
2. Clock gating technique is manually added by adding gates to the specific modules or register.

The clock gating logic used in the proposed architecture is known to be low-power clock-gated RTL code which belongs to the first method stated above. In this type, an enable pin is given to the components to which clock gating has to be applied. A new clock is defined to these components such that it is the result of AND logic between original clock and the enable signal. Thus, whenever the enable signal is logic zero, the component will not consume power and whenever it is logic one, the components works according to its logic.

### 4 Simulation Results and Analysis

We have simulated the design using ModelSim 10.4 with Verilog and analyzed the power using Cadence Encounter v.11.20. The simulations for each sub-module are given below. Figure 4, 5, 6, 7 and Fig. 7 show the outputs for 128/64 point FFT/IFFT.

The above figure shows the serial output when a serial input is given for 128 point FFT, i.e., when  $a = 0$  and  $b = 0$ .

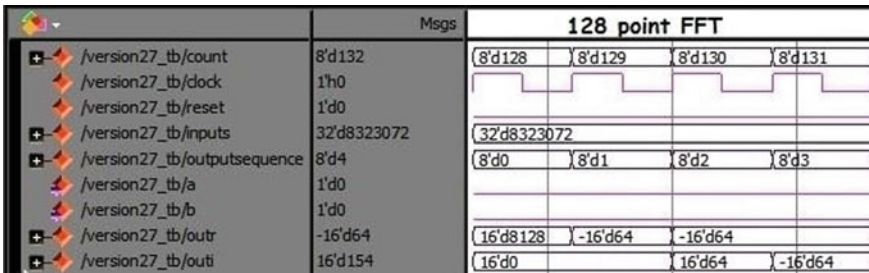


Fig. 4 128 point FFT output

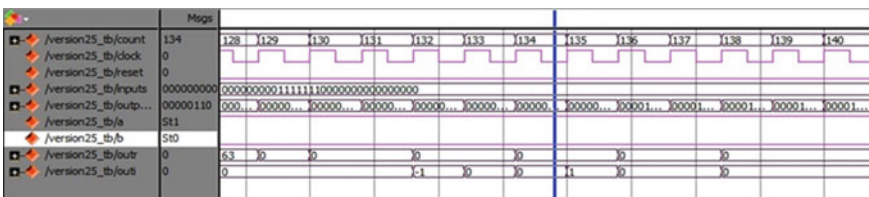


Fig. 5 128 point IFFT output



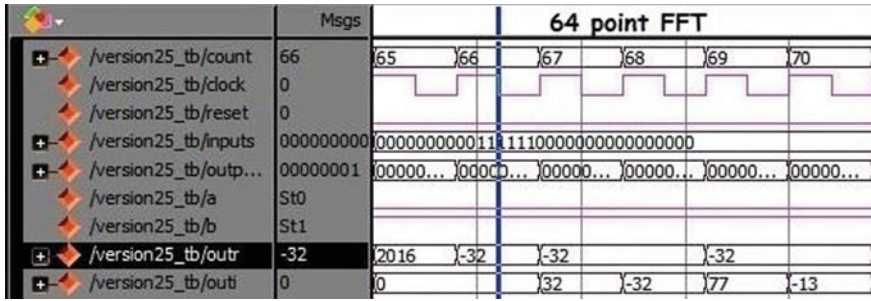


Fig. 6 64 point FFT output



Fig. 7 64 point IFFT output

The above simulation result shows the output of 128 point IFFT operation, i.e., when  $a = 1$  and  $b = 0$ .

The above screenshot shows the serial, 64 point, FFT output, i.e., when  $a = 0$  and  $b = 1$ .

The above simulation result shows the output of 64 point IFFT operation, i.e., when  $a = 1$  and  $b = 1$ .

The FFT/IFFT processor design without clock gating and with clock gating power reports are as shown in Figs. 8 and 9, respectively. It is synthesized using 180 nm technology standard cell library with 1.8 supply voltage, cadence encounter tool, and the results have been tabulated in Table 3.

## 5 Conclusion and Future Scope

The proposed processor performs FFT and IFFT for variable lengths with low complexity due to the implementation of radix- $2^5$  algorithm. Implementation of digit slicing multiplier and clock gating technique makes the processor to consume less power than the existing design. The power consumed by the processor without clock gating is 58.27 mW and with clock gating is 2.13 mW. This FFT/IFFT processor design can be extended to generic size of  $N$  point.

```

=====
Generated by:      Encounter(R) RTL Compiler v11.20-s017_1
Generated on:     Apr 12 2016  04:17:39 pm
Module:          test20
Technology library:  tsmc18 1.0
Operating conditions: slow (balanced_tree)
Wireload mode:   enclosed
Area mode:       timing library
=====

          Instance                Leakage      Dynamic      Total
          Cells Power(nW)      Power(nW)    Power(nW)
-----
test20                35284 32171.987 58246418.000 58278589.987
    
```

Fig. 8 Power report without clock gating

```

=====
Generated by:      Encounter(R) RTL Compiler v11.20-s017_1
Generated on:     Apr 23 2016  10:28:10 am
Module:          test21
Technology library:  tsmc18 1.0
Operating conditions: slow (balanced_tree)
Wireload mode:   enclosed
Area mode:       timing library
=====

          Instance                Leakage      Dynamic      Total
          Cells Power(nW)      Power(nW)    Power(nW)
-----
test21                36254 41419.269 2093247.754 2134667.023
    
```

Fig. 9 Power report with clock gating

Table 3 FFT/IFFT processor power report

	Data word length	Power
64 point FFT/IFFT [15]	16	9.79 mW at 20 MHz
64 point FFT/IFFT [10]	16	6.45 mW at 20 MHz
64/128 FFT/IFFT proposed design	16	2.13 mW at 20 MHz

## References

1. He S, Torkelson M (1996) A new approach to pipeline FFT processor. In: Proceedings of the 10th international parallel processing symposium, IPPS, pp 766–770, 15–19 April 1996
2. Hsiao C-F, Chen Y, Lee C-Y (2010) A generalized mixed-radix algorithms for memory based FFT processor. IEEE Trans Circ Syst II Express Briefs 57:26–30
3. Oh J-Y, Lim M-S (2005) New radix-2 to the 4th power pipeline FFT processor. IEICE Trans Electron E88-C(8)

4. Lin YW et al (2005) A 1-GS/s FFT/IFFT processor for UWB applications. *IEEE J Solid-state Circ* 40(8):1726–1735. MDF
5. Yang C-H, Yu s-H, Markovic D (2012) Power and area minimization of reconfigurable FFT processors 3GPP-LTE example. *IEEE J Solid Circ* 47(3):756–768
6. Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex Fourier series. *Math Comput* 19:297–301
7. Cho T, Lee H (2012) A high-speed low-complexity modified radix-2<sup>5</sup> FFT Processor for high rate WPAN applications. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 1063–8210
8. Lee T-Y, Huang C-H, Chen W-C, Liu M-J (2016) A low-area dynamic reconfigurable MDC FFT processor design. *Microprocess Microsyst* 42:227–234
9. Yang K-J, Tsai S-H (2013) MDC FFT/IFFT processor with variable Length for MIMO-OFDM systems. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 21(4)
10. Yu C, Yen M-H, Hsiung P-A, Chen S-J (2011) A low-power 64-point pipeline FFT/IFFT processor for OFDM applications. *IEEE Trans Consum Electron* 57(1)
11. Mario G, Grajal, Sanchez MA, Gustafsson O (2013) Pipelined radix-2<sup>k</sup> feedforward FFT architectures, *IEEE transactions on very large scale integration systems*
12. Vennila C, Lakshminarayanan G, Ko S-B (2011) Dynamic partial reconfigurable FFT for OFDM based communication systems. *Circ Syst Signal Process* 1–18
13. Sarada V, Vigneswaran T (2015) Low power complex multiplier based FFT processor. *Int J Eng Technol (IJET)*
14. Yeap G (2002) *Practical low power digital VLSI design*. Kluwer Academic Publishers, Springer US
15. Arunachalam V, Raj ANJ (2014) Efficient VLSI implementation of FFT for orthogonal frequency division multiplexing applications. *IET Circ Devices Syst* 8(6):526–531

# Color Image Watermarking in DCT Domain Using SVR



Shabnam Thakur, Rajesh Mehta, and Geeta Kasana

**Abstract** A robust digital image watermarking scheme in the DCT domain for color images using support vector regression is presented in this paper. This scheme attempts to achieve a fair balance between imperceptibility and robustness using an array of techniques including machine learning. The host image is transformed from RGB to YCbCr color space which gives a luma component used for embedding the watermark. This scheme utilizes Lagrangian Support Vector Regression (LSVR) for learning image features; entropy for selecting the image blocks in which watermark bits are embedded; discrete cosine transform to de-correlate the luminance component pixels. The low-frequency coefficients scanned in a zigzag manner are used for training and testing of LSVR model, and the output acquired from LSVR is used for embedding the watermark. The proposed method is applied to general image processing dataset images and found to yield fairly good results as shown by the comparison with existing methods.

**Keywords** YCbCr · Lagrangian support vector regression · Entropy · Discrete cosine transform

## 1 Introduction

The advent of the Internet has heralded in a new era in the domain of data sharing. Innovators can share their creations and ideas with the rest of the world faster than was ever possible before. While largely advantageous, this method of information sharing

---

S. Thakur (✉) · R. Mehta · G. Kasana  
Thapar Institute of Engineering and Technology, Patiala, India  
e-mail: [shabnamt1@gmail.com](mailto:shabnamt1@gmail.com)

R. Mehta  
e-mail: [rajesh.mehta@thapar.edu](mailto:rajesh.mehta@thapar.edu)

G. Kasana  
e-mail: [gkasana@thapar.edu](mailto:gkasana@thapar.edu)

has its drawbacks. With the right knowledge and skills, all of the information on the World Wide Web can be accessed and distributed freely without the owner's consent. This has led to a surge in the number of copyright infringement, piracy, and plagiarism cases. Digital watermarking has emerged as a promising field to ensure intellectual content protection. The owner can embed their watermark in the product and claim rights over it, if and when required. Watermarking can be done on images, audio [1], video [2], and documents [3]. Watermarking is essentially the embedding of specific data or information into a carrier (or host) in a robust manner. Our concern, in this work, is digital image watermarking which is inserting information into a host image [4]. Image watermarking is mostly performed in two domains [5]—the spatial domain and the transform domain. Spatial domain watermarking requires no pre-processing of the host image as the bits of the watermark are directly embedded into the host. Methods like least significant bit substitution and additive watermarking fall under this category. Su et al. [6] proposed a sturdy method in the spatial domain using the features of the Direct Current (DC) and generation principle. The watermark is then subdivided into four which are embedded four times into the host image. Although computationally less complex and easy to perform, the techniques in this domain show little to no resistance to geometric attacks. The transform domain, on the other hand, converts the host image from spatial to frequency domain after which it embeds the watermark bits into the coefficients obtained. This process, while computationally intricate, shows more robustness in the face of attacks. Discrete Fourier Transform (DFT) [7], Discrete Cosine Transform (DCT), Integer Wavelet Transform (IWT) [8], Discrete Wavelet Transform (DWT) [9], and Singular Value Decomposition (SVD) [8] belong to this domain of watermarking.

While the watermarked image is being transferred over the Internet, it can be attacked by persons with malicious intent. To measure the efficacy of any watermarking technique, the criteria of robustness and imperceptibility are used as parameters. However, there is a trade-off between the two with one likely to decrease as the other increases. This makes obtaining equilibrium between imperceptibility and robustness an optimization problem. Optimization problems are best solved using machine learning algorithms as has been seen in the recent literature of diverse fields ranging from finance and medicine to digital watermarking. Machine learning algorithms like support vector machines and deep learning networks have been extensively used in digital watermarking. In our study, Lagrangian Support Vector Regression (LSVR) and entropy [10] are employed to achieve better results. LSVR has fast learning speed and good generalization property. It is used to find the relationship between the middle pixel and its eight neighbors to estimate an appropriate location for watermark embedding. On the other hand, entropy yields the most favorable features for embedding the watermark.

Although recent color image watermarking the literature has shown promising results, it is still widely under-explored area as compared to gray image watermarking. Wang et al. [11] present a method wherein the watermark, and host images are transformed from RGB space to YIQ using the Haar wavelet. After applying one-level DWT and three-level DWT on the watermark and the host, respectively, four coefficient matrices are procured. Watermark embedding into the carrier image

is done via a predetermined algorithm, and on the resulting image, three-level DWT is applied to finally obtain a watermarked image. Ko et al. [12] present a transparent watermarking method which uses the difference between the DCT coefficients of two blocks and depending on the watermark bit modifies it. A zigzag sequence is employed to determine the extent of modification based on the median of AC coefficients and the DC coefficient. Yuan et al. [13], in original research, selected some fixed blocks from the middle frequency band after applying DCT on them. Watermark embedding and extracting are done by changing the size association between the chosen middle frequency coefficients according to a predefined set of rules. Patvardhan et al. [14] transformed the image from RGB to YCbCr applied wavelet domain transformation and used SVD to select the bits to be embedded. For enhanced security, Quick Response (QR) code has been used as the watermark. Giri et al. [15] apply level-1 DWT on the carrier image, thereby dividing it into four sub-bands-LL, HL, LH, and HH. HL is considered to be less sensitive to HVS; therefore, it is an ideal location for embedding. The modulus operator is applied on the watermark which results in three different digit streams. These are then embedded into three coefficient streams of the host. Inverse DWT is performed to get the watermarked image.

Our objectives in this study are:

- (a) To present an innovative and robust image watermarking technique suitable for color images.
- (b) To achieve an equilibrium between robustness and imperceptibility as their values vary inversely to each other.

This paper is organized as the prerequisites are described in Sect. 2; Sect. 3 presents the proposed scheme; the experimental results are presented in Sects. 4 and 5 bring a conclusion to the paper.

## 2 Preliminaries

### 2.1 YCbCr

Human Vision System (HVS) is much more sensitive to grayscale (or black and white) images than it is to color images. YCbCr is a family of color spaces which take advantage of this fact by essentially separating the grayscale and color components into a Luminance (or Luma) component, Y, and two chrominance (or chroma) components, Cb, and Cr. Cb represents the color of chrominance to blueness, and Cr represents the color of chrominance to redness. This color space is more advantageous for embedding because errors can be introduced in perceptually meaningful ways in YCbCr space via subsequent operations. As compared to RGB, YCbCr can also achieve better robustness against a large number of attacks and does not store too much redundant information.

## 2.2 Arnold's Transform

It [16] is a chaotic transformation after whose application the pixels of the image appear to be randomly rearranged. It is used on the watermark in the algorithm to provide an additional layer of security. For an image of  $N \times N$  dimensions, Arnold's transform is defined as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \text{ mod } N \quad (1)$$

where  $x, y, x', y' \in \{0, 1, 2, \dots, N - 1\}$ ,  $x'$  and  $y'$  are the coordinate of a position in the scrambled image,  $x$  and  $y$  are the coordinates of a position in the original image, and  $N$  is the size of the watermark image. The watermark is scrambled, and an unrecognizable watermark is obtained.

## 2.3 Discrete Cosine Transform

DCT is a linear transform which is popularly utilized as compression and signal decomposition technique. It transforms an image from the spatial domain to the frequency domain by converting the host image into its cosine coefficients. This conversion results in three frequency bands-high, middle, and low. The amount of image information contained in each band decreases from lower to higher bands. The most significant image information is stored in the low-frequency band, and as a result, this band is most robust against potential attacks. Operations performed on this band might affect the image quality. The high-frequency band contains redundant or non-important data which can potentially be compressed but watermarking here is not as robust. DCT has been employed in the low-frequency band; due to the robustness, it provides against common attacks such as JPEG compression [17].

## 2.4 Lagrangian Support Vector Regression

Machine learning is very effective in tackling problems on forecasting. SVM has the advantage of better generalization performance in comparison to other methods and provides a unique solution. Regression aims to predict the correlation between given input variables and their output values. SVM methods have been effectively applied to regression by the usage of  $\mathcal{E}$ -insensitive error loss function proposed by Vapnik [18]. This led to the formulation of LSVR. The training for learning the image features and testing for embedding the watermark using LSVR model [19, 20] used in color image watermarking in this paper reveals the efficacy of LSVR.

### 3 Proposed Scheme

The experiments in this paper are performed using a cover (or host) image,  $I$ , of dimensions  $512 \times 512$  and a grayscale watermark,  $W$ , of dimensions  $32 \times 32$ . The embedding process has been presented in Fig. 1.

#### 3.1 Embedding Procedure

- (i) Convert  $I$  into YCbCr color space. This gives us three components- $Y$ ,  $Cb$  and  $Cr$ .  $Y$  Component is selected for embedding.
- (ii) Arnold’s chaotic map is applied on  $W$  to offer an additional layer of security by scrambling the image. After applying Arnold transform, a scrambled watermark- $Wm$  is obtained.
- (iii) The  $Y$  component is separated into sub-blocks of size  $8 \times 8$ . Fuzzy entropy of each block is calculated, and the blocks are arranged in descending order in the spatial domain.
- (iv) DCT is applied to every selected sub-block, and then, features (low-frequency DCT coefficients) are selected. Zigzag traversal is made on each of the sub-blocks till the tenth index position. From these ten DCT coefficients, nine coefficients are selected except for the first DC coefficient to form a size  $3 \times 3$

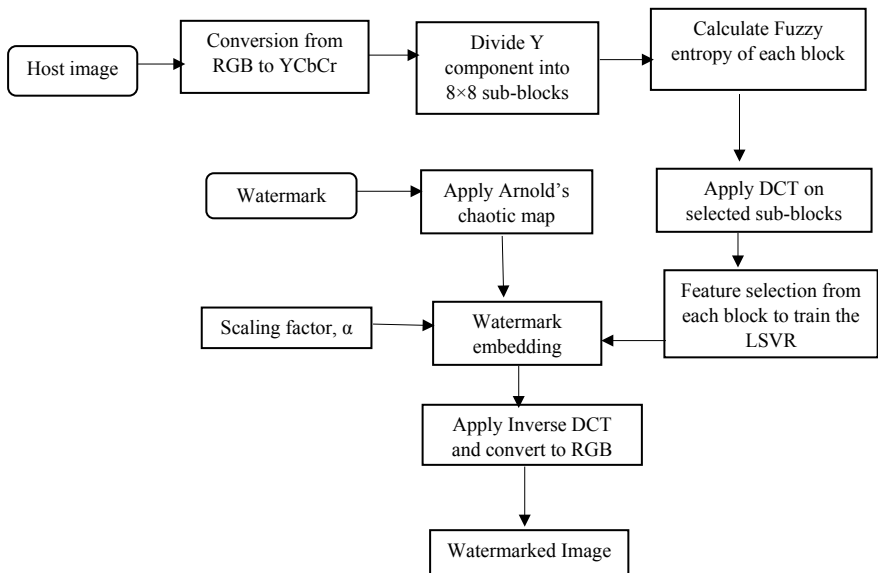


Fig. 1 Block diagram of the embedding process



window. This yields a vector which is to be used as the dataset to train the LSVR model.

- (v) The centrally located coefficient, i.e., the fifth coefficient is the anticipated output of the model, and its neighbors act as input. From the dataset, odd indexed values are used for training and even for testing. The predicted output is obtained corresponding to the required output, and in this watermark is embedded according to:

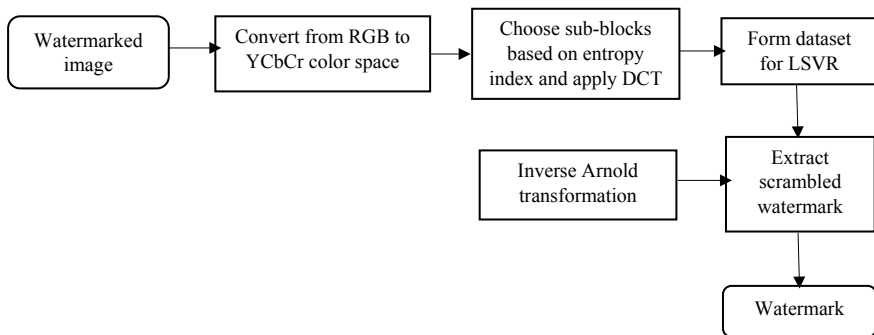
$$\begin{aligned}
 & \text{if } wmBitValue == 1 \\
 & \quad \text{mod } Value = \max(trgCmp, lsvrOut + \alpha) \\
 & \text{else} \\
 & \quad \text{mod } Value = \max(trgCmp, lsvrOut - \alpha)
 \end{aligned}$$

where  $wmBitValue$  is the pixel value at the  $(i, j)$ th location of the watermark;  $modValue$  is the modified central value;  $trgCmp$  is the anticipated output;  $lsvrOut$  represents the output of the trained LSVR; and  $\alpha$  is the scaling factor whose value is considered to be 15 in the proposed work.

- (vi) Original blocks of the host image are now replaced with the modified blocks which have the embedded watermark. Inverse DCT is used to produce the watermarked luminance component.
- (vii) The watermarked Y component is now merged with the two chroma components, i.e., Cb and Cr, to obtain YCbCr image which is then transformed into an RGB to yield a last RGB watermarked image. Then, the value of RGB watermarked image is tested by the Peak-Signal-to-Noise Ratio (PSNR).

### 3.2 Extraction Procedure

The extraction process has been presented in Fig. 2.



**Fig. 2** Block diagram of the extraction process

- (i) A transformation is applied on the watermarked image,  $WI$ , to YCbCr color space to obtain  $Y$ ,  $Cb$  and  $Cr$  components. Since the legitimate user knows which component to use for extraction,  $Y$  component is used.
- (ii) Sub-blocks of  $Y$  component are made according to  $indEntropy$  used in the embedding process. To each of these blocks, DCT is applied.
- (iii) Zigzag traversal is made to obtain a dataset for the LSVR model similar to the one used for embedding. Odd indexes are used for training and even for testing. The even indexed predicted values are compared with the expected values, and watermark bits are obtained using:

$$\begin{aligned}
 & \text{if } exOut > preOut \\
 & \quad wmBitValue = 1 \\
 & \text{else} \\
 & \quad wmBitValue = 0
 \end{aligned}$$

where  $exOut$  is the expected output;  $preOut$  is the output predicted by LSVR; and  $wmBitValue$  is the watermark bit value.

- (iv) Arnold's chaotic map is applied to  $Wm$ , resulting in the original watermark,  $W$ .

## 4 Experimental Results

The experiments in this paper are performed on MATLAB 2019a using Intel(R) Core(TM) i5-2450 M CPU @ 2.50 GHz 2.50 GHz system. The scheme was tested on several images including Lena, Airplane, and House from USC-SIPI image dataset as shown in Fig. 3 along with the binary watermark. The effectiveness of the scheme is figured out based on imperceptibility and robustness. Figure 4 presents the watermarked images along with the extracted watermarks.



**Fig. 3** Host images: Lena, airplane, and house with the watermark



Fig. 4 Watermarked images corresponding to Fig. 3 along with recovered watermarks

#### 4.1 Imperceptibility

Imperceptibility is an important factor in analyzing the strength of a technique. It can be ascertained using the PSNR value.

PSNR is obtained between the watermarked image and original image using:

$$\text{PSNR} = 10 \log_{10} \left( \frac{255^2}{\text{AMSE}} \right) \quad (2)$$

where average mean square error (AMSE) is expressed as:

$$\text{AMSE} = \frac{\text{MSE(RC)} + \text{MSE(GC)} + \text{MSE(BC)}}{3} \quad (3)$$

where MSE(RC) is the mean square error of the red channel and is calculated by,

$$\text{MSE(RC)} = \frac{\sum_{M,N,M,N,M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \quad (4)$$

where  $I_1$  represents the original image,  $I_2$  is a watermarked image,  $M$ , and  $N$  are image dimensions;  $(m, n)$  are the coordinates of a pixel in the image. The values of MSE(GC), for the green channel, and MSE(BC), for blue channel, are found similarly. PSNR values are measured in decibels (dB). Values greater than 40 dB are considered good. High PSNR value shows the watermarked image is not much distinguishable from the original to the human eye. The resultant PSNR values calculated for the test images are shown in Table 1. The values obtained are all greater than 50 which demonstrates that our scheme can hold up the requirement of imperceptibility quite well.

**Table 1** PSNR values of the watermarked images

Host image	PSNR
Lena	53.5952
Airplane	53.7048
House	53.7048

### 4.2 Robustness




























Robustness is the effectiveness of the scheme against potential attacks. It is evaluated by subjecting the watermarked image to many attacks and then extracting the watermark. The quality of the watermark is tested using the Normalized Cross-Correlation (NCC).

NCC: It is obtained between the extracted watermark and the original watermark.

$$NCC = \frac{\sum_x \sum_y OW(x, y) \times EW(x, y)}{\sum_x \sum_y [OW(x, y)]^2} \tag{5}$$

where OW is the original watermark; EW is the extracted watermark; (x, y) is the location of a pixel. The closer the NCC value is to 1 the more similar the extracted watermark is to the original watermark. Values above 0.85 can be considered effectual. As can be seen from Table 2, the NCC values of the extracted watermarks obtained via the proposed scheme came out to be greater than 0.85 even after performing some attacks which prove this scheme to be quite resilient.

**Table 2** Extracted watermarks with corresponding NCC values

Attacks	NCC		
	Lena	Airplane	House
Median filtering(3×3)	 (1)	 (1)	 (0.98)
Average filtering(3×3)	 (0.93)	 (0.97)	 (0.97)
JPEG compression(QF=20)	 (0.98)	 (1)	 (1)
Sharpening	 (0.97)	 (1)	 (0.99)
Histogram equalization	 (0.99)	 (1)	 (0.98)
Gaussian blurring	 (0.93)	 (0.97)	 (0.97)
Resizing	 (0.98)	 (1)	 (0.93)
Cropping from center	 (0.99)	 (1)	 (0.98)
Cropping from top left	 (0.99)	 (1)	 (0.99)

**Table 3** Comparison between the NCC values obtained from the suggested method and [15]

Attack type	Parameter	NCC			
		Airplane		Pepper	
		Proposed	[15]	Proposed	[15]
No attack		1	1	1	1
Median filter (window size)	$3 \times 3$	1	0.8675	0.9957	0.7893
Gaussian noise (variance)	0.001	0.9828	0.8207	0.9485	0.8063
	0.002	0.9056	0.8124	0.9099	0.8041
	0.003	0.8841	0.8085	0.8884	0.8797
JPEG compression (quality factor)	90	1	0.7902	0.9871	0.7923
	80	1	0.7933	0.9614	0.8772
	70	1	0.7902	0.9871	0.7880
Salt and pepper (density)	0.001	0.9957	0.9754	0.9914	0.9760
	0.002	0.9871	0.9743	0.9828	0.9732
	0.003	0.9785	0.9715	0.9700	0.9695

### 4.3 Comparative Analysis

The success of the scheme is measured by comparing it with [15]. Robustness is taken as the criteria for measurement which is estimated by the values of NCC of the extracted watermark after attacks such as Gaussian noise, salt and pepper, and more. The corresponding NCC values have been presented in Table 3 with images airplane and peppers taken as hosts for a fair comparison. The proposed method and [15] both give NCC as 1 when no attack is performed. Median filtering with a window size of 3 is performed which results in the proposed method giving better results as NCC values are closer to 1. Gaussian noise with different values of variance, JPEG compression with varying quality factor values and salt and pepper attack with diverse densities—all see the proposed method yield higher NCC values which mean that this method is more robust to attacks as compared to [15].

## 5 Conclusion

In this work, an original digital image watermarking scheme for color images in transform domain has been proposed. Low-frequency DCT coefficients have been used as features for training the LSVR and then embedding the binary watermark in the selected image regions based on the randomness measure of each region. LSVR for watermark embedding is employed and extraction in varied textured and colored images. LSVR has significant learning capability of image features, and its effective generalization ability reveals the successful extraction of watermark even after performing a sequence of attacks. Various images are used in experimental

work to verify the efficacy of the suggested approach and the results obtained clearly signify the strength of the proposed method as compared to existing work. Further improvement can be made upon this scheme to enhance its strength against geometric attacks and find multiple scaling factors for watermark embedding in each selected region using metaheuristic approaches.

## References

1. Bassia P, Pitas I, Nikolaidis N (2001) Robust audio watermarking in the time domain. *IEEE Trans Multimedia* 3(2):232–241
2. Asikuzzaman M, Pickering MR (2017) An overview of digital video watermarking. *IEEE Trans Circ Syst Video Technol* 28(9):2131–2153
3. Borges PVK, Mayer J (2006, May) Document watermarking via character luminance modulation. In: 2006 IEEE international conference on acoustics speech and signal processing proceedings, vol 2, pp II–II. IEEE
4. Shukla D, Tiwari N, Dubey D (2016) Survey on digital watermarking techniques. *Int J Signal Process Image Process Pattern Recogn* 9(1):239–244
5. Araghi TK, Manaf ABA., Zamani M, Araghi SK (2016) A survey on digital image watermarking techniques in spatial and transform domains. *Int J Adv Image Process. Technol IJIPT*, 3(1):6–10. Yorozu Y, Hirano M, Oka K, Tagawa Y (1987) Electron spectroscopy studies on magneto-optical media and plastic substrate interface. *IEEE Trans J Magn Jpn* 2:740–741 [Digests 9th annual conference on magnetics Japan, p 301, 1982]
6. Su Q, Chen B (2018) Robust color image watermarking technique in the spatial domain. *Soft Comput* 22(1):91–106
7. Cedillo-Hernández M, García-Ugalde F, Nakano-Miyatake M, Pérez-Meana HM (2014) Robust hybrid color image watermarking method based on DFT domain and 2D histogram modification. *SIViP* 8(1):49–63
8. Makbol NM, Khoo BE, Rassem TH, Loukhaoukha K (2017) A new reliable optimized image watermarking scheme based on the integer wavelet transform and singular value decomposition for copyright protection. *Inf Sci* 417:381–400
9. Bajracharya S, Koju R (2017) An improved DWT-SVD based robust digital image watermarking for color image. *Int J Eng Manuf* 7(1):49
10. Kumar R, Das RR, Mishra VN, Dwivedi R (2010) Fuzzy entropy based neuro-wavelet identifier-cum-quantifier for discrimination of gases/odors. *IEEE Sens J* 11(7):1548–1555
11. Wang J, Du Z (2019) A method of processing color image watermarking based on the Haar wavelet. *J Vis Commun Image Represent* 64:102627
12. Ko HJ, Huang CT, Horng G, Shih-Jeng W (2020) Robust and blind image watermarking in DCT domain using inter-block coefficient correlation. *Inf Sci* 517:128–147
13. Yuan Z, Liu D, Zhang X, Su Q (2019) New image blind watermarking method based on two-dimensional discrete cosine transform. *Optik* 164:152
14. Patvardhan C, Kumar P, Lakshmi CV (2018) Effective color image watermarking scheme using YCbCr color space and QR code. *Multimedia Tools Appl* 77(10):12655–12677
15. Giri KJ, Peer MA, Nagabhushan P (2015) A robust color image watermarking scheme using discrete wavelet transformation. *IJ Image, Graphics and Signal Process* 1:47–52
16. Wu L, Zhang J, Deng W, He D (2009, December) Arnold transformation algorithm and anti-Arnold transformation algorithm. In: 2009 first international conference on information science and engineering, pp 1164–1167. IEEE
17. Garg P, Dodeja L, Dave M (2019) Hybrid color image watermarking algorithm based on DSWT-DCT-SVD and Arnold transform. In: *Advances in signal processing and communication*. pp 327–336, Springer, Singapore

18. Vapnik V (2013) The nature of statistical learning theory. Springer science & business media
19. Balasundaram S (2010) On Lagrangian support vector regression. *Expert Syst Appl* 37(12):8784–8792
20. Mehta R, Rajpal N, Vishwakarma VP (2018) Robust image watermarking scheme in lifting wavelet domain using GA-LSVR hybridization. *Int J Mach Learn Cybernet* 9(1):145–161

# Leaf Detection Using Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Classifying with SVM Utilizing Claim Dataset



Kazi Sultana Farhana Azam, Farhin Farhad Riya, and Shah Tuhin Ahmed

**Abstract** In the field of science, the area of leaf in progressed pictures plays an exceedingly significant portion inside the field of science. The distinguishing proof of leaf may be an uncommonly basic key to maintain a strategic distance from an overpowering hardship of resign and the sum of rustic things. Our paper broadens the organization approach to arranging and testing images by local binary pattern descriptor on the image and store Histogram of Oriented Gradients is related to the expelled LBP features, then divided into two specific classes: organized images and oriented images, tolerating the support vector machine. The organized illustration on the databases implies rate of revelation generally 95.25% on our very own dataset. The time complexity to boot essentially diminished and the methodology is found to be working well underneath diverse lighting up assortment conditions. Every so often plants and leaves are used to prepare drugs for the resolution of numerous human diseases, and the selection of the accurate leaf and plant is compulsory in the search for precise medication. So the advantage of our project comes here that will help producers of such medicines to locate the specific leaf and plant for the precise preparation of medicines and thereby avoid errors.

**Keywords** LBP · HOG · SVM · Dataset · Train data · Test data · Leaf detection

---

K. S. F. Azam (✉) · F. F. Riya  
Department of Computer Science, Primeasia University, Dhaka, Bangladesh  
e-mail: [tahia.pau@gmail.com](mailto:tahia.pau@gmail.com)

F. F. Riya  
e-mail: [farhinfarhad@gmail.com](mailto:farhinfarhad@gmail.com)

S. T. Ahmed  
Department of Computer Science, North South University, Dhaka, Bangladesh  
e-mail: [tuhin.nsu13@gmail.com](mailto:tuhin.nsu13@gmail.com)



## 1 Introduction

An unused concept of quick developing has been displayed where the field conditions are controlled and checked utilizing the self-operating systems. The self-recognition of leaf is based on the recognizable verification of the characteristics of leaf, so that information of the leaf occasions may well be quickly and absolutely given to the agriculturists, pros, and examiners. Concurring to the leaf the features may alter. The features that are removed from the picture are color, shape, texture, etc. In few cases for displayed patterns, more features are removed and these removed features would increase the overall toll. This helps manipulating the inside complexity and the computation of time. Hence, it is fundamental to diminish the incorporate data. Computational models of leaf recognizing confirmation must address some complex issues. This inconvenience develops from the truth that clears out leaf information to recognize a must specific leaf from all other features to be appeared in a way that best utilizes the available leaf information to recognize classification with various other techniques based on leaf picture which is utilized for this kind of classification comparison, such as cell and molecule techniques [1]. It is conceivable once the effortless exchange of the leaf picture takes place with a computer and a computer can evacuate features from that picture ordinarily by utilizing the picture taking care of procedures [2]. It is not basic to remove and trade those features from the picture to a computer actually. By utilizing a single reasonable calculation alone for recognizing the leaf sample and telling the result to be the finest and productive at recognizing for any and all assortment of the same leaf is a troublesome job. And to legitimize the same calculation is more difficult to differ between diverse leaves [3]. There are numerous investigations that have been done for the leaf acknowledgment with distinctive surface including extraction strategies [4]. We have connected different calculations to the dataset we are working on. Main objective of our proposed method is to develop an identification system of various kinds of leaves using different algorithms in image processing approaches and to design a model that can detect exact plant or leaf to use without any expert's knowledge.

## 2 Literature Review

Different strategies of picture preparing have been created for discovering leaves happening on plants that clears out, stems, injury, etc., by the analysts. The sooner characteristics shows up on the leaf it ought to be identified, recognized and comparing measures ought to be taken to maintain a strategic distance from misfortune. Subsequently a speedy, exact, and less expensive system needs to be made. The analysts have embraced different strategies for location and recognizable proof of leaf character precisely. Utilizing SVM clustering sectioned leaf parcel is obtained. Nearby twofold design (LBP) [5] is fundamentally a surface degree administrator which is grayscale invariant. At to begin with it was found to be capable of including

surface classification, but its application ought to be distant coming as well as inside the ranges of facial expression affirmation, go up against affirmation, etc. In LBP-based portrayals, features can be effectively chosen, combines the shape, the surfaces, and all the stream in consolidate vector. Over the picture to the grayscale the essential step of local binary pattern is changing. In 2005, Triggs and Dalal presented “Histogram of Oriented Gradients” (HOG) [6]. HOG [6] is broadly utilized in different question discovery areas, overwhelmingly in discovery errand of pedestrian. HOG computes in a picture the rise of the slant presentation in an adjacent fix. The thought is to assess the neighborhood concentration and introduction conveyance that may delineate the neighborhood protest shape and appearance [6].

### 3 Methodology

#### 3.1 Feature Extraction

To make a definitive decision and detection of a leaf, extraction of the feature must be performed on both train data and test data. This is the most essential part of our current interpretation and we have used LBP and HOG on our own dataset according to that extraction method.

The framework we have proposed includes three utilitarian squares. The basic work is to choose the input outline from the dataset which is considered. The work-piece is to obtain the features using LBP descriptor and after that applying HOG on the LBP extricated features. The endmost work step is to prepare and constitute classification by the SVM classifier.

#### 3.2 Features

Differing features are chosen to portray particular properties of the clears out. Some clears out are with outstandingly specific shape, and some have especially unmistakable surface plans.

#### 3.3 Texture Analysis

Surface examination basically focuses on the computational talk to a normal insight of the surface and to empower modified planning of the surface. The strategy of surface examination as a run the show produces kind of numeric portrayals. The strategy of computing the surface incorporate is known as incorporate fetching. There

are colossal number of surface examination techniques underneath this category in showing disdain toward the truth that none wins.

Procedures that we have utilized for classification is portrayed here.

### 3.4 *Feature Descriptor*

An include descriptor may be a representation of a picture or a picture fix that streamlines the picture by extricating valuable data and tossing absent unessential information. Typically, a highlight descriptor changes over a picture of measure.

### 3.5 *Histogram of Oriented Gradient (HOG)*

The descriptor HOG concentrates on an object's structure or design. In the context of surface functions, we determine if the pixel is an edge or otherwise. HOG will include the path of the tip, too. This is achieved by eliminating the edges' gradient and orientation. These orientations are measured in segments' localized. It means the entire picture is divided into smaller domains, and the gradients and orientation are determined for each domain. The HOG will eventually produce a unique histogram for each of these domains. The histograms are generated using the gradients and pixel values orientations, hence the term "Oriented Gradient Histogram".

Hog could be a thick included extraction strategy for pictures. Thick implies that it extricates features for all areas within the picture or a locale of intrigued within the picture.

The shape of the structures in the locale is tried to capture by it.

- Let us compute HOG descriptor for the following  $256 \times 256$  image.
- We start with computing  $X$ -gradient and  $Y$ -gradient of the image.
- For each pixel  $(i, j)$ .
- And another matrix with the gradient orientations.

For each pixel  $(i, j)$ .

### 3.6 *HOG Vector*

Now as we start computing the HOG vector, let us start with the first block sized  $16 \times 16$  pixels and we divide this block into  $2 \times 2$  cells sized,  $8 \times 8$  pixels each.

Let us start with first cell sized  $2 \times 2$  pixels (Figs. 1, 2, 3).

In the above graph, the x-axis represents different cell angles and the y-axis represents changes in magnitude for different cell angles. From the graph above, it can said that when the cell angle is in between 0 and 19, the magnitude is around 40



Fig. 1 Block cell

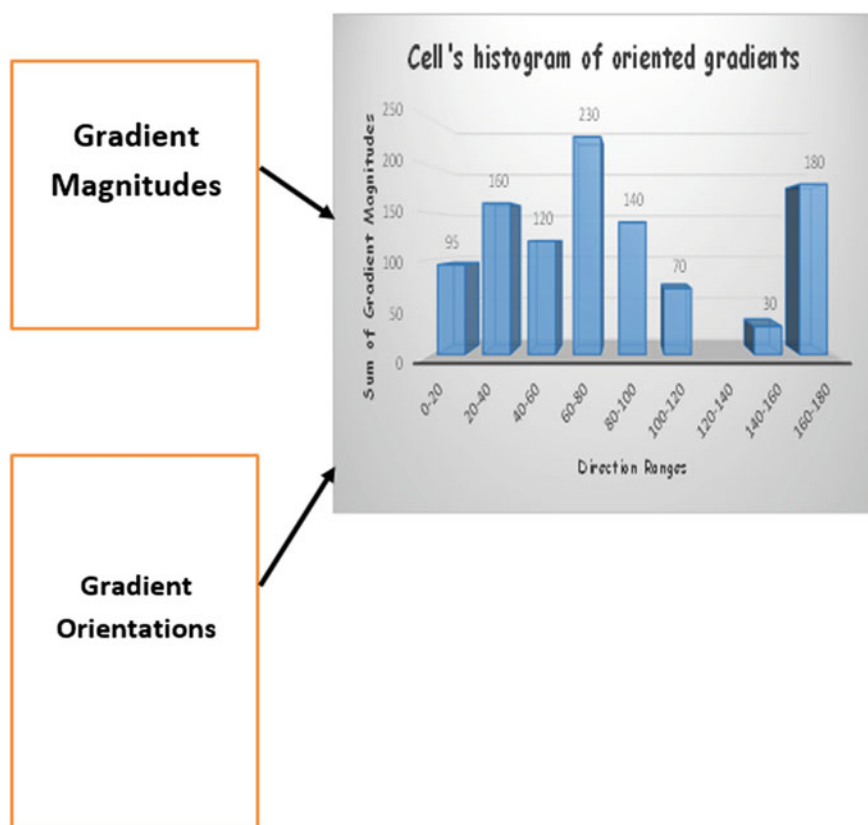


Fig. 2 Cell's HOG

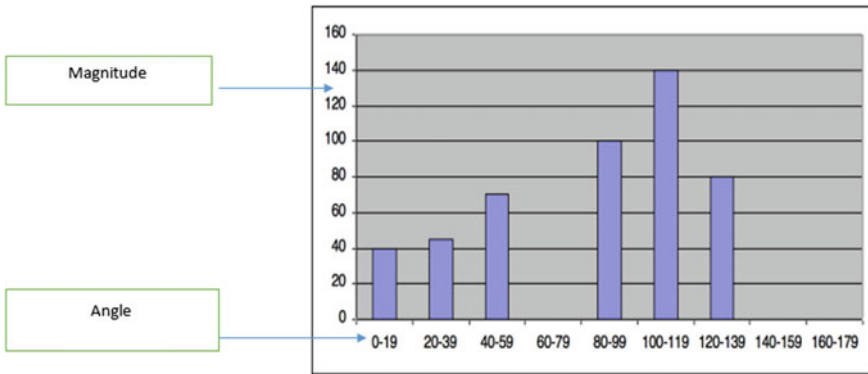


Fig. 3 Magnitude versus angle

and the magnitude keeps increasing with the increasing value of cell angles but the magnitude decreases after it hits to 140. The magnitude is highest when the angle is between 100 and 119 but after that the magnitude falls and for the cell angles between 120 and 139 it is 80 (Figs. 4 and 5).

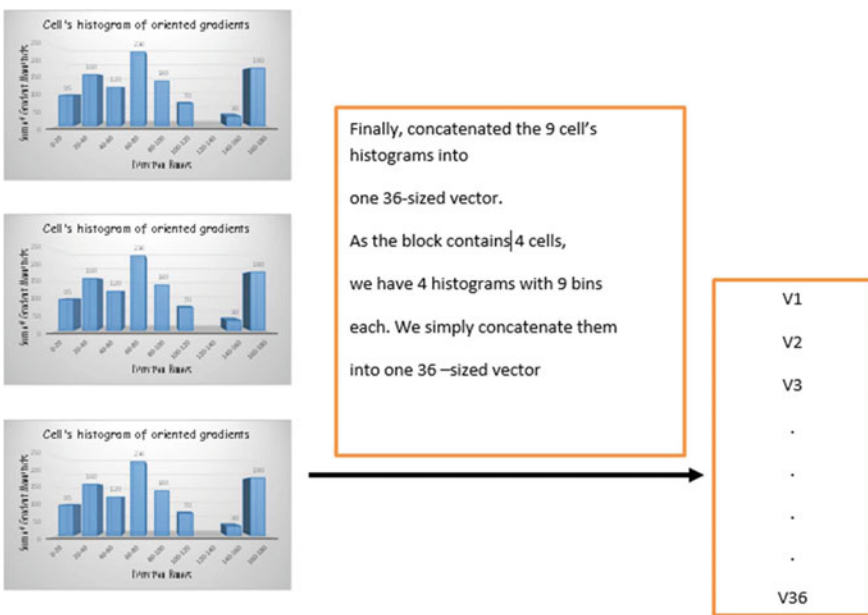


Fig. 4 Concatenations of 9 cell's histogram and 36 sized vector

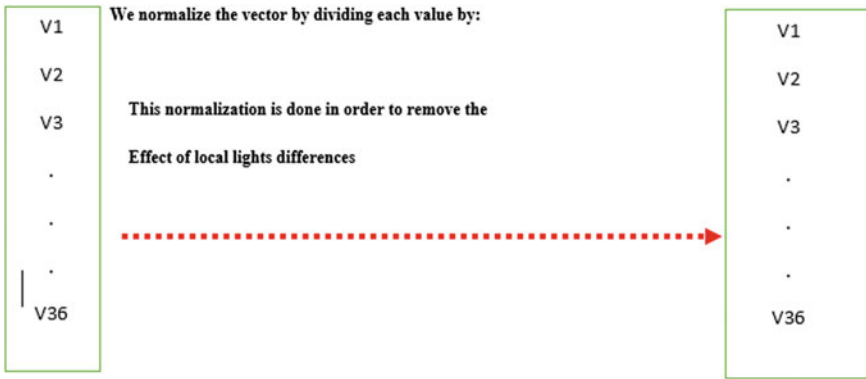


Fig. 5 Normalize the concatenated block’s vector

### 3.7 Local Binary Pattern

The actual LBP operator which is named LBP codes substitutes the pixel value of an image with decimal numbers which encrypts the local structure around each pixel. Every central pixel is contrasted to its eight peers; peers with a lower value than the central pixel will have a bit 0, and other peers with a value significantly higher than the central pixel will have bit 1. Of each provided central pixel, a binary number can be obtained, which is generated by concatenating all such binary bits in a clockwise fashion, beginning from one of its top-left peers. The subsequent decimal value of the derived binary number substitutes the central pixel value. The histogram of LBP labels determined over a region or image is often used as a structure representation of that image. Local binary pattern is an important texture descriptor for images that benchmarks the adjacent pixels based on the existing pixel value.

The LBP descriptor is given for every pixel as

$$LBP(X, Y) = \sum x = 0X - 1e(hx - hd)2x$$

where  $hd$  and  $hx$  denote the intensity of the neighboring and current pixel, respectively.  $X$  is the number of neighboring pixels chosen at a radius (Fig. 6).

### 3.8 Support Vector Machine

In the domain of pattern recognitions, extensive applications have been discovered by the support vector machine [7]. SVM may be a directed machine learning calculation that can be utilized for classifying a strategy called the kernel trick which is used by SVM to convert information and it finds an ideal boundary between the conceivable yield based on these changes. Basically, it does a few amazingly complex information

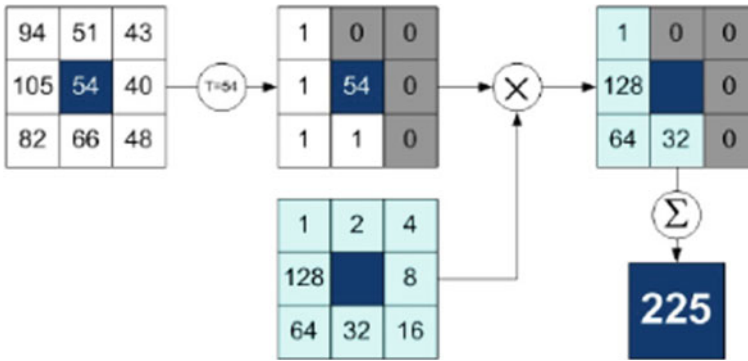


Fig. 6 Working process of local binary pattern

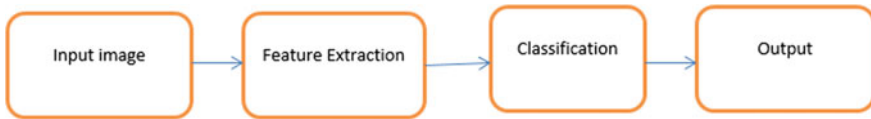


Fig. 7 Proposed method

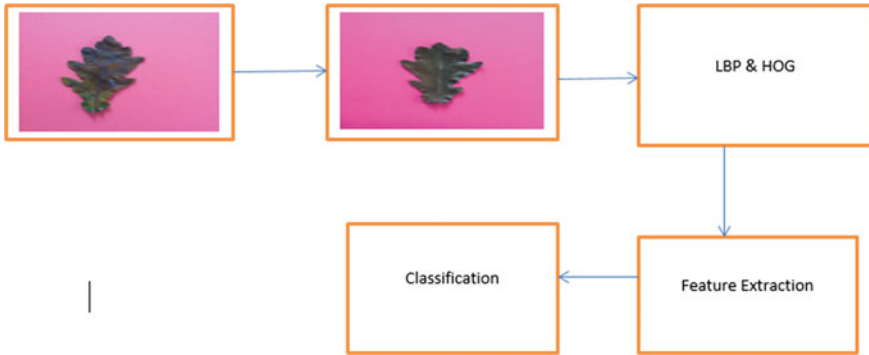
changes, at that point figures out how to separate information based on the outputs [8].

### 4 Applications of SVM

- SVMs are supportive in content and categorization as their application can basically diminish the requirement for the labeled planning events in both the standard settings.
- Classification of pictures can in addition be performed utilizing SVMs. Exploratory comes around to show up that SVMs finish basically higher to see precision than ordinary request refinement plans after reasonable three to four rounds of congruity input. In general, the originality of the picture division systems checks those by utilizing a modified shape of SVM.
- Hand-written characters can be recognized utilizing SVM (Fig. 7).

### 5 Result

Different algorithms have provided different accuracies and their performances were measured. We have calculated the accuracy of each of the algorithms. Now we will be



**Fig. 8** Classification steps in SVM method

representing all the algorithms below that we were interested in. Then we would be able to calculate the accuracy level for each algorithm and put them in comparison. For the aggregate view, we have put all the values in one table. The algorithms we have applied show different performance measures. Support vector machine has provided the highest accuracy (95.25%) with our proposed dataset.

The try is planned to demonstrate the execution of two included extraction strategies, LBP and HOG calculations, for taking off classification reason. The steps of classification are outlined in Fig. 8. With the utilization of the HOG surface highlight fetching, the test starts after changing the colored picture to a gray-level picture. We took 63 images in the training dataset and then categorized those images into 7 distinct groups. Then we arbitrarily took a total of 37 images from each group (Fig. 8; Table 1).

To measure the performance metrics, the use of confusion matrix is useful and convenient. A confusion matrix is introduced to represent the relation between the actual and predicted class (Fig. 9; Table 2).

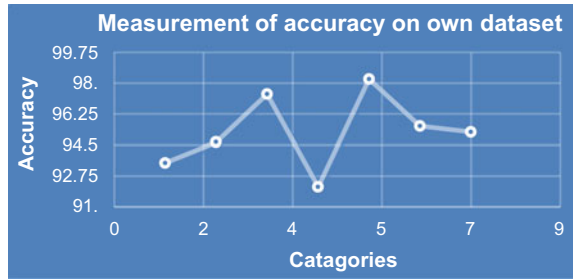
It can be said from the graph above, that the level of accuracy for each group varies from one another. In the very first group, the accuracy level is the lowest which is 93.5. There can be evident a slight fluctuation between grades 2, 6, and 7. Group 5 holds the highest level of accuracy that is approximately 98.23.

**Table 1** Confusion matrix

Actual	Prediction	
	Positive	Negative
Total = TP + TN + FP + FN		
Positive	TP	FN
Negative	FP	TN



**Fig. 9** Accuracy measurement on various categories



**Table 2** Measurement of accuracy

Categories	Accuracy (%)
1	93.50
2	94.67
3	97.38
4	92.15
5	98.23
6	95.57
7	95.23

## 6 Conclusion

The classification that has done recognizing based on the clear out pictures with cleared surface features was proffered and actualized. The LBP and the HOG calculations removed the surface features from the 50 pictures in dataset and with mutilated or advanced leaf pictures for testing. In development, the LBP and HOG technique were utilized and it was found out to be 95.25% exactness. Considering the time of recognizing a picture as one of the foremost classification criteria, it can be said that the HOG is removed prevalently than LBP methodology for making the dataset of image vectors for one time and it will not be the tremendous point to be identified. Our extreme objective is to assist the common community with respect to the clears out discovery. To ensure better commitment in cultivation, we have involved ourselves in this execution to create our nation distant better; a much better; a higher; a stronger; an improved “a much better put for planting.”

## References

1. Graham J (2000) Application of the Fourier-Mellin transform to translation-, rotation and scale-invariant plant leaf identification. McGill University, Montreal
2. Wu SG, Bao FS, Xu EY, Wang Y, Chang Y, Xiang Q (2007, July 29) A leaf recognition algorithm for plant classification using probabilistic neural network. arXiv, 0707, 4289v1, [cs.AI]

3. Wang Z, Chi Z, Feng D, Wang Q (2000) Leaf Image retrieval with shape features. In: Laurini R (ed) *Visual 2000*, LNCS 1929, pp 477 – 487
4. Tzionas P, Papadakis SE, Manolakis D (2005) Plant leaves classification based on morphological features and a fuzzy surface selection technique. In: *Fifth international conference on technology and automation*, Thessaloniki, Greece, pp 365–370
5. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28:2037–2041
6. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: *IEEE computer society conference on Computer vision and pattern recognition*, vol 1, pp 886–893. IEEE
7. Hsu C-W, Chang C-C, Lin C-J (2003) A practical guide to support vector classification, pp 1396–1400
8. Li G, Wu Q, Tu D, Sun S (2007) A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD. In: *2007 IEEE International conference on multimedia and expo*, pp 1750–1753

# Prediction of Congestive Heart Failure (CHF) ECG Data Using Machine Learning



Manthan S. Naik, Tirth K. Pancholi, and Rathnakar Achary

**Abstract** Machine learning can learn a complex system using a large amount of data and has the potential in predicting critical medical emergencies like congestive heart failure (CHF). In this paper, a machine learning algorithm is proposed for predicting CHF using convolutional neural network (CNN). The system uses a VGG16 model, which is dependent on ImageNet and utilizes ECG signal. The CHF information is obtained from the Beth Israel Deaconess Medical Centre (BIDMC) CHF dataset and also the typical dataset from the FANTASIA database. Various models of CNN were prepared and tested on the dataset to obtain the information about CHF or normal for a patient until the required accuracy is achieved. The performance is assessed based on the accuracy of the code and precision. The ECG signal of a patient recorded for 24 hours is utilized in this research. The experimental result obtained by this research has given accuracy of 100% on applying the VGG16 Model to the dataset.

**Keywords** Congestive heart failure (CHF) · ECG · CHF prediction · VGG16 · Convolutional neural network (CNN) · Deep learning

## 1 Introduction

Among the many heart diseases, CHF is one of the chronic diseases, which affects the normal function of the heart. Generally, this is referred to as heart failure. The CHF specifically referred to a stage in which there is a change around the heart and causes it to pump the blood inadequately. CHF cannot be cured, as a result, there is an adverse effect on the quality of life of the peoples suffering from it. Since there is no definitive

---

M. S. Naik · T. K. Pancholi · R. Achary (✉)  
Alliance College of Engineering and Design, Alliance University, Bengaluru, India  
e-mail: [rathnakar.achary@alliance.edu.in](mailto:rathnakar.achary@alliance.edu.in)

M. S. Naik  
e-mail: [naikmanthan580@gmail.com](mailto:naikmanthan580@gmail.com)

T. K. Pancholi  
e-mail: [tirthkamleshpancholi602@gmail.com](mailto:tirthkamleshpancholi602@gmail.com)

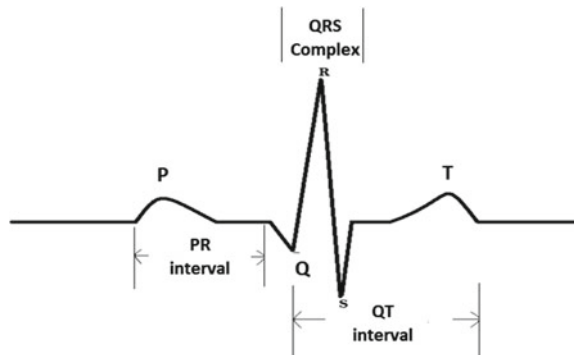
diagnostic test for CHF, in most of the cases, the patients are advised based on the observation and results of physical examination, radiograph, echocardiography, and electrocardiogram. Out of these multiple techniques, the electrocardiogram (ECG) is considered as the best methods to detect the heart-related problem in a patient.

The ECG waveform obtained from a patient suffering from CHF is divided into multiple complex waves as in Fig. 1. It is a combination of four different waves represented as  $P$ ,  $Q$ ,  $R$ ,  $S$ , and  $T$ . The complete ECG signal is divided into three sections as  $P$  wave,  $QRS$  complex, and  $T$  wave. To analyze the heart function of an individual, the ECG signal is examined. The parameters considered for the verification are shape, the duration, and the relationship with each other between the  $P$  wave,  $QRS$  complex, and  $T$  wave components and  $R$ - $R$  interval.  $P$  wave is the first wave that indicates the excitation of the heart, which is due to the atrial depolarization. It is originated in the sinus mode and then propagated to the atrium. The depolarization of the right atrium is indicated by the first left half and for the left atrium by the other half.  $QRS$  is a combination of three waves, in which  $Q$  is a down-ward wave and the high tip vertical wave is  $R$ . It is triangular in shape and is about 2.5 mV amplitude and continued for a period of 40 mS. The down-ward wave immediately after  $R$  wave is known as  $S$  wave. The combination of these three waves is responsible for the excitation of the ventricular electrical process. The left and right ventricular depolarization processes are due to the  $QRS$  complex section.  $T$  wave section of the ECG is of much lower amplitude and longer period generated by ventricular repolarization. Let us consider the dynamic model of the ECG to obtain its quasi-periodic property.

The  $P$ ,  $Q$ ,  $R$ ,  $S$ , and  $T$  points on the ECG are considered by a set of exponential functions. These points are denoted by five angles as  $\theta_P$ ,  $\theta_Q$ ,  $\theta_R$ ,  $\theta_S$ ,  $\theta_T$ . The dynamic equation of motion and the movement of the trajectory around an attracting limit cycle is represented by a set of three differential equations as;

$$\dot{x} = ax - wy, \quad \dot{y} = ay + wx$$

**Fig. 1** Components of ECG signal



$$z = \sum_{i \in \{P, Q, R, S, T\}} ai \Delta\theta i \exp\left(\frac{-\Delta\theta i^2}{2bi^2}\right) - (z - z_o) \text{ where } \alpha = 1 - \sqrt{x^2 - y^2}$$

$$\Delta\theta = (\theta - \theta i) \text{ mod } (2\pi), \theta = \alpha \tan 2(y, x)$$

Are the arctangents with  $-\pi \leq a \tan 2(y, x) \leq \pi$  and  $w = 2\pi f$  denotes the angular frequency of the signal. The magnitude of the peaks and the width of each peak are controlled by the parameters  $\alpha_i$  and  $b_i$ , respectively. The introduction of baseline phenomenon by coupling a baseline value  $z_0$  to  $f_2$ (respiratory frequency) using  $z_o(t) = A \sin(2\pi f_2 t)$  where  $A = 0.15$  mV. Solving this numerically yields the output for an ECG signal  $s(t)$  as  $s(t) = z(t)$ , which represents the vertical component of the 3D ordinary differential equation.

For analyzing the ECG of a patient suffering from CHF, its noise must be filtered. The inherent noise in a raw ECG signal is filtered by using any suitable filters such as extended Kalman filter. In this paper, a failure detection model is developed using ECG signals to indicate the criticality of the issue to an individual. This predictive model advises the individuals to know about the criticality of their heart functions. The given ECG dataset of the patients is analyzed and given to an algorithm for classifying the data to represent the patient suffering from CHF or normal, by using machine learning techniques [1, 2].

The other section of this research paper is structured as follows: In Sect. 2, the research work on CHF is analyzed and related work of other authors about the CHF detection techniques and their performance. The proposed technique of CHF detection and its implementation methodology is in Sect. 3. The result analysis of the proposed algorithm and its performance is in Sect. 4 and the conclusion in Sect. 5.

## 2 Related Works

Many researchers have published their works on CHF detection, using different machine learning algorithms and observed the best accuracy by using an SVM with an accuracy of 97%. In [3], the authors performed the detection of CHF by using an artificial neural network with an accuracy of 92% [4]. The authors in [5] proposed the detection methodology by power spectral densities analysis for the detection of CHF in patients, with an accuracy of around 90%. This was not very successful as there was already a system with 92% accuracy.

The authors in [6] propose the use of SVM algorithm with only 15 features provided an accuracy of 97% for classifying the patient with CHF. The main constraint of this method is its complexity in handling a large amount of minor data. Hence, there is a challenge in handling a large volume of data that may be needed for an improvement in the system and to study the various changes in the ECG signal of a person suffering from CHF. This disadvantage can have a major issue while handling large quantities of data. Large amounts of data might have, but

the system may not work properly. Also, another issue is that the accuracy achieved was only 97%. So, with the increase in data, it might not work with the same amount of accuracy.

Different models are considered for analysis, among these the support vector machine and random forest are the nonparametric models. The complexity of these models increases as the number of training samples increases. Training a nonparametric model consumes more processing resources, hence using them becomes more expensive than any generalized linear model, because a random forest becomes more expensive as the number of trees increases. In case of support vector machines also, there are as many numbers of support vectors in the training set. An alternative for this is a multiple class support vector machine, its implementation for multiple class-classification is one-vs-all [7]. This represents that each class is trained by a support vector machine, whereas the random forest and decision tree can handle multiple classes out of the box [8]. Using  $k$ -nearest classifier, the accuracy is 96.39%, using a value of  $k$  as 5 or 7 with 8–10 features [9].

### 3 Methodology

In this paper, a framework has been developed to analyze whether a patient is experiencing congestive heart failure or not by utilizing a basic ECG signal and VGG16 CNN model [10]. The dataset of the ECG signal is obtained from 60 different samples, which is preprocessed by wavelet analysis to obtain the .csv files [11]. The sample ECG data for a patient in .csv form is shown in Fig. 2. The dataset

Fig. 2 Sample ECG in .csv form

	A	B		
1	0.635	-0.64	1	8.46
2	0.68	-1.28	2	8.592
3	0.63	-1.92	3	8.352
4	0.57	-2.535	4	8.152
5	0.535	-2.8	5	8.34
6	0.51	-2.9	6	8.528
7	0.515	-2.99	7	8.312
8	0.52	-3.03	8	8.084
9	0.525	-3.025	9	8.248
10	0.495	-2.975	10	8.444
11	0.43	-2.825	11	8.308
12	0.275	-2.565	12	8.084
13	0.195	-2.305	13	8.184
14	0.2	-2.05	14	8.42
15	0.26	-1.835	15	8.34
16	0.355	-1.695	16	8.1
17	0.37	-1.59	17	8.132
18	0.315	-1.465	18	8.356
19	0.155	-1.29	19	8.38
20	-0.035	-1.075	20	8.136

ECG in .csv form for a patient with CHF

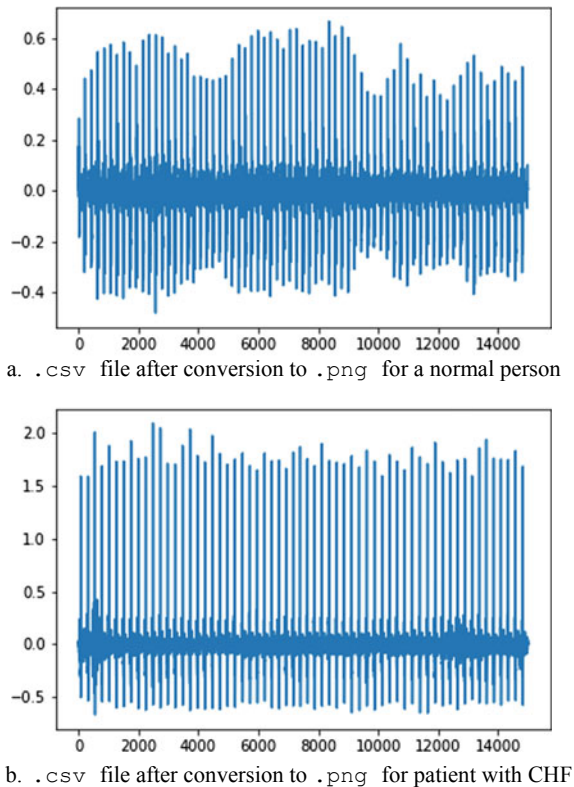
ECG in .csv form for a normal person

obtained for our analysis is from FANTASIA database and BIDMC CHF database of `archive.physionet.org`, which provide the data samples of 40 normal peoples of different age groups and also data of the patients suffering from CHF in `.csv` file.

### 3.1 Data Preprocessing

This is a major step in any machine learning algorithm. A `.dat` file was obtained which was then converted to `.csv` by using the `physionet` convertor. Generally, a file contains a lot of noise and they must be removed in this stage. There are 60 such files considered for this procedure. The reduction of the features was done and only the ECG value was maintained and the rest of them are omitted. After filtering all these unnecessary data, a proper file containing only the ECG values was obtained. Each file generates one ECG signal which was converted to an image in `.png` form as shown in Fig. 3. This file is generated by plotting the ECG using a python code to get a `.png` picture of the same data. This is then used for prediction.

**Fig. 3** `.csv` file after conversion to `.png`



### 3.2 Training and Testing

The training process is also known as fitting the model. It is carried out on the labeled data to make the machine to learn the algorithm for classification by itself [7]. Generally, a very large amount of data is feed to the model during the training process. For the best result, the ratio between the train data to test data is about 60:40 or 80:20. In this research, this ratio has been considered as 60:40. There are two methods for training the model in Keras:

- Using `fit()`: This function is used when there is less amount of data to be trained, and the data are categorized in  $x$  and  $y$  labels.
- Using `fit_generator()`: This is used when a large amount of randomly generated data is used. Generally, developers prefer `fit_generator()` method because it is a very practical approach for implementation.

The training was done using the `fit_generator()` function. After training the model on the labeled data, the model has to be tested to assess the data, predicts well on the test cases. `Predict_generator()` function was used to test the model on several test images or test data.

### 3.3 Training of VGG16 Model

VGG16 is the model for image classification. This pre-trained model consists of 16-layer architecture as in Fig. 4, which is perfectly suitable for the classification of over 1000 categories of images [12]. Last layer (dense) of this architecture is modified to give only two outputs (as “CHF” or “NORMAL”) instead of inbuilt 1000 outputs. A random image was chosen with target size (224, 224) because VGG16 gives output in this size only. Then the image was converted to an array of pixels which can then be predicted using the trained model. The `relu()` and `softMax` convert the image into fractions. Because of the `softmax` in the final classification layer, `predict()` function is used to give the probabilistic value of the classification [7].

In the test set as in Fig. 5, the value 1.0 represents the patient is suffering from CHF and 0.1 for normal.

Finally, an `if-else` logic is applied in such a way that if the prediction value is more than or equal to 0.5 then it will display “CHF” as output along with the probabilistic value, and “NORMAL” if another way around. As there were only two categories to be classified, the value of the results was kept as 0.5, and the results were appealing good. The prediction value was 98% (almost perfect prediction), and it shows CHF. VGG16 found to be the best model to fit the dataset for classification among all five models considered for comparison which were implemented because of the `LOS (CATEGORICAL_CROSSENTROPY)` and almost near to zero (0.1646), along with an accuracy of 100% [13].



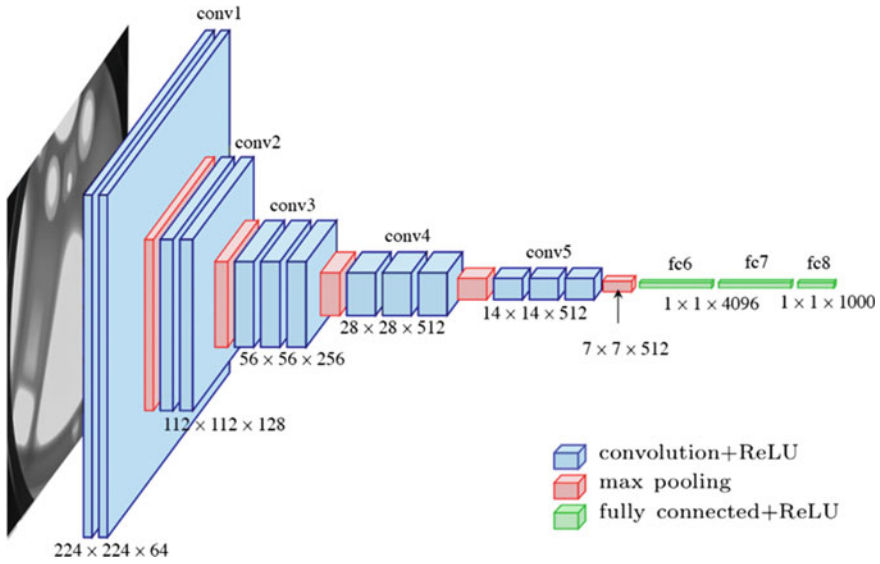


Fig. 4 VGG16 architecture

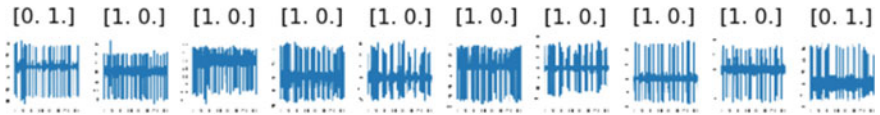


Fig. 5 Test set to represent CHF or normal

### 3.4 Testing

The algorithm is tested by using a combination of BIDMC CHF dataset and FANTASIA normal database. The data has been split into train, test, and valid directory. All of them having CHF as well as normal data. Upon entering an image, our model predicts if the person has CHF or not with an accuracy of 100%. Some ECG signals of normal and CHF patients have been set aside that were used in the prediction model [1]. The prediction model gave a correct output with an accuracy of 100%. Using the trained model for prediction, 100% accuracy is obtained on the training data. Testing sample data is also classified with 100% accuracy and Validation accuracy was 93%.

**Table 1** Accuracy of different models tested

Model	k-nearest neighbour	SVM	11-layer CNN model	AlexNet CNN model	VGG 16 CNN model
Accuracy	96.39%	97.59%	60.00%	56.67%	100.00%

## 4 Results Analysis

The 16 layers of the VGG16 models help us to classify the images with the best possible result. An accuracy of 100% is attained on the test data and a validation accuracy of 93%. The trained model is then tested with the test set. After the completion of this procedure, there is another model. This model is the prediction model. An ECG signal is added to the model after converting to an image. The results are generated by feeding one random image file to `predict()` method, which gives the classified output based on the processed image's probabilistic value, i.e., if there are two categories of classification such as "a" and "b" and its probabilistic value is more than 0.5 then it will display as "a" else "b". The `softmax` function was used to convert the data to images and probabilistic values in the VGG16 model. Table 1 listed the accuracy obtained from each algorithm.

## 5 Conclusion

The proposed model consists of VGG16 architecture which fits both training and testing data well. Over-fitting is not shown when 16 layers architecture is used for 20 epochs over the training data. The accuracy over training data is 100% and the validation accuracy is 93%, which can be improved by providing more data to the algorithm and validation accuracy can be increased further by using the system with higher graphics processing unit (GPUs), with better benchmarking. Most of any machine learning implementation is about the data preprocessing. A large amount of data is required to make the system to perform even better. For future enhancement, a system can be built to predict, if a person will suffer from CHF in the future by learning the pattern of his ECG signal over a period of time.

## References

1. Ibaida A, Khalil I, van Dhiah AS (2010) Embedded patients confidential data in ECG signal for health care information system. In: Proceedings of the 32nd annual international conference of the IEEE EMBS Buenos Aires Argentina 31, pp 3891–3894
2. Chen S-T, Guo Y-J, Huang H-N, Kung W-M, Tsung K-K, Yu S-Y (2014) Hiding patients confidential data in the ECG signal via a transform-domain quantization scheme. Springer Science and Business Media New York

3. Clerk Maxwell J (1982) A treatise on electricity and magnetism, vol 1, 3rd edn. Clarendon, Oxford, pp 68–73
4. Automatic localization of casting defects with convolutional neural networks available from: [https://www.researchgate.net/figure/fig-A1-The-standard-VGG-16-network-architecture-as-proposed-32-note-that-only\\_fig3\\_322512435](https://www.researchgate.net/figure/fig-A1-The-standard-VGG-16-network-architecture-as-proposed-32-note-that-only_fig3_322512435). (12 Aug 2019)
5. Hossen A, Al-Ghunaimi B (2008) Identification of patients with congestive heart failure by recognition of sub-bands spectral pattern. World Academy of Science, Engineering and Technology, pp 21–24
6. Hossen A, Al-Ghunaimi B (2008) Identification of patients with congestive heart failure by recognition of sub-bands spectral patterns. World Academy of Science, Engineering and Technology, pp 21–24
7. Application of computational intelligence techniques for cardiovascular diagnostics—scientific figure. [https://www.researchgate.net/figure/The-components-of-the-ECG-signal\\_fig1\\_24830806](https://www.researchgate.net/figure/The-components-of-the-ECG-signal_fig1_24830806). Accessed 11 Aug 2019
8. Nambakhsh MS, Ahmadian A, Ghavami M, Dilmoghani RS, Karimi-Fard S (2006) A novel blind watermark of ECG signal on medical Image using EZW algorithm. In: 28th annual international conference of the IEEE engineering in medicine and biological society (EMBS's)
9. Yu SN, Lee M-Y (2012) Conditional mutual information—based feature selection for congestive heart failure recognition using heart rate variability. *Computer Methods Programs Biomed* I(8):299–309
10. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN towards real-time object detection with region proposal network. In: *Advances in neural information processing system*, pp 91–99
11. Engin M, Cindam O, Engin EZ (2005) Wavelet transformation (ECG). *J Med Syst* 29:589, 594
12. Russakovsky O et al (2015) ImageNet large scale visual recognition challenge. *Int J Comput Vis* 115(3):211–252
13. Isler 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

# Aspect-Based Unsupervised Negative Sentiment Analysis



Mainak Ghosh, Kirtika Gupta, and Seba Susan 

**Abstract** Twitter is a social media platform where users post their opinions on various events, products, services and celebrities. Automated analysis of these public posts is useful for tapping into public opinion and sentiment. Identifying negative public sentiment assumes importance when national security issues are at stake or when critical analysis of a product or policy is required. In this paper, a method is introduced that classifies tweets based on their negative content, without any prior training. Specifically, an unsupervised negative sentiment analysis is presented using an aspect-based approach. Phrase and keyword selection criteria are devised after identifying fourteen valid combinations of part-of-speech tags listed in a prioritized order, that are defined as phrase patterns. A sliding text window is passed through each sentence of the tweet to detect the longest valid phrase pattern. The keyword indicating the aspect information is detected using a dependency parser. SentiWordNet lexicon is used for scoring the terms in the detected keyword and phrase combination. The scores are summed up for each sentence of the tweet and transformed nonlinearly by a modified sigmoid function whose output is in the range  $[-2, 2]$ ; this value comes out to be negative for negative tweets. The utility of our method is proved by superior results as compared to the state of the art on the benchmark SemEval 2013 twitter dataset.

**Keywords** Sentiment analysis · Negative sentiment · Aspect-based sentiment analysis · Unsupervised method · Dependency parser

## 1 Introduction

There is a continuous growth of textual content in social media platforms which has led to research on finding automatic ways for processing this information. Analysing the sentiments of social media posts allows us to understand public opinion behind

---

M. Ghosh (✉) · K. Gupta · S. Susan  
Department of Information Technology, Delhi Technological University, Bawana Road, Delhi  
110042, India  
e-mail: [mainakgh1@gmail.com](mailto:mainakgh1@gmail.com)

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021  
J. Hemanth et al. (eds.), *Intelligent Data Communication Technologies and Internet of Things*, Lecture Notes on Data Engineering and Communications Technologies 57,  
[https://doi.org/10.1007/978-981-15-9509-7\\_29](https://doi.org/10.1007/978-981-15-9509-7_29)

335

crucial topics [1]. Sentiment analysis involves the classification of tweets as negative, positive and neutral. In this paper, exclusively focusing on negative sentiment analysis because, in recent times, an increase in the number of online posts that promote hatred and discord in society is observed.

There are two types of learning methodologies employed for sentiment analysis, namely supervised and unsupervised. Examples of supervised machine learning tools are naïve Bayes, logistic regression, support vector machines (SVM) and long short-term memory (LSTM) [2–5]. Examples of unsupervised methods are fuzzy logic-based approaches that interpret the SentiWordNet scores of tweets to make a decision [6, 7], and aspect-based analysis [8–11] which is the approach undertaken in this paper. A brief discussion on aspect-based works in the literature is included in Sect. 2 prior to introducing the proposed model that is based on deriving phrase patterns and keywords from tweets and scoring them with the help of the SentiWordNet lexicon. This approach generally involves determining the head aspect or issue subject of opinion phrases in a sentence. The organization of this paper is as follows. The proposed work is introduced in Sect. 2, the experimentation and the results are discussed in Sect. 3, and the conclusions are given in Sect. 4.

## 2 Proposed Work

The aspect-based analysis aims to determine the subject issue or head aspect in the given text to understand the topic of discussion. One of the significant works in this regard is that of Canonical Conditional Random Fields (CRF) [12] that locates an opinion phrase in every sentence where the head aspect occurs. However, it is not necessary that all opinion phrases, detected in this manner, would contribute to useful information. Some other related works are: [13] that extracts subjective verb expressions, and [14] that extracts keyphrases in a supervised manner. Mukherjee [11] tried to work around the shortcomings by adopting a feature-based approach for representing the head aspect. The semantic features derived in the generative phase are fed into the discriminative phase for issue-based sentence classification. A labelled dataset of aspect-based opinion phrases was generated for the purpose. In [8], the sentiment score of a sentence was obtained by averaging the individual sentiment scores of each aspect at the clause level. This was done under the assumption that a single sentence may express different sentiments towards different aspects of the movie. The focus in [9] was to determine user sentiments of different aspects of a product from tweets. A method that finds special mention is the unsupervised aspect analysis for sentiment detection in [10] since the theme is closely related to our method. They used a specially labelled dataset with annotations of both aspect and sentiment. The top-ranked words for each aspect were shortlisted based on the probabilities generated by a latent Dirichlet allocation (LDA) model. A polarity score was assigned to each noun–adjective pair to compute the overall sentiment. In our aspect-based approach, a sentiment-scoring function for tweets is used that is explained in more detail below.

### 2.1 Process Flow for the Negative Sentiment Analysis Task

The block diagram in Fig. 1 shows an overview of the generic negative sentiment analysis task for the SemEval 2013 twitter dataset. Text preprocessing, feature extraction and sentiment computing are the major chunks of the generic negative sentiment analysis task as noted from Fig. 1. The text preprocessing shown, prior to the feature extraction, is common to our task. Synsets from the SentiWordNet lexicon [15] provide positive, negative and neutral scores for each word in a tweet. The maximum of these three scores has been taken for each word assuming that the maximum score reflects the real nature and context of a word. The labels of tweets have been considered as ‘negative’ and non-negative’ with the ‘neutral’ category considered as ‘negative’. The first phase is the preprocessing phase, and the first step is slang and abbreviation removal from each sentence. A list of abbreviations along with their full forms is maintained. This list was taken from the slangs and abbreviation list of webopedia and can be accessed online at [16].

Each tweet is scanned for words in the abbreviation list that are replaced by its full form. Once all the abbreviations are replaced, each word is converted to lower-case followed by tokenizing the text. The next step is spelling correction. The *symspell* checker using the *symspell* inbuilt dictionary [17] with a maximum distance of 2

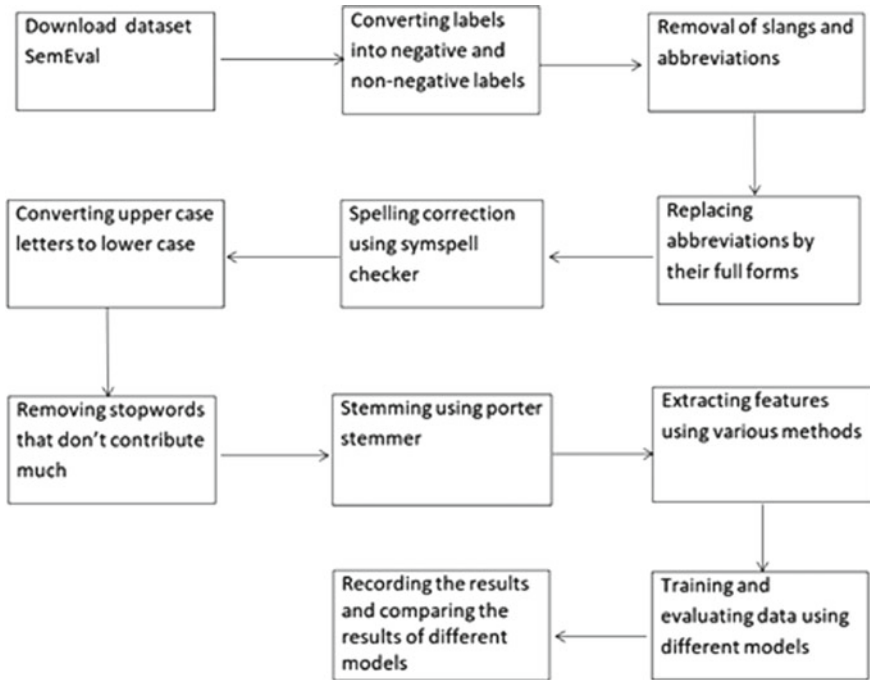


Fig. 1 Process flow of the negative sentiment analysis task

**Table 1** Our fourteen POS tag categories for detecting sentiment phrases

1	VERB
2	NOUN
3	ADJECTIVE
4	ADVERB + NOUN
5	ADVERB + ADJECTIVE
6	ADVERB + VERB
7	MODAL + ADJECTIVE
8	MODAL + VERB
9	ADJECTIVE + NOUN
10	ADJECTIVE + ADJECTIVE
11	ADVERB + ADVERB + ADJECTIVE
12	ADVERB + ADVERB + VERB
13	VERB + ADVERB + ADVERB
14	COORDINATING CONJUNCTION

is used, to edit changes in the spelling. The words are then tagged along with their part-of-speech (POS) tags using the NLTK library [18]. POS tags that do not give any sentiment information are identified and remove them from our text corpus. These tags are identified as ['EX', 'FW', 'LS', 'NNP', 'WP', 'WP\$', 'NNPS', 'POS', 'PRP', 'PRP\$', 'TO', 'WDT', 'WRB'] according to [19]. The remaining POS tags are the ones that are based on which the sentences can be scored. An aspect is a subject that is spoken about in the sentence. For example, the sentence 'The food was very good but the ambience was not very nice', here the aspect is 'food' and 'ambience' and the words describing them are 'very good' and 'not very nice', respectively. Since a single sentence contains both positive and negative sentiments, a prioritized scoring system is devised by identifying 14 types of phrases containing combinations of POS tags ADJECTIVES, VERBS, NOUNS, MODALS and ADVERBS, as shown in Table 1.

The idea is to capture the above phrases in a sentence and then score it. A sliding text window is iterated over the tokenized sentence until the variable token reaches the last token in the sentence. There is a possibility that multiple tag patterns may be associated with the same keyphrase in the given sentence. If more than one pattern in Table 1 exists in a sentence then the longest one is chosen, as the phrase 'not very nice' in the context of 'ambience' keyword, in the example cited above. Preference is given in the order from bottom to top in Table 1. In case, the length and preference order of two patterns is the same, and the first pattern is chosen.

For example, consider the sentence *She was not very good at basketball*. The phrase patterns derived are—(i) *very good* → ADVERB+ADJECTIVE (ii) *not very good* → ADVERB + ADVERB + ADJECTIVE. Since both (i) and (ii) are valid tag patterns present in Table 1, our system selects the longest possible tag pattern in the sentence, i.e. ADVERB + ADVERB + ADJECTIVE (*not very good*) as the phrase pattern to be scored. The SentiWordNet scores of the terms (*not, very, good*)

are multiplied to get a sentiment score for the phrase pattern which comes out to be a negative value due to the presence of *not*.

A corpus of verbs, adverbs, modal, conjunctions, noun and adjectives were compiled using wordnet [20]. An ordered dictionary was kept for each category of verbs, adverbs, modal, conjunctions, noun and adjectives so that the searching can be performed faster. The category of a word is identified by examining the allowed POS tags in that category. For POS tags have been referred to the English Penn Treebank [21]. The allowed POS tags for each category are compiled in Table 2 for verbs and Table 3 for adjectives and adverbs, respectively.

The dependency parser [22] has been used to get the syntactic dependency between the tokens in each statement. The resulting aspects need to belong to this class of dependency parsing tags shown in Table 4. For more reading, the readers are referred to [23] that enlists examples for these dependency parsing tags. Finding links between tokens and relationship-tagging gives more meaningful classification than single keyword-based classification systems [24] or bag-of-word features where the order in which words occur is not known [25]. A sentence when processed by the dependency parser may detect zero or more of these dependency tags.

The preference of selection of these tags, as the keyword, is from top to bottom in Table 4. Further sub-categories or extended POS tagging for each of these dependency tags are kept in our accept states as shown in Table 5. The keyword that matches a category in both Tables 4 and 5 qualifies as the aspect information and is included for sentiment-scoring along with the detected phrase.

**Table 2** Acceptable POS tags for the verb category

VB	BASE FORM OF A VERB
VBD	PAST TENSE OF BASE FORM OF VERB
VBG	PRESENT PARTICIPLE OF BASE FORM OF VERB
VBN	PAST PARTICIPLE VERB
VBP	VERB, NON-3RD PERSON SINGULAR PRESENT
VBZ	VERB, 3RD PERSON SINGULAR PRESENT
MD	MODAL AUXILIARY (could, cannot etc.)

**Table 3** Acceptable POS tags for the adverb and adjective categories

JJ	ADJECTIVE
RB	ADVERB
RBR	COMPARATIVE ADVERB
RBS	SUPERLATIVE ADVERB



**Table 4** Acceptable syntactic dependency tags

nsubj	NOMINAL SUBJECT
dobj	DIRECT OBJECT
iobj	INDIRECT OBJECT
pobj	PREPOSITION OBJECT

**Table 5** Results on the SemEval 2013 dataset

Model	Accuracy (%)	Precision	Recall	F-score
Logistic regression + bag of words [28]	52	0.29	0.65	0.4066
SVM+ bag of words [29]	49.6	0.51	0.49	0.487
SVM-NGRAM (mix of uni- and bi-gram) [30]	57.5	0.58	0.49	0.51
LSTM with embedding, convolution layer and max pooling [31]	77.5	0.5	0.5	0.5
Aspect-based method (unsupervised) [ours]	80.0	0.67	0.535	0.52

## 2.2 Sentiment-Scoring of Tweets

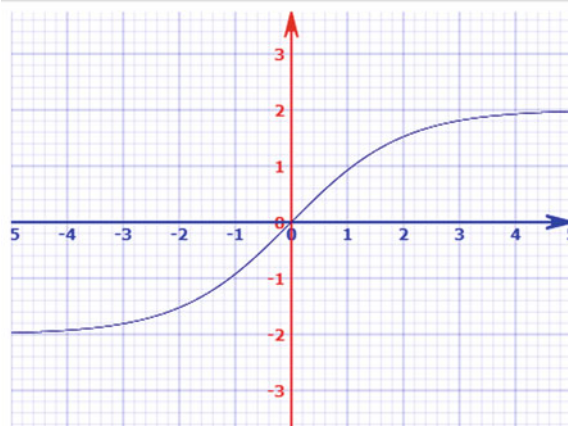
Once our phrase pattern from Table 1 and the keyword from Table 5 have been selected, the phrase + keyword combination for sentiment-scoring and classification is used. *Synsets* from the SentiWordNet lexicon provide us with the positive, negative and neutral scores for each word in the phrase + keyword combination. The maximum of these three scores has been taken for each word assuming that maximum score reflects the real nature and context of a word. The polarity is (+) for positive scores and (−) for negative scores. The labels have been considered as ‘negative’ and non-negative’ with all non-positive labels including ‘neutral’ labelled as ‘negative’. The procedure of sentiment-scoring for keywords is the same as for the phrases. The software implementation of our code is available for reference at <https://github.com/mkaii/bug-/tree/gh-pages>. All valid scores derived from sentences in the twitter post are added up to form the variable  $x$  that is nonlinearly transformed into a decision variable  $f(x)$  as explained below.

The polarized SentiWordNet score is used to replace each word of our phrase + keyword combination. The summation of all these scores, denoted by  $x$ , is fed as input into our modified sigmoid function shown in Eq. (1).

$$f(x) = \frac{4}{1 + e^{-x}} - 2 \quad (1)$$

The sigmoid function in Eq. (1) is plotted in Fig. 2. The range of the sigmoid function in (1) is from  $-2$  to  $2$ . When the sentiment score  $x$  is  $0$ ,  $f(x) = 0$ . If the score  $x$  is greater than  $0$ , then  $f(x) > 0$  and the overall sentiment is determined to be positive. Likewise, if the score  $x$  is less than or equal to  $0$ , then  $f(x) \leq 0$  and the overall sentiment is determined to be negative. So the final sentiment score of any

**Fig. 2** Sigmoid function  $f(x)$  versus  $x$



phrase + keyword combination would be between  $-2$  (for negative sentiment) and  $2$  (for non-negative sentiment).

### 3 Experimentation and Results

The dataset that has been used in this project is SemEval 2013 [26]. It is a twitter-based dataset having 1650 samples. It contains 3 labels, positive, negative and neutral. ‘0’ is for neutral, ‘-1’ is for negative and ‘1’ is for positive sentiment. As the dataset consists of three labels (‘1’ for positive, ‘0’ for neutral and ‘-1’ for negative), for our application of negative sentiment analysis, two labels have been used (converted from three labels by considering neutral tweets as negative). In the two-label approach, label ‘1’ represents non-negative or positive sentiment and label ‘-1’ represents negative sentiment. The experiments were performed in Python 3 software on an Intel i-5 processor with a 2.6 GHz clock. Our code took less than one minute to execute. The software implementation of our code is available online at [27].

Our method is the aspect-based method for negative sentiment analysis which is unsupervised, and in this approach, there is no need for training the model. In the preprocessing step, spelling correction was done, slangs and abbreviations are removed. Preprocessing includes text normalization, stemming, removal of stop-words, spelling correction. After the preprocessing step, the words are tagged along with their correct POS tags using the NLTK library. POS tags that do not give any sentiment information are identified and removed them from our text corpus. The remaining tags are the ones based on which the sentences can be scored. Since a single sentence contains both positive and negative sentiments, a prioritized scoring system has been made by identifying the aspect as well as the phrases containing the acceptable POS tags, in a priority order selected by us.  $k$ -fold cross-validation has also been performed for all our experiments, where  $k = 10$ . In this process, the

input is divided into  $k$  subsets. The model is trained on  $k-1$  subsets and evaluated on the remaining 1 subset. This experiment is repeated  $k$  times, each time choosing different  $k - 1$  subsets, and the final results are averaged. Apart from accuracy (in %), the precision, recall and F-score metrics are also computed. Table 5 shows the comparison between accuracy, precision, recall and F-score values of different classification models for the SemEval 2013 dataset. The highest accuracy and F-score values are got by using this method, as observed from the results summarized in Table 5.

For comparison, several baseline approaches are also implemented for sentiment analysis and detection that are supervised approaches. The least mean square (LMS) loss function is used for the training phase of all the supervised techniques. The grid search was used for determining support vector machine (SVM) parameter settings. Using n-gram, as a feature extraction method, delivered better results as compared to using the bag-of-word (BOW) features. Long short-term memory (LSTM) model also performed well. The Adam optimizer, with a batch size of 50 and 100 epochs, is the hyperparameter settings for LSTM. The GLOVE word embedding was used. The best result, however, was delivered by our unsupervised aspect-based method for detecting negative sentiment tweets.

## 4 Conclusions

Hatred in society has spread, in recent times, to social media platforms that are now required to be monitored continuously to detect negative online posts that may trigger mishappenings like riots. In this work, twitter posts have been classified based on their negative content, using an aspect-based approach based on the sentiment scoring of a phrase and keyword combination, detected separately. Various text preprocessing techniques are initially used such as normalization, stemming, removal of stop words, spelling correction and replacing abbreviations with their full forms. Phrases and keywords are selected from a prioritized list involving POS tags, from each sentence. The SentiWordNet lexicon provides us with the sentiment scores of phrases and keywords that are summed up for each sentence of the tweet and nonlinearly transformed into a decision score for detecting the sentiment of the tweet. A negative decision score indicates a negative tweet. Higher classification accuracies on the benchmark SemEval 2013 dataset prove the superiority of our approach as compared to the state of the art. Inclusion of multilingual slang to detect negative content of tweets is the future scope of our work.

## References

1. Giachanou A, Crestani F (2016) Like it or not: a survey of twitter sentiment analysis methods. *ACM Comput Surv (CSUR)* 49(2):1–41

2. Prabhat A, Khullar V (2017) Sentiment classification on big data using Naïve Bayes and logistic regression. In: 2017 international conference on computer communication and informatics (ICCCI). IEEE, pp 1–5
3. Devi DVN, Kumar CK, Prasad S (2016) A feature based approach for sentiment analysis by using support vector machine. In: 2016 IEEE 6th international conference on advanced computing (IACC). IEEE, pp 3–8
4. Vashishtha S, Susan S (2019) Sentiment cognition from words shortlisted by fuzzy entropy. *IEEE Trans Cogn Dev Syst*
5. Kumar KLS, Desai J, Majumdar J (2016) Opinion mining and sentiment analysis on online customer review. In: 2016 IEEE international conference on computational intelligence and computing research (ICCIC). IEEE, pp 1–4
6. Vashishtha S, Susan S (2019) Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Syst Appl* 138:112834
7. Vashishtha S, Susan S (2018) Fuzzy logic based dynamic plotting of mood swings from tweets. In: International conference on innovations in bio-inspired computing and applications. Springer, Cham, pp 129–139
8. Thet TT, Na J-C, Khoo CSG (2010) Aspect-based sentiment analysis of movie reviews on discussion boards. *J Inf Sci* 36(6):823–848
9. Lek HH, Poo DCC (2013) Aspect-based twitter sentiment classification. In: 2013 IEEE 25th international conference on tools with artificial intelligence. IEEE, pp 366–373
10. Brody S, Elhadad N (2010) An unsupervised aspect-sentiment model for online reviews. In: Human language technologies: the 2010 annual conference of the North American chapter of the association for computational linguistics. Association for Computational Linguistics, pp 804–812
11. Mukherjee A (2016) Extracting aspect specific sentiment expressions implying negative opinions. In: International conference on intelligent text processing and computational linguistics. Springer, Cham, pp 194–210
12. Yang B, Cardie C (2014) Joint modeling of opinion expression extraction and attribute classification. *Trans Assoc Comput Linguist* 2:505–516
13. Li H, Mukherjee A, Si J, Liu B (2015) Extracting verb expressions implying negative opinions. In: Twenty-ninth AAAI conference on artificial intelligence
14. Berend G (2011) Opinion expression mining by exploiting keyphrase extraction. In: Proceedings of 5th international joint conference on natural language processing, pp 1162–1170
15. Esuli A, Sebastiani F (2006) Sentiwordnet: a publicly available lexical resource for opinion mining. *LREC* 6:417–422
16. [https://www.webopedia.com/quick\\_ref/textmessageabbreviations.asp](https://www.webopedia.com/quick_ref/textmessageabbreviations.asp). Accessed 22 Mar 2020
17. Python port of SymSpell (2019) [online]. Available at: <https://github.com/mammothb/symspellpy>. Accessed 22 Mar 2020
18. Loper E, Bird S (2002) NLTK: the natural language toolkit. In: Proceedings of the ACL-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics, pp 63–70
19. Singh S, Rout JK, Jena SK (2016) Construct-based sentiment analysis model. In: Proceedings of the international conference on signal, networks, computing, and systems. Springer, New Delhi, pp 171–178
20. Miller GA (1998) WordNet: an electronic lexical database. MIT Press, Cambridge, MA
21. Marcus M, Kim G, Marcinkiewicz MA, MacIntyre R, Bies A, Ferguson M, Katz K, Schasberger B (1994) The Penn Treebank: annotating predicate argument structure. In: Proceedings of the workshop on human language technology. Association for Computational Linguistics, pp 114–119
22. Eisner JM (1996) Three new probabilistic models for dependency parsing: an exploration. In: Proceedings of the 16th conference on computational linguistics, vol 1. Association for Computational Linguistics, pp 340–345
23. Lamontagne L, Plaza E (eds) (2014) Case-based reasoning research and development. In: 22nd international conference, ICCBR 2014, Cork, Ireland, 29 Sept 2014–1 Oct 2014. Proceedings, vol 8765. Springer

24. Susan S, Zespai S, Sharma N, Malhotra S (2018) Single-keyword based document segregation using logistic regression regularized by bacterial foraging. In: 2018 4th international conference on computing communication and automation (ICCCA). IEEE, pp 1–4
25. Susan S, Keshari J (2019) Finding significant keywords for document databases by two-phase Maximum Entropy Partitioning. *Pattern Recogn Lett* 125:195–205
26. Wilson T, Kozareva Z, Nakov P, Rosenthal S, Stoyanov V, Ritter A (2013) SemEval-2013 Task 2: sentiment analysis in Twitter. In: Proceedings of the international workshop on semantic evaluation SemEval '13, Atlanta, Georgia, June 2013
27. <https://github.com/mkaii/bug-/tree/gh-pages>. Accessed 23 Mar 2020
28. Salazar DA, Vélez JI, Salazar JC (2012) Comparison between SVM and logistic regression: which one is better to discriminate? *Revista Colombiana de Estadística* 35(SPE2):223–237
29. Dias C, Jangid M (2020) Vulgarity classification in comments using SVM and LSTM. In: Smart systems and IoT: innovations in computing. Springer, Singapore, pp 543–553
30. Abbass Z, Ali Z, Ali M, Akbar B, Saleem A (2020) A framework to predict social crime through Twitter tweets by using machine learning. In: 2020 IEEE 14th international conference on semantic computing (ICSC). IEEE, pp 363–368
31. Chen Q, Ling Z-H, Zhu X (2018) Enhancing sentence embedding with generalized pooling. In: Proceedings of the 27th international conference on computational linguistics, pp 1815–1826

# A System for Anxiety Prediction and Treatment Using Indian Classical Music Therapy with the Application of Machine Learning



G. Kruthika, Padmaja Kuruba, and N. D. Dushyantha

**Abstract** Anxiety and stress are the problems of all times. Extensively, technology and art go hand in hand. Music therapy has been a sought-after technique for restoring one's mental health problems including anxiety, stress and blood pressure. The body symptoms such as palpitations, age, sex, sweat, EEG signals, blood pressure and heartbeat rate help in easy investigation of stress. Manual analysis of these parameters can be cumbersome and time-consuming. Hence, a more contemporary approach to aid the health specialists in the treatment of the patients is by using efficient music along with computing technology of machine learning algorithms. This work is a new product used to help health specialist treat patients with mental disorder using Indian classical music and machine learning. In this work, a dataset of mental disorder parameters and music datasets are mapped together as inputs. Music datasets include Indian classical ragas suggested by renowned health specialists. Machine learning algorithms, namely logistic regression, SVM, K-nearest neighbor (KNN), random forest (RF classifier) and decision tree classifiers, are used to train the models, and the accuracy performance of each model is compared. From the various testing datasets, the efficiency seen by our experiments that support vector machine (SVM) is best suited for our work which has an accuracy of 87.23%. To make it simple for health specialist and caretaker to handle this with ease, a web application using Flask, a python framework has been designed. The application enables the user to listen to the audio files and keep a track of one's stress status.

---

G. Kruthika · P. Kuruba (✉)

Department of Electronics and Communication Engineering, Global Academy of Technology, Bengaluru 560098, India

e-mail: [padmajamtech@gmail.com](mailto:padmajamtech@gmail.com)

G. Kruthika

e-mail: [kruthikag17@gmail.com](mailto:kruthikag17@gmail.com)

N. D. Dushyantha

Department of Electronics and Communication Engineering, KS School of Engineering and Management, Bengaluru, Karnataka 560109, India

e-mail: [dushyanth\\_crp@rediffmail.com](mailto:dushyanth_crp@rediffmail.com)

**Keywords** Machine learning · SVM · Indian classical music · Music therapy · Confusion matrix · Accuracy · Raga · Anxiety · Stress · Sensitivity · Specificity

## 1 Introduction

“Fear is inevitable, but it cannot paralyze you”. So is anxiety. Spielberger defined that ‘anxiety is an emotional state involving subjective feelings of tension, apprehension, nervousness and worry experienced by a person’ [1]. He defines anxiety as an unenthusiastic state which includes abstract sentiments of dread. Anxiety can come in various forms. But there is a lot of difference between being anxious and being emotional. A study from Concordia University shows that for millions of suffers from generalized anxiety disorder, anger is more than an emotion, it is analogous to a conduit that intensifies anxiety. Anxiety can either be psychosocial or be physiological. But it is assured that physiological stress is a result of psychosocial stress. Anxiety can affect our lives in one way or the other. The root cause of all mental ailments is **stress**. If mental health is not taken into consideration, stress backbones depression and anxiety. However, it is the psychological parameters that determine stress and anxiety. Stress kills the human brain unknowingly [2]. It can arouse due to work pressure, peer pressure, money crisis, family problems, etc. Anxiety disorders come in various forms. A few of them include generalized anxiety disorder, agoraphobia, panic disorder, social anxiety disorder (also known as SAD) and post-traumatic stress disorder (known as PTSD) [3]. Diagnosis of stress, anxiety and other mental health-related disorders is very challenging and a tedious process since it can vary from an individual to individual. One of the best parameters to measure anxiety in a person is hypertension. It is a common disease and leads to cardiovascular diseases which are a great concern of risk for human health. Few of the physiological parameters determining stress include heart rate variability (HRV), pupil dilation, hypertension and skin conductance. To obtain the neural response to anxiety/stress, non-invasive electrophysiological technique, known as electroencephalography (EEG), is used. It is one of the most efficient ways to measure the oscillations of brain waves [4].

India is a country with a very high prevalence of anxiety, depression and other mental health disorders. There are a lot of conventional approaches and methods to tame down anxiety. The most expected, antibiotic medicines are the conventional approaches to maintain the anxiety level in a person. The medicines have a lot of side effects. Another approach to reducing anxiety with zero side effects is music therapy. ‘Deep within us is the silent sound of our vibrations, which may be ignored but which is the musical core within us all’.—Yehudi Menuhin. A study found that people exposed to Indian classical music felt more relaxed, experiencing a reduction in stress oscillations in the brain and decreased heart rate which in turn reduces anxiety and cortisol levels. Music therapy practiced in ancient India called nada yoga had its effect on the mind, and its feelings estimate its conceivable clinical applications in the advanced time. Even though the foundation of music is based on

the cadenced (rhythmic) and recurrent developments of music, it is eventually the combination of the different sound examples that impact all our physiological and psychological actions [5]. Darbari kaanada has also helped a young girl in a coma to regain her consciousness back after listening to it for two weeks. Indian classical music is a great source for music therapy. Ragas resonate the human body. Raga in Sanskrit means 'color' or 'mood'. A raga is presented at a particular time in a day to bring out explicit mind-sets within the audience [6]. The use of music therapy is an evolving, acceptable practice for healing of psychological and psychosomatic ailments [7–11]. The rhythm called taala in Indian classical music also has an impact on emotions [12]. Nada anusandana is the practice of employing swaras and ragas to evoke a resonant response from targeted body chakras. Chakras are the energy centers in the human body. Music also has the power to create a state of ecstasy, distress and also pain. Carnatic classical music is replete with the effect of music in sometimes miraculous anecdotes [13]. Music therapy is not a cure to anxiety but an adjuvant measure to prevent anxiety.

Diagnosis of stress and its treatment through the sensor data is a tedious job and not efficient method [14]. Therefore, the development of computing systems (machine intelligence) to support the medicinal field in making accurate decisions is very important. Machine learning (ML) minimizes the task that requires human labor by providing tools and frameworks, enhancing accuracy. ML can detect and analyze depression. *Machine learning* is that field of computing that helps a machine to gain the capacity to learn without being programmed explicitly. ML is a data analytic technique in which the machines learn and produce accurate results. ML algorithms can be combined to identify patterns in data in the model. The advantage of such algorithms from the identified data provides a more robust and efficient diagnosis. ML is classified into supervised techniques, unsupervised algorithms and reinforcement learning methods. In this work, a supervised learning technique is used. A larger number of supervised ML algorithms are used for classification as well as regression. Most common algorithms are SVM, RF classifier, decision tree classifier, KNN and logistic regression [11]. The better way to have solutions for our problems specialized for us, effectively and instantly without human intervention, is by using machine learning techniques. Undoubtedly, now is the era of big data, and with such humungous set of anxiety parameters along with the corresponding set of ragas, ML proves to be the best approach to this ever-longing problem of mental health. For all the processing of the data, Jupyter notebook is used as an IDE. It is a web-based IDE. It is adaptable and enables the UI to support a wide scope of work processes on data analytics, computing and ML applications. In this project, the main parameters used for prediction of anxiety are the blood pressure, age and sex of the subject. Blood pressure, as well as heartbeat rate, gets affected by music waves. Hence, music therapy is proved as a significant part of hypertensive recovery. A simple front-end application is created to access the audio tracks (with reference to the raga therapy). Flask, a python framework is used for creating the front end. The subject can log in and choose any track he/she wants to listen to.



## 2 Related Work

The author [14] compares the performance of three ML methods: case-based reasoning, artificial neural networks and support vector machine. A supervised algorithm is used which with the parameters being finger temperature and heart rate variability (HRV) [15]. Two datasets, training datasets and testing datasets, are considered. ECG signals are also used to measure the datasets. In [4], electroencephalogram (EEG) signals determine the blink rate, facial gestures, dilation, skin conductance and heart rate which are the parameters to determine anxiety. There are certain mental arithmetic tasks (MAT) to which the subject is subjected. Certain parameters of the EEG signal determine anxiety. Ragas influence the emotions of an individual by altering and affecting the resonance and reverberation of the body. For healing with music, the cells of the body have to vibrate, and a resonant response is evoked. The three parameters to measure music are pitch, loudness and quality. Amygdala is the region of the brain that evokes emotions. Stress parameters are affected when subjected to pleasant music. Few ragas used in music therapy include darbari kaanada, miyan ki todi and hindola [6]. In [16], the author mainly emphasizes on a few body parameters to determine stress. All the ragas are stored onto a database. The therapy is then examined in the subjects' blood pressure. Clustering techniques are used to cluster patients based on age and BP. Later patterns are derived from clusters to analyze the behavior of blood pressure. Music therapy is moving from social to neuroscience perspective [17]. Appropriate scale, note, pitch and beat form the component ingredients of music as a therapeutic tool [13]. In [12], the author examines heart-beat rate, in which the spectral analysis of the RR tachogram is made. It was proven that tuning to Indian classical music (Hindustani or Carnatic) induces the impacts of excitement and sedation, hence making a person stress-free. The author [2, 18] uses EEG signals to examine stress. EEG signals are exact, precise and accurate and hence can be dependable for examination. A case of the different classification algorithm is used out of which SVM obtained the highest accuracy. In [19], the audio features are obtained from the audio signal and EEG datasets are used. Then some other existing features, such as absolute power is used for EEG signal analysis, are also analyzed. At last, these features are consolidated, and these two datasets are combined and RF classifier is used for classification. The proposed methodology has a front-end web application, where the subject registers himself with his login credentials. After the login, the subject can access the audio tracks required for the taming of anxiety or stress. Sensors are used to quantify the stress parameters. To confirm the effect of music therapy, IOT [20] is used. A predictive model is developed to diagnose anxiety and depression among older patients. Random forest showed the highest accuracy rate of 89%. When tested with another dataset, the RF model showed an accuracy rate of 91%. Patients were made to listen to music for 9 min. Decision tree classifier was implemented to anticipate the effect of music on stress-free conditions. An accuracy of 79% was obtained. The resulting tree enables it to find the prescient predictive variables that affect remedial music listening results [21]. In [22, 18, 23] EEG signals were used for stress detection. Many classifier algorithms were used for prediction.

They were SVM, RF, decision tree, ANN and hybrid methods. Hybrid methods and SVM obtained the highest accuracy. The related work uses two different methods for determining the applied models’ prediction performance. They are accuracy score and the area under the curve (AUC) score [22]. Blood pressure (hypertension) [24] is one of the main indicators of stress. The effect of raga malkauns (hindola) on the blood pressure parameters is summarized, and it shows that different tempos of the raga affect high blood pressure and low blood pressure readings [25].

In [13], the author describes Raga Arabhi to enhance the alertness of the brain by reducing negative emotions like jealousy, ego and toxic feelings. It is also helpful for overcoming depression or violent urges in children. Bhairavi is known as Sarva Swara Roga Nivarani Ragam. When a person is not able to identify what is wrong with oneself, however, feels unwell, or unhappy, and just knows that something is wrong, then Bhairavi is the raga that works on any problem. This raga enhances the ‘feel good’ factor. Darbari kaanada, an anti-stress raga gives stability to the heart chakra. This raga helps in bringing out inner strength. The inner strength that exists in rudimentary form in oneself is strengthened by singing the notes of this raga. Raga Mayamalavagowla releases tension and gives a feeling of peace which is very effective when the mind is agitated, and the person is unable to concentrate. It is also effective for people suffering from stammering due to anxiety, especially when they face crowds, or interviews [13]. The further sections of the paper include proposed work, methodology, scheme of work and implementation, results and discussion, conclusion and references.

### 3 Proposed Work

Music tames anxiety without causing any side effects on the subjects. Table 1 shows the seven swaras on classical music is associated with seven different emotions [13]. The first step of the project includes mapping the stress elements with the raga (chosen ragas are described in Sect. 4 of the paper). Exploratory data analysis and the data visuals are obtained, and the algorithms are applied to the model after a

**Table 1** Stress parameters

Attributes	Description
Age	Age in years
Gender	Male represent by 1 Female represented by 0
Blood pressure	Resting blood pressure Normal conditions: 120–140 mm Minimum: 90 Maximum: 200
Goal	Value 1: absence of anxiety/stress Value 2: presence of anxiety/stress Also includes a raga along with the values

sufficient amount of data cleaning. Five different algorithms are applied, logistic regression, KNN, random forest, SVM and decision tree algorithm. The accuracy of each model is obtained and compared with the highest accuracy obtained. A front-end web application is created. The application consists of the few Indian classical ragas which are used for music therapy. The user can register onto the page and log in. It consists of a prediction dashboard, where the user should be able to fill in the stress parameters and the dashboard would suggest the raga based on the patients' stress status. The patient can keep a track of the ragas.

### 4 Methodology

An ML program learns from *experience*, with reference to a set of *tasks* and measures the *performance* which improves through experience.

- *EXPERIMENT*—Collecting all the datasets (anxiety and music parameters). This collection will describe all the possibilities of whether the subject is under anxiety or not.
- *TASKS*—Clustering and segregating the collected data into different visualizations to understand the effect of whether the 'X', 'Y', 'Z' or 'XY', 'YZ', 'XZ' or 'XYZ' parameters help in predicting the detection of anxiety.
- *PERFORMANCE*—Different ML algorithms should be applied to the cleaned and the visualized data to measure the accuracy, confusion matrix, sensitivity and specificity. The algorithm which provided the highest accuracy is considered.

Figures 1 and 2 depict a general workflow of an ML model.

(1) Data block represents the dataset used for the prediction.

Dataset includes music elements and stress parameters in this methodology. The datasets of both music and stress elements are mapped into one CSV file

(2) *Data pre-processing*: A procedure which is used to convert the given raw data into an understandable format.

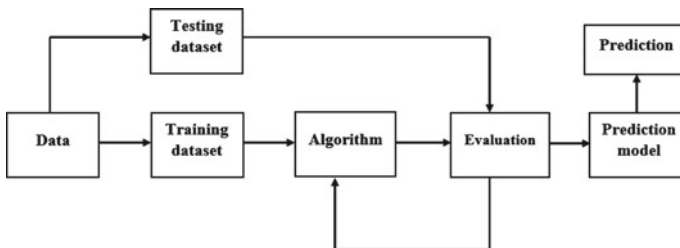
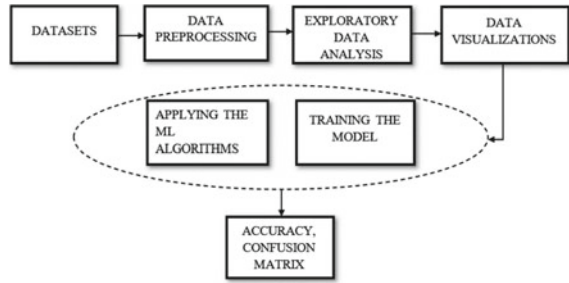


Fig. 1 Workflow of the proposed model

**Fig. 2** Procedure of an ML project



*DATA CLEANING*

- The data pre-processed can have many redundant and empty rows. To get rid of this, data cleaning is performed.

*DATA TRANSFORMATION*

- This process is done in order to normalize and scale the data in the dataset.

*DATA REDUCTION*

- Data analysis is a tedious process while working with a large number of datasets. To prevent this, this method is applied. It increases the storage capacity.
- The entire dataset is classified into training and testing dataset. Any algorithm applied to the dataset is on these training datasets only. Testing datasets as the name says are used only for testing purposes.
- Exploratory data analysis is done on the training datasets (Sect. 4 of the paper).

EDA is tied in with comprehending information close by, before getting messy with it. It involves analyzing the dependent variables. It helps in identification of categorical data and Numerical Data. It mainly portrays the flow of initial examination on data to identify patterns, detect anomalies and detect outliers, through statistics. A visual EDA is called ‘data visualization’.

- (5) A suitable algorithm is applied to the data and training accuracy is evaluated.
- (6) The same algorithm is applied to the testing dataset, and the testing accuracy is evaluated.
- (7) Accuracy, precision scores and sensitivity are calculated using the confusion matrix.

Software Requirement: Anaconda, Jupyter notebook IDE, Language: Python 3.7, Framework used for front end: Flask.

**Table 2** Music parameters

Raga	Benefit
1. Amruthavarshini	Dilutes the negative emotion
2. Arabhi	Eliminating depression and violent urges
3. Darbaari kaanada	Anti-stress raga
4. Hanumat Todi	Reduces fear
5. Mayamalavagoula	Releases tension
6. Neelambari	Reduces blood pressure

## 5 Scheme of Work and Implementation

### 5.1 Dataset Description

The parameters listed in Table 1 are the main parameters responsible for anxiety and are taken into consideration for prediction.

Table 2 depicts the ragas used for the front end and the dataset.

### 5.2 Feature Selection Flowchart

Tables 1 and 2 show the music and the stress parameters in the dataset. The source of stress datasets is the Kaggle repository, and the music datasets are manually added to the same file after discussion with a health specialist. After data cleaning, the dataset was found to have 298 rows and 15 columns. Figures 3 and 4 represent the flow of the proposed project.

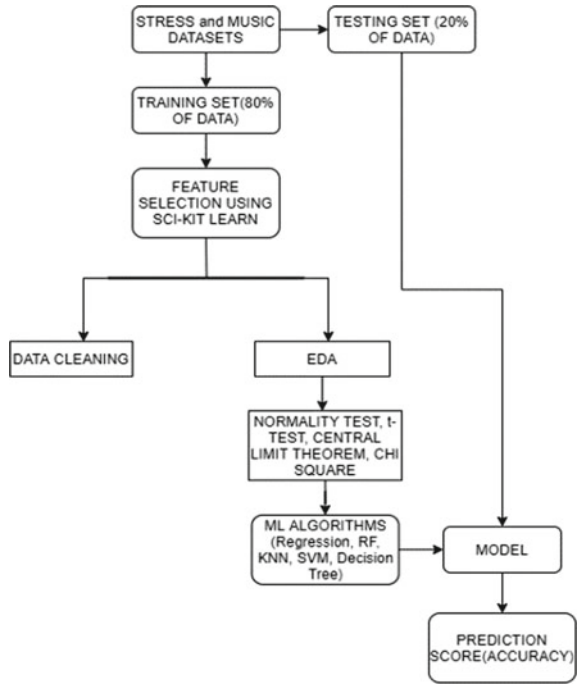
There are five different algorithms applied.

**LOGISTIC REGRESSION:** It a supervised classification technique. The yield variable ‘Y’ can take just a discrete arrangement of estimation of the information which is the input ‘X’. This algorithm is used to depict data information and to describe and clarify the connection between one binary variable which is dependent and ratio-level variables which are independent.

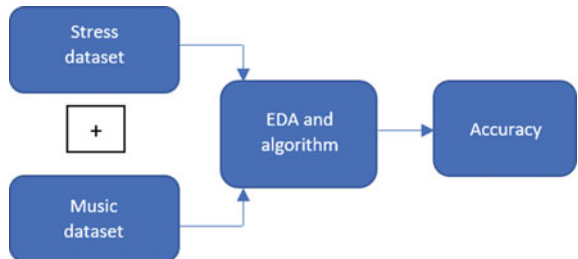
**KNN:** K-nearest neighbor (KNN) is a very simple and easy to use algorithm used in the machine learning domain for either regression or classification. This algorithm uses the given data and classifies new information points in comparison with the similarity measures of the given data.

**SUPPORT VECTOR MACHINE:** A support vector machine (SVM) belongs to a set of deep learning algorithms which perform supervised learning method for both classification and regression of the given datasets. In artificial intelligence and ML techniques, supervised learning techniques give both input and the expected output information or data, which are used mainly for classification.

**Fig. 3** Flowchart describing feature selection



**Fig. 4** Basic flow



*DECISION TREE CLASSIFIER:* Decision trees are also supervised machine learning techniques where the input data is indefinitely split based on certain parameters in the dataset.

*RANDOM FOREST:* Random forest which is also known as random decision forests is a learning method used for either classification or regression and several other processes that are performed by constructing a large number of decision making trees during the process of training and providing the output as the class. The output can be the prediction means all the individual trees or the mode of the output class.

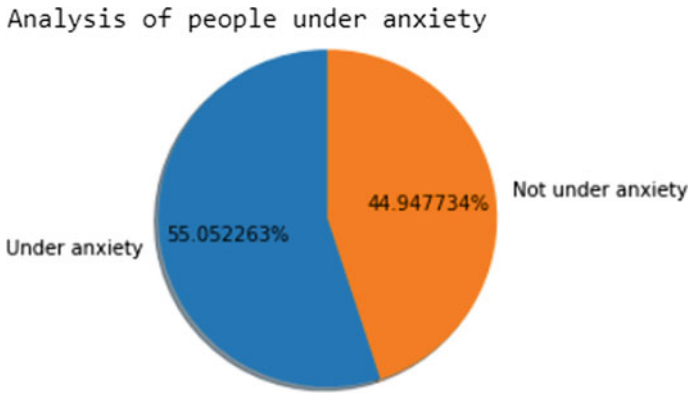


Fig. 5 Description of people under anxiety

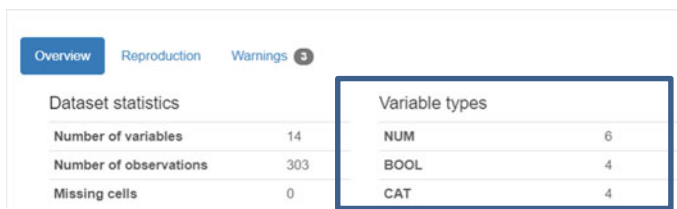


Fig. 6 Types of variables

### 5.3 Exploratory Data Analysis

Figure 5 shows that 55% of the people in the dataset are under anxiety and the rest of them not under anxiety. Figure 6 shows the three different types of variables present in the dataset. There are six numerical variables, four Boolean variables and four categorical variables (obtained through pandas profiling).

*Analysis of the correlation matrix:* A correlation matrix shows relationship (correlation) coefficients between the set of factors. This enables us to see which sets have the highest relation which is the correlation. From the analysis of the correlation matrix, it was found that:

*Positive correlation* (as D increases, E also increases)

>> As the age increases, blood pressure (Trestbps) also increases.

>> As the age and heart rate variability (HRV) of a person increases, there is a high chance of the subject suffering from anxiety (goal in the dataset)

*Negative correlation* (as D increases, E decreases or vice versa)

>> As the age increases, the heart rate decreases.

*Normalization test, central limit theorem, t-tests.*

*Null hypothesis for the project:* There is no mean difference between mean of the subjects subjected to stress or anxiety and subjects under stress-free status.

*Alternate hypothesis for the project:* There is a difference between the mean of the subjects subjected to stress or anxiety and subjects under stress-free status.

**Normality test** is used to identify if the collected blood pressure, age and other data are well demonstrated by a normal distribution and to find out how it is for an arbitrary factor to be normal in the given dataset. The **central limit theorem** is used, when independent arbitrary factors are added, and their sum which is normalized goes to a normal distribution under the circumstances in which the arbitrary random variables do not have a normal distribution. A **t-test** is used when the probability statistics of the test follows a normal distribution if the value of the scaled variables in the probability statistics of the test was known (described in Sect. 6).

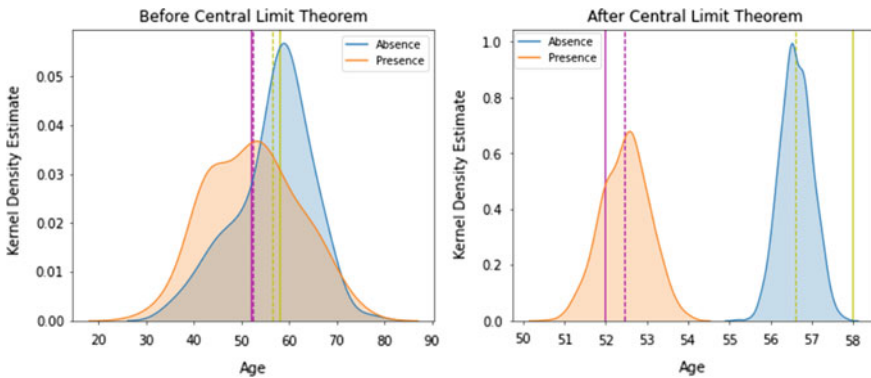
## 6 Results and Discussion

*Central limit theorem* is applied to eliminate the overlapping of the people affected with anxiety and people not under anxiety.

For example, considering age, in Fig. 7, the elimination of the overlapped cohorts can be easily visualized.

**AGE:** After the application of central limit theorem, mean and median of the people not under anxiety is 56 and 58 whereas the mean and median of people under anxiety is 52.4 and 52. Mean of people not under anxiety is less than the mean of the people under anxiety; hence, null hypothesis can be rejected. From Fig. 8,  $p < 0.05$ , and hence, aged people have higher chances of having an anxiety disorder.

**BLOOD PRESSURE:** Mean and median of the people not under anxiety are 133.2 and 133 whereas the mean and median of people under anxiety are 129.2 and 130. Mean of people not under anxiety is less than the mean of the people under anxiety; hence, null hypothesis can be rejected. From Fig. 9,  $p < 0.05$ , and hence, people with high blood pressure have higher chances of having an anxiety disorder. Similarly,



**Fig. 7** Central limit theorem applied to age



**Fig. 8** *t*-test for age

```
In [20]: #t test to analyse the mean difference
t, p = stats.ttest_ind(presenceofdisease,absenceofdisease)
print("t = " + str(t))
print("p = " + str(2*p))

t = -3.9303547512561936
p = 0.00021305762108129946
```

**Fig. 9** *t*-test for blood pressure

```
In [19]: #t test
t, p = stats.ttest_ind(presenceMeans,absenceMe
print("t = " + str(t))
print("p = " + str(2*p))

t = -99.0630634591247
p = 0.0
```

the same observation can be concluded for **heart rate** and **gender** as well. Hence, it can be concluded that blood pressure, age and gender are good predictive features.

Table 3 gives a comparison between the accuracies of the training and the testing dataset, specificity and sensitivity of the testing dataset.

**SVM was found to give the highest accuracy of 87.91%.**

**Front end**

Flask was used to create the front end. Flask is a web structure framed with Python. It is named a microframework because it need not have or require any particular libraries to be installed externally. It does not contain any data deliberation layer, form validation or other segments where already existing libraries provide a few similar functionalities (Figs. 10 and 11).

The proposed system consists of a web-page application named Kill Anxiety. The subject can register, login and fill in the stress details required. The application suggests the raga based on the input data. The user can also choose the other ragas in

**Table 3** Accuracy comparison of all the models

Model	Training accuracy (%)	Testing accuracy (%)	Sensitivity (testing set) (%)	Specificity (testing set) (%)
Logistic regression	86.79	86.81	87.91	86.53
K-nearest neighbors	86.79	86.81	88.00	88.00
<b>Support vector machine</b>	<b>93.39</b>	<b>87.91</b>	<b>88.00</b>	<b>89.79</b>
Decision tree classifier	98.59	78.02	74.00	84.09
Random forest classifier	98.98	82.41	84.00	84.00

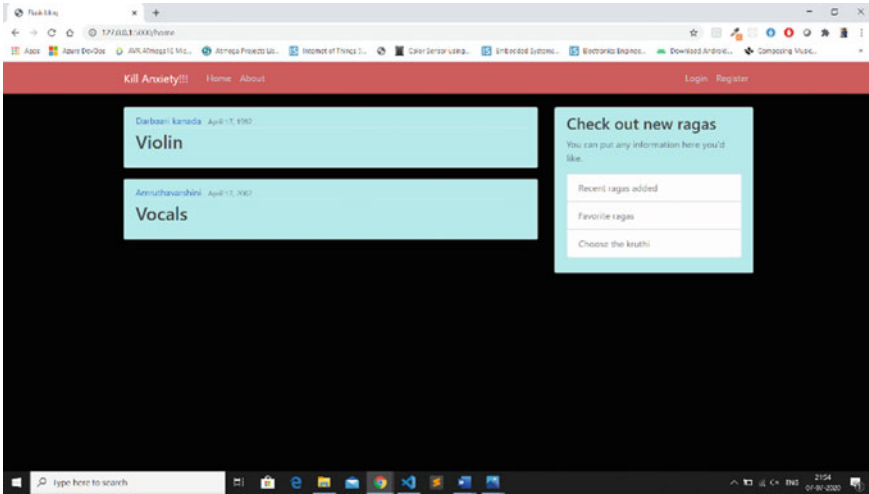


Fig. 10 Home page

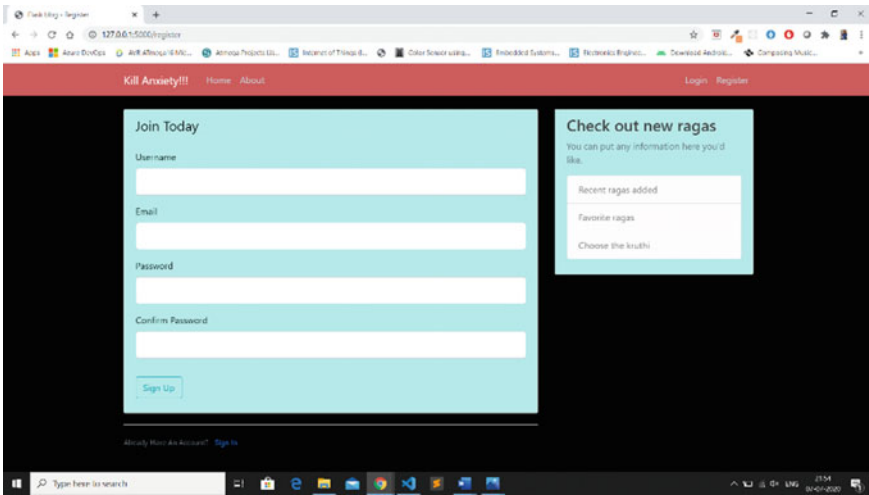


Fig. 11 Register page

the application. The audio track consists of both instrumental and non-instrumental audio. The subject can keep a record of one's favorite tracks.

The application consists of the audio files of the ragas described in Sect. 4.

## 7 Conclusion

This proposed methodology helps in detecting anxiety in humans. Anxiety can be tamed through raga therapy. The applied algorithms are logistic regression, KNN, support vector machine, random forest and decision tree classifier.

>> Out of the applied algorithms, support vector machine provided the highest accuracy of 93.39%. Accuracy can be increased by further data cleaning. Accuracy can be tested by the application of convolutional neural networks or other networks. Audio feature extraction can be done to understand the biological aspect of music on the human nervous system. Music therapy can have application either with medicine or as medicine. A person with limited musical knowledge may listen to any music of their choice for their wellness or mood management. This is an individual choice. It is important to distinguish the role of music for one's satisfaction and the role of a music therapist, who employs music for the wellness of the subject.

## References

1. Spielberger CD, Gonzalez-Reigosa F, Martinez-Urrutia A (1971) Development of the Spanish edition of the state-trait anxiety inventory. *Int J Psychol* 5:145–147
2. Sharma R, Chopra K (2020) EEG signal analysis and detection of stress using classification techniques. *J Inf Optim Sci* 41:230–232
3. American Psychiatric Association (1971) Diagnostic and statistical manual of mental disorders, vol 5. American Psychiatric Publishing, Washington, DC
4. Rauf Subhani A, Mumtaz W, Bin Mohamed Sad MN, Kamel N, Saeed Malik A (2017) Machine learning framework for detection of stress at multiple levels. *IEEE* 5:13545–13546
5. Pandian MD (2019) Sleep pattern analysis and improvement using artificial intelligence and music therapy. *J Artif Intell* 1:54–62
6. Bardekar AA, Gurjar AA (2016) Study of Indian classical ragas structure and its influence on human body for music therapy. In: 2nd international conference on applied and theoretical computing and communication technology (iCATccT). *IEEE*, pp 119–121
7. Ragilo A, Vico F (2017) Music and technology: the curative algorithm. *Front Psychol* 8:1–4
8. Sundar S, Parmar Parin N (2018) Music therapy clinical practice and research initiatives in India: bridge between the experiences of traditional music healing practices and its scientific validations. *IJAPR* 6:1–4
9. Jacob IJ (2019) Capsule network based biometric recognition system. *J Artif Intell* 1(02):83–94
10. Suma V (2019) Computer vision for human-machine interaction—review. *J Trends Comput Sci Smart Technol (TCSST)* 1(02):131–139
11. Gururaj V, Shriya VR, Ashwini K (2019) Stock market prediction using linear regression and support vector machines. *Int J Appl Eng Res* 14(8). ISSN 0973-4562
12. Chiu HW, Lin LS, Kuo MC, Hsu CY (2013) Using heart rate variability analysis to assess the effect of music therapy on anxiety reduction of patients. *IEEE*, pp 470–472
13. Shankar R (2019) The healing power of Indian Ragas
14. Barua S, Begum S, Uddin Ahmed M (2019) Supervised machine learning algorithms to diagnose stress for vehicle drivers based on physiological sensor signals. In: 12th international conference on wearable micro and nano technologies for personalized health, pp 1–3
15. Bashar A (2019) Survey on evolving deep learning neural network architectures. *J Artif Intell* 02:73–82

16. Prabhakar Pingle Y, Bhagwat A (2015) Music therapy and data mining using Indian ragas as a supplementary medicine. In: 2015 2nd international conference on computing for sustainable global development (INDIACom), pp 347–349
17. Hegde S (2017) Music therapy for mental disorder and mental health; the untapped potential of Indian classical music. *Bjpsych Int* 14:31–33
18. Sourina O, Liu Y, Nguyen MK (2011) Real-time EEG based emotion recognition for music therapy. *J Multimodal Interfaces* 27–35
19. Zhang F, Meng H (2016) Emotion extraction and recognition from music. In: 12th international conference on natural computation, fuzzy systems and knowledge discovery. IEEE, pp 1728–1733
20. Pingle Y (2016) IOT for music therapy. In: International conference on computing for sustainable global development (INDIACom). IEEE, pp 1453–1455
21. Ragilo A, Imbriani M, Imbriani C, Baiardi P (2019) Machine learning techniques to predict the effectiveness of music therapy: a randomized controlled trial. *Comput Methods Prog Biomed* 2–6
22. Pintelas EG, Kotsilieris T, Livieris IE, Pintelas P (2018) A review of machine learning prediction methods for anxiety disorders. In: 8th international conference on software development and technologies for enhancing accessibility and development, DSAI, pp 8–14
23. Sau A, Bhakta I (2017) Predicting anxiety and depression in elderly patients using machine learning algorithms. *Healthc Technol* 4:238–243
24. Yu JY, Huang DF, Li Y, Zhang YT (2009) Implementation of MP3 player for music therapy on hypertension. In: 31st annual international conference of the IEEE EMBS, Minneapolis, Minnesota, USA, pp 6444–6447
25. Kulkarni P, Jalnekar R (2017) Impact of raga malkauns on blood pressure measurement. In: Third international conference on computing, communication, control and automation (ICCUBE). IEEE, pp 1–3

# Design of a Water and Oxygen Generator from Atmospheric Pollutant Air Using Internet of Things



D. K. Niranjana and N. Rakesh

**Abstract** Reducing the contaminated water and obtaining fresh oxygen is difficult in these environmental conditions. Water is in the form of vapour and moisture mixture in the air. It is generally noted that 30% of water is almost wasted. The moisture in air is used in the proposed system. This system considers moisture air from atmosphere and converts it directly into drinking water. Condense process is used in the proposed system to convert moisture into water droplets. This device also generates oxygen ( $O_2$ ) from water through electrolysis process. This system consumes less power. The system sends the status of the system to the mobile application.

**Keywords** Water pollution · Oxygen generation

## 1 Introduction

Nowadays, the environment is affected in several ways, and the major reason for the effect is the pollution. The pollution impacted all the living lives, where it increases the health risk. The health risk has affected the whole world, and to reduce the life risk, the environmental conditions should be changed, but practically, it is not possible to change or control the environment. The reason why the environment is remaining unaltered is because the humans cannot stop their daily needs. The two major areas where the pollution adversely affected are air and water. Due to pollution and increased number of industries, the water is polluted and caused the health risk. So, the water scarcity has also been increased. The reason for water scarcity can be lack of rain or droughts and pollution. As there are a lot of industries in the world, where their wastages are dropped into freshwater, the natural water resources are contaminated. The water scarcity can be reduced by recycling it, but

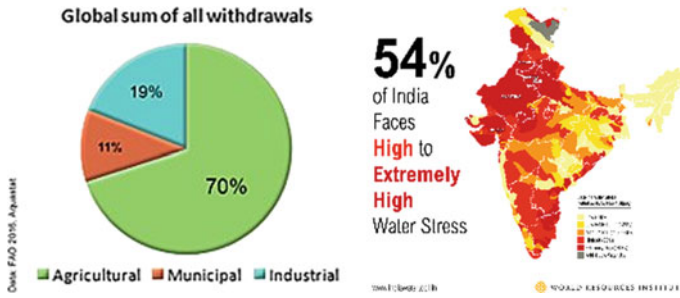
---

D. K. Niranjana (✉) · N. Rakesh  
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita  
Vishwa Vidyapeetham, Bengaluru, Bengaluru 35, India  
e-mail: [niranjandk6@gmail.com](mailto:niranjandk6@gmail.com)

N. Rakesh  
e-mail: [n\\_rakesh@blr.amrita.edu](mailto:n_rakesh@blr.amrita.edu)

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021  
J. Hemanth et al. (eds.), *Intelligent Data Communication Technologies and Internet of Things*, Lecture Notes on Data Engineering and Communications Technologies 57,  
[https://doi.org/10.1007/978-981-15-9509-7\\_31](https://doi.org/10.1007/978-981-15-9509-7_31)

361



**Fig. 1** Usage of water and raising water scarcity

the water quality will be reduced drastically. Even from the recycled water, the health risk will be initiated and many people will be facing the issue. On an average 1.2 billion, which is approximately one-fifth of the population in the world, are affected by water scarcity. One more reason for water scarcity is agriculture. In the process like irrigation, the water consumption is more, and this has started depletion of water from the sources. The whole world is left with only 3% of the earth's freshwater, which is not sufficient for the present population [1–3].

Figure 1 describes the reason for water scarcity, where it shows 70% of water is used for agriculture purpose and remaining 30% is used for municipal and industrial purposes.

In the metropolitan cities, the fresh air is also less, and proportionally, the health risk is increasing day by day. So, people have started changing their lifestyle such that they can reduce the pollution and also increase the life span. The drinking water in many areas is contaminated by chemical pollutants and the bio-waste. As the water contamination is increasing, the health-related infection and diseases are also increasing at an unprecedented rate. The most common diseases generated from the water contamination are skin infection, transferrable-fever, dengue, diarrhoea, etc. Nearly 21% of infection is generated from the contaminated water. The water pollution is increasing every day, so to reduce the health risk and increase the life span, freshwater and fresh oxygen are required. This work is mainly motivated to eradicate the health risk and water crisis across the globe with a minimal cost [4–7].

The air is also contaminated, and oxygen concentration level is also remaining less in the atmosphere. So, the system is designed, where the water is generated from the moisture of atmospheric air by filtering the fine dust from the air. The system is built to store the filtered water. The system also generates the oxygen from the generated water. The generated oxygen will be stored in the tank so that it can be used whenever required.

This paper is organized in the following order. In Sect. 2, the background work is explained. In Sect. 3, the proposed system is explained. In Sect. 4, hardware description is deployed. Section 5 describes the set-up and the result of practical experiments. Finally, in Sect. 6, conclusion and future work of the paper are explained.

## 2 Background Work

In [8], the researchers explain that the freshwater is generated for irrigation, through desalination process. The system is designed to use thin air for generating water for harvesting, whereas the device is very expensive. The system generates freshwater and store for agricultural use. The device where the water is stored is less and also the cost of generating water is too expensive where the cost of water generated by this system is of two dollars per  $m^3$  of land to be watered. For harvesting we need more water for irrigation so, the water generator device not much effective, and also cost of the device is more. In [9], this paper explains that they have integrated atmospheric water generator with distributed clean energy sources so that it can change the way to develop a new path to net zero. The energy consumed by the system is less, and also the device will reduce the carbon emission. The carbon emission is reduced to achieve net-zero water from the buildings and industries. In [10], authors explain that the desalination is the process used to clean a water. It is the global challenge for cleaning the contaminated water, especially in the industrial areas natural freshwater is contaminated. So, the communities rely on the desalination processes. The water is generated by atmospheric air, in the winter season only. In this season, the air moisture is more, so they can collect more water. The cost of water generation is same as the cost of bottled water. So, this device is more economical, but it works only in the winter season only. In [11], this paper explains that the atmospheric water generator units are more efficient only at coastal regions. The relative humidity is high in the coastal regions. They have obtained that specific temperature to condense water with the help of Peltier devices. The system works efficiently, where the humidity percentage is low. The cost of the device is more. In [12], this paper explains that it is a portable water generator system, where the water generator device is generated from the atmospheric air. They also used as a system to increase the temperature for evaporating technology. It is low productivity and also consumes more power and also consumes more time to fill water.

## 3 Proposed System

The system is designed to reduce the water crisis by generating water from the atmospheric air. The atmospheric air consists of moisture droplets in air; by these air droplets, we are generating freshwater. The freshwater is generated, where it depends on the air value data. If the air quality is worst or not the best quality, then the water will not be generated. If the air quality is fine or best in quality, it starts the generating of freshwater. Before starting the water generation, the air is filtered by the filters so that it filters all the dust particles and the particulate matters in the air. The quality of air is measured and analysed by our previous air purifier system. The air is purified based on the database stored in the system model so that the system purifies air and generates the water. Generation of water is done by the condense process. The cold

water is circulated continuously depending on air quality, for collecting the moisture in the air to generate fresh water. The stored water will be passed into the copper tube by that the fan will start by collecting the moistures from air and start generating water. The water is stored in a container, for future use of water. The water can be used for drinking purpose. In our proposed system, the oxygen is generated by the electrolysis process. The electrolysis process is done by the generated water. In our system, both hydrogen and oxygen are generated separately. The oxygen generated will help sick person as well as old age people. Both water and oxygen are generated by the atmospheric air.

Figure 2 shows the block diagram of the water generator system from atmospheric air. In our system, the air purification is done by the Raspberry Pi model, the data of air is sent to the water generator device so that it can decide and generate the water automatically, and the device also can be operated by the mobile application such that we will get notification and also we can ON and OFF the device automatically as well as manually. The device can be turned off instantly in case of emergency, and the device process will help all the old person so that we can control the water crisis in and around the world.

The generated water is filtered by the water purifying filters, and these filters are of three stages, where all the particles from the generated water will be filtered. The filters used in this device are of three stages. These are reverse osmosis membrane, sediment filter cartridge and mineraliser. This device is the less power consumption; in the winter and rainy seasons, this device works more accurately and generating more water than the summer time.

In Fig. 3, it explains the process of generating oxygen, where the device generates oxygen from the generated water by the process called electrolysis. Until the water storage tank, all the process is same. After that, the system consumes small amount of water from the generated water. The process is where the electrolysis process for generating the oxygen. The system will add little amount of sodium oxide. By electrolysis process, the solution will split the oxygen and hydrogen. The split oxygen and hydrogen are stored in the tank such that we can use whenever we want.

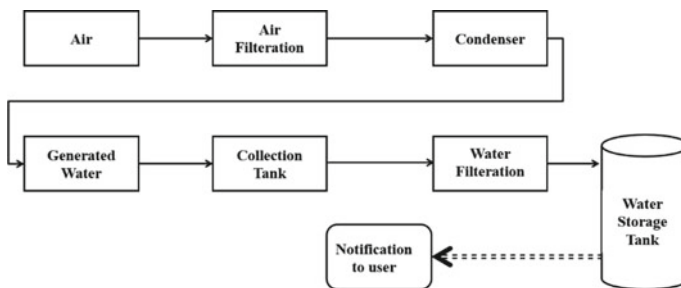
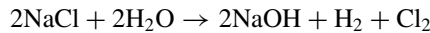


Fig. 2 Block diagram of water generator system



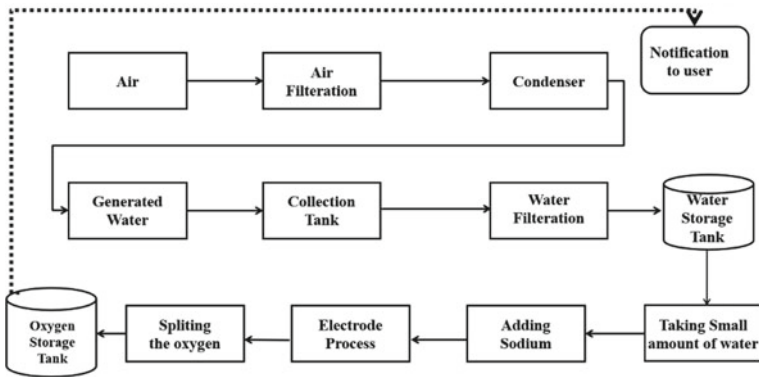
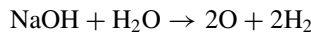


Fig. 3 Oxygen generator block diagram



In our device, the two stages of electrolysis process are done wherein first process is where sodium chloride is added with water and gives the sodium hydroxide solution is generated, and the second process of electrolysis is where sodium oxide and adding the generated water to it. It generates the solution where it splits the solution of the oxygen and hydrogen.

Figure 4 shows the flow chart of the proposed system.

In Fig. 4, the process of generating water and oxygen is started by collecting air quality data. The collected air data is stored in the device and to the server. Based on measured air data, the system sends data to the generator device by the decision taken the water generator device. If the air quality is less (less polluted air), then the system starts purifying. If the air quality is high (high polluted air), then the system will not start the water generator. When the generator starts, the condense process for generating the water is started and purifying the water through the water filters. Some amount of water from the water generator is taken and sent for generating oxygen through electrolysis process.

The system is most cost-effective than the existing models, and the existing models are designed only for generating water, but the proposed system is designed for purifying air and generating water as well as oxygen. The system is detachable and carryable with us, and the air purifier is the system operated by the battery (rechargeable). The operation of device is done by purifying the polluted air. Purifying the air is done by the five-stage air purifier and purification of air done based on the air quality (controlling the fan speed, three fans are used to purify the air).

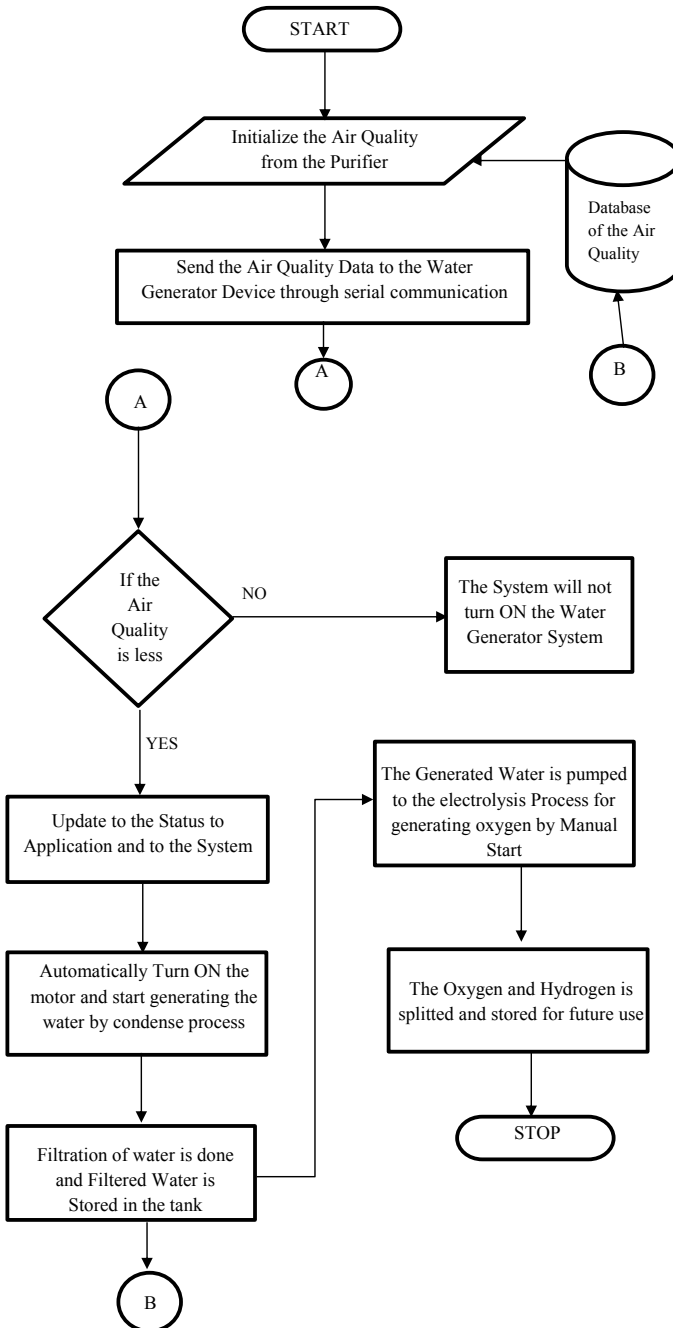


Fig. 4 Flow chart of the proposed system

### 4 Hardware Description

In our system, we have used Raspberry Pi 3B+ for doing all the process we want at less power so that we can configure the system perfectly where the Raspberry Pi has the built-in Wi-Fi so that we can directly upload the data values to the database. Raspberry Pi is of 1.4 GHz speed which is quad-core processor of 64-bit, it is of dual-band Wireless Local Area Network, in-built Bluetooth is of 4.2/BLE, it has Ethernet where it is fast, and it has Power-over-Ethernet support (with separate PoE HAT), so we have used this device for our work (Fig. 5).

A gas sensor is used to detect the polluted air and purify the air, and sensors are used for detecting MQ-135 and MQ-02 sensors. Figure 6 shows the MQ135 sensor and its pin details. MQ-135 detects or measures NH<sub>3</sub>, NO<sub>x</sub>, alcohol, benzene, smoke, CO<sub>2</sub>, etc. It has a wide detecting scope, stable, fast response and high sensitivity. It is of longer durability and consumes only 5 V. It can be used for both analog and digital depending on the work, and the digital output consists of TTL logic.

The proposed system has employed the most commonly used gas sensor MQ-02. The sensor works based on the variable resistance, as the chemiresistor as it

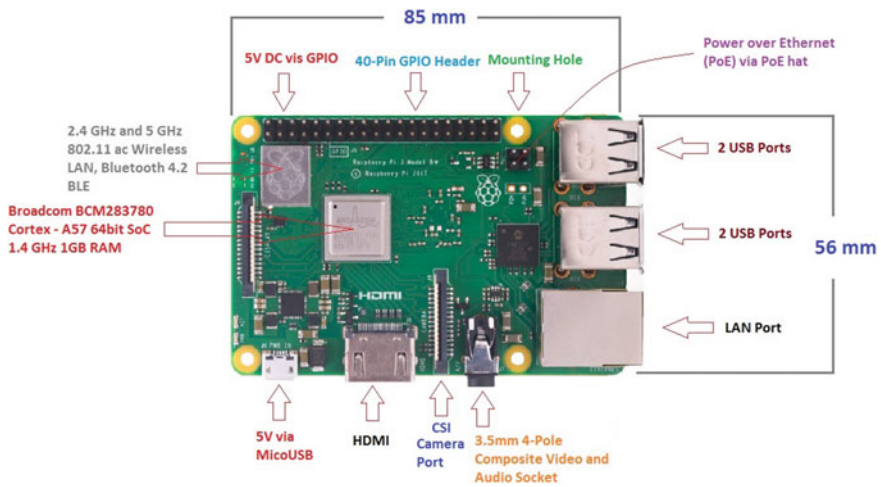
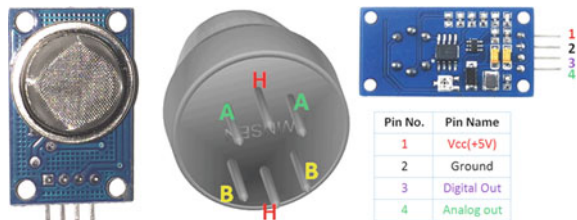
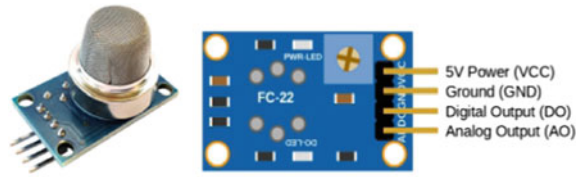


Fig. 5 Raspberry Pi 3B+

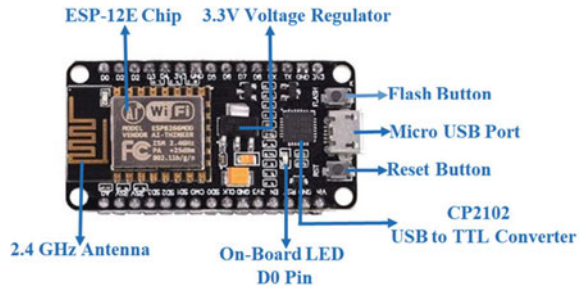
Fig. 6 MQ-135 sensor and pin details



**Fig. 7** MQ-02 sensor and pin details



**Fig. 8** Node MCU



detecting the gas, it varies resistance. When the gas is detected with the metal oxide semiconductor. The value of the analog signal output varies as the gas concentration varies. The sensor consists of two-layer steel mesh, where this mesh is called an ‘anti-explosion network’. This steel mesh is used to protect from explosion from the flammable gases. This sensor is used to detect the LPG, butane, propane, methane, alcohol, hydrogen and smoke concentration from 200 to 10,000 ppm (Fig. 7).

NodeMCU is a low-cost device where it is of open source for IoT platform. Firmware for the system is already initialized to run the ESP8266 Wi-Fi SoC from Espressif Systems. The module is of ESP-12 (Fig. 8).

The module is equipped with the ESP-12E module, and Tensilica Xtensa 32-bit LX106 RISC microprocessor is the system in the ESP8266 chip. The module operates from 80 to 160 MHz. The microprocessor supports real-time operating system. The frequency is an adjustable clock frequency.

It consists of 128 KB RAM and 4 MB of the flash memory to store programs and the data. It consists of high power processing with built-in Wi-Fi, Bluetooth and has a special operating feature of deep sleep to make it ideal for Internet of things projects. The module is powered using Micro-USB. It supports UART, SPI and I2C interface.

The proposed system has used ultrasonic sensor, where it measures the distance or level of the water. The sensor converts the sound into electric signal. The sensor transmits and receives the sound waves and later converts it into electric signal. This sensor transmits and receives the signal, which is faster than the speed of the sound that is audible to humans. This sensor has two components: transmitter and receiver. The distance between the sensor and the object is detected.

$$D = 0.5 * T * C$$

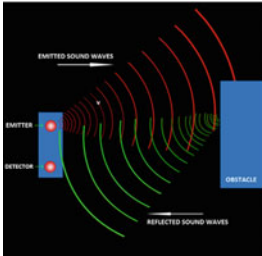


Fig. 9 HC-SR04 sensor

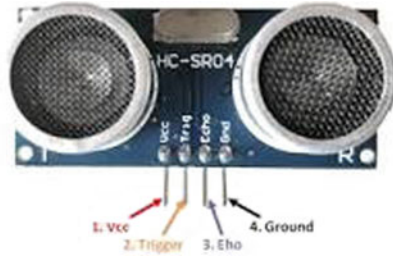


Fig. 10 ADC1015

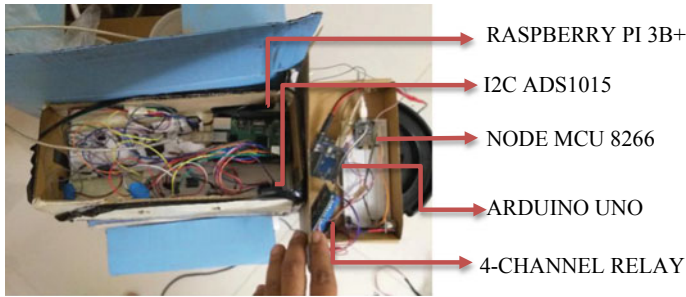


$D$ —distance  
 $T$ —time  
 $C$ —speed of the sound (Fig. 9).

The ADS1015 is used in our project for converting analog to digital data. Texas Instruments manufactured the ADS1015. It is the 12-bit precision at up to 3300 readings per second. The readings can be programmed. The board can be configured to accept four sensors of the type we will be using single-ended, or two differential channels, where two varying signals instead of a single signal and a ground. In ADS 1015, it consists of programmable gain amplifier. The programmable gain amplifier is built in and with up to  $\times 16$  gain. The gain amplifier will help to amplify smaller signals to the full range. This module will operate in a voltage range from 2 to 5 V. This voltage will be applied to the VDD pin (Fig. 10).

## 5 Results

The system consists of Raspberry Pi GPIO pins which are collected to the I2C, for converting the analog signals to the digital signal. The gas sensor is interfaced with GPIO of the Pi board. The python coding is done for simulation of the value and storing the record the value. The air quality values are recorded based on the 24-hour format, and every minute it updates the air quality data to the system, and also it



**Fig. 11** Circuit deployment

updates the voltage which is consumed by the system. Based on these values, an analog waveform is generated by the digital values in the offline mode web server. The water generator system starts working based on the air quality data. The designed system will generate the water if the air quality is less and the fresh for filtering the air. The system works based on the principle of the condense process. If the air quality is fresh, then the system will filter and start generating the atmospheric air into the fresh consumable water. The water can be used for drinking purpose. The generated water will be stored in a small container (Fig. 11).

The stored water in the small container will be pumped automatically and start filtering process of the water. The filtered water will be stored in a storage container for usage of drinking water. The generated water from the system will be pumped for generating oxygen. The oxygen is generated from the electrolysis process so that we can use for our future use, whenever we want.

The generated water is converted into oxygen using the electrolysis process by further adding the sodium hydroxide solution to the generated water. The process is done with 12 V DC supply so that the solution will split and generate the required amount of oxygen and hydrogen. This system will help to reduce the water scarcity in the metro cities. Proposed system has the capability to generate one litre of water in less than 30 min. The drinking water is stored in the tank for later use. The firebase data is used for generating the system application. Figure 12 depicts the air quality of the system, and if the air quality is less than the 50%, then the system will start generating water. As shown in Fig. 12, the air quality remains as 22%. The water is then generated and stored in a storage tank. It also shows whether the air quality is safe or not safe or it is required to check and operate the system manually (Fig. 13).

The air quality of the environment has been recorded through SQLite, and for every 1 min, from the collected data of air quality, the water generator system generates the water accordingly.

Figure 14 shows the MIT app inventor, where some basic information as well as the turning ON and OFF switch manually is designed. Figure 15 shows the screenshot of the designed application. Figure 15a shows the first page of the system. It displays “Water and Oxygen Generator”, start button are used to proceed for the next page. Figure 15b shows the second page, after proceeding the first page. The second page



Fig. 12 Air quality data and checking is it safe or not safe

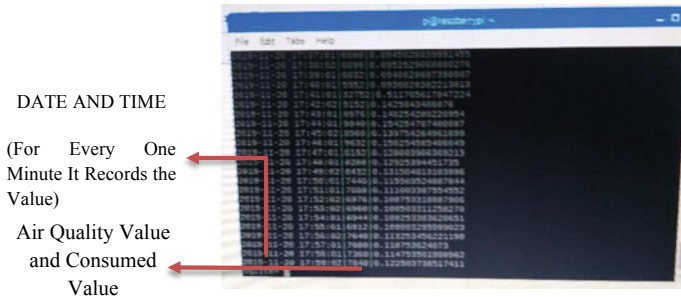


Fig. 13 Air quality database

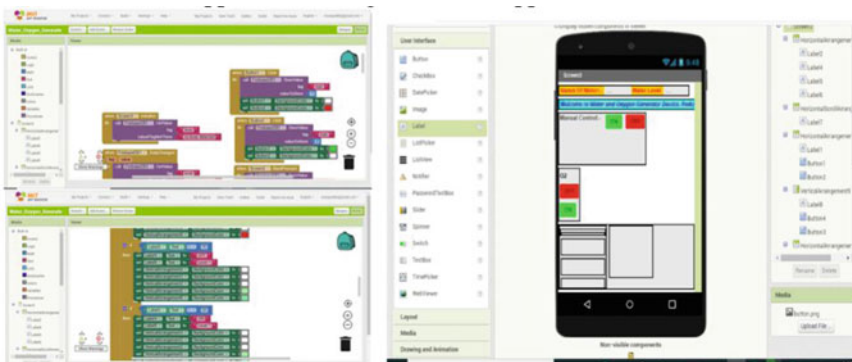
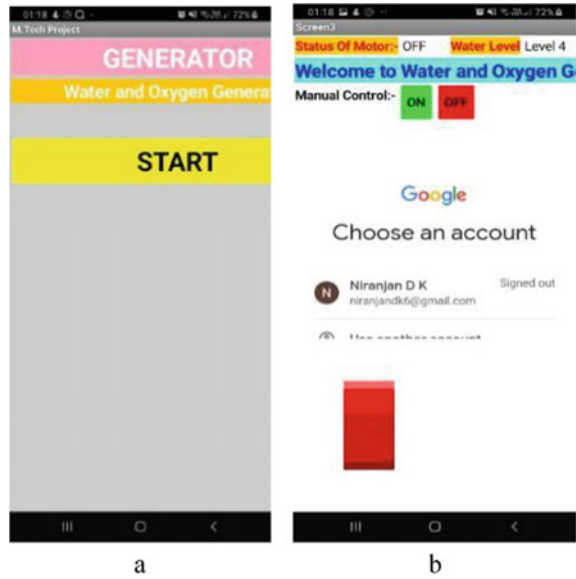


Fig. 14 MIT app inventor, designing the application

of the application shows the status of the motor, where it automatically performs ON or OFF operations. The water level from the generated water system can also be viewed by using a graphical representation. The red scroll bar shows the level of water in the system. It also has a control where the motor can be manually controlled



**Fig. 15** Screenshots of the designed application. **a** First page; **b** second page



by turning ON and OFF in the case of emergency, and if the system does not turn, the system turns OFF automatically [13, 14].

In the system process model, the process of water generation and filtration is shown in Fig. 16. The process of the filtration is of three stages: reverse osmosis membrane, sediment filter cartridge and mineraliser.

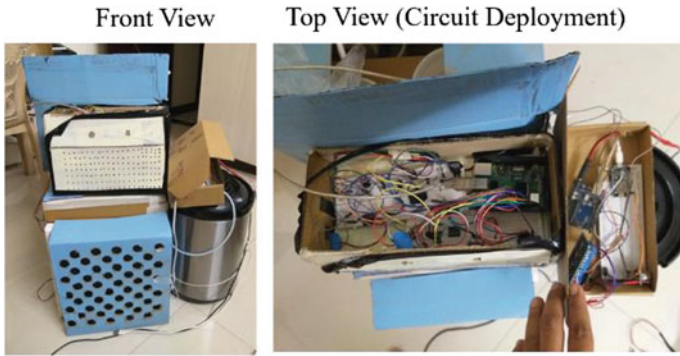
In reverse osmosis membrane, it removes the bacteria and viruses and reduces the dissolved salts. In sediment filter cartridge, it removes the fine impurities from the generated water. In mineraliser, it consists of minerals for adding to the generated water (Fig. 17).

The system generates the water based on the air pollutant data, and the generated water is from the atmospheric air. The collection of data is done by the air purifier part and sends the data value to the Arduino by the serial communication and updates the status of the system in the mobile application.



**Fig. 16** System process model





**Fig. 17** System model (front view and top view)

**Fig. 18** Electrolysis process of the oxygen generator (red colour specifies oxygen and black colour specifies hydrogen)



The oxygen is generated by the process of the two-stage electrolysis process, where the red colour specifies the oxygen generated and stored. The black colour specifies the hydrogen generated and stored, as shown in Fig. 18.

## 6 Conclusion and Future Scope

### 6.1 Conclusion

The real-time analysis of the air purification, water and oxygen generation is done on the basis of the at-present air quality and the dataset stored in the system. Depending on the variation of the quality of air, the system generates purified air, and the data recorded will help for monitoring the environmental conditions. Python language

is used for processing the system, and SQLite is employed for recording data from the air. Air is purified on the basis of the region, previous air data and the present air quality. The system generates freshwater and oxygen from the atmospheric air and generates oxygen by utilizing the electrolysis process. The proposed system will generate one litre of water in around 30 min and stored for drinking purpose. The system can be operated through mobile application so that it can be utilized globally. The details of the system will be updated in mobile application and can be controlled even by using the manual option. This device is low powered device and can be detachable for the air purifier and portable in nature. This device is mostly cost-effective compared to existing models, where the air filter and the copper tube (2-metre) are only the cost. With this system, the water crisis will be reduced, and also the health risk will be reduced by employing the fresh air and oxygen. The major disadvantage of this proposed model is the limited storage of water, and the time taken to generate oxygen is also increased (approx. 2-hour for small tank).

## 6.2 Future Scope

The system will be upgraded so that its storage capacity will be increased. The number of connections of connectors will be reduced so that the weight remains very less. Solar power will also be integrated with this system.

## References

1. Maag B, Zhou Z, Thiele L (2018) A survey on sensor calibration in air pollution monitoring deployments. *IEEE Internet Things J* 5
2. Cerro C (2018) Developing solutions for dealing with water and food scarcity. In: *Advances in science and engineering technology international conferences (ASET)*, Dubai
3. Reddy V (2017) Deep air: forecasting air pollution in Beijing, China
4. Curtis L, Rea W, Smith-Willis P, Fenyses Pan Y (2006) Adverse health effects of outdoor air pollutants. *Environ Int* 32:815–830
5. Hastie T, Tibshirani R, Friedman JH (2016) *The elements of statistical learning data mining, inference, and prediction*. Springer, New York, NY, pp 417–458
6. Rohde RA, Muller RA (2015) Air pollution in China: mapping of concentrations and sources. *PLoS ONE* 10(8). <https://doi.org/10.1371/journal.pone.0135749>
7. Ravi Subrahmanyam B, Gautam Singh A (2018) Air purification system for street level air pollution and roadside air pollution. In: *2018 international conference on computing, power and communication technologies (GUCON)*
8. Pontious K, Weidner B, Guerin N (2016) Design of an atmospheric water generator: harvesting water out of thin air. In: *IEEE systems and information engineering design conference 2016*
9. Adeoye OO, Shephard LE (2016) Atmospheric water generation: a path to net-zero. In: *IEEE green technologies conference, 2016*
10. Asiabanpour B, Summers M (2019) Atmospheric water generation and energy consumption: an empirical analysis. In: *IEEE systems and information engineering design conference 2019*
11. Bharath A, Bhargav K (2017) Design optimization of atmospheric water generator. *Int J Res Appl Sci Eng Technol (IJRASET)* 5(XII)

12. Kabeel AE, Abdulaziz M, El-Said EMS (2014) Solar-based atmospheric water generator utilisation of a fresh water recovery: a numerical study. *Int J Ambient Energy*
13. Vijnatha Raju P, Aravind RVRS, Sangeeth Kumar B (2013) Pollution monitoring system using wireless sensor network in Visakhapatnam. *Int J Eng Trends Technol (IJETT)* 4(4)
14. Rakesh N (2016) Performance analysis of anomaly detection of different IoT datasets using cloud micro services. In: *International conference on inventive computation technologies (ICICT 2016)*
15. Shitole S, Nair D, Pandey N, Suhagiya H (2018) Internet of things based indoor air quality improving system. In: *3rd international conference for convergence in technology (I2CT)*
16. Muthukumar S, Sherine Mary W (2018) IoT based air pollution monitoring and control system. In: *International conference on inventive research in computing applications (ICIRCA 2018)*

# Use of LSTM and ARIMAX Algorithms to Analyze Impact of Sentiment Analysis in Stock Market Prediction



Archit Sharma, Prakhar Tiwari, Akshat Gupta, and Pardeep Garg

**Abstract** The stock markets are considered to be the most sensitive and volatile financial institutions. Investment-related decisions are made by looking at historical data and observing the patterns, and sometimes these turn out to be profitable and sometimes not. Investment involves making predictions based on various factors and later combining all of these to conclude. The market as a whole is vulnerable to news and leaks which in turn decide the sentiment of buyers and hence directly affect the price of a given stock. Nowadays, machine learning techniques are being used to forecast the trend of a given stock. This paper presents a relative analysis of the prediction of the stock price using algorithms long short-term memory (LSTM) and auto-regressive integrated moving average exogenous (ARIMAX), without and with sentiment analysis. It has been observed from results that both algorithms attain a considerable improvement in the forecasting when sentiment analysis is applied.

**Keywords** Auto-regressive integrated moving average exogenous · Long short-term memory · Sentiment analysis · Stock market index

---

A. Sharma · A. Gupta · P. Garg (✉)  
Department of Electronics and Communication Engineering, Jaypee University of Information  
Technology, Wanknaghat, Solan, India  
e-mail: [prakhartiwari.1711@gmail.com](mailto:prakhartiwari.1711@gmail.com)

A. Sharma  
e-mail: [archit04.sharma@gmail.com](mailto:archit04.sharma@gmail.com)

A. Gupta  
e-mail: [garg.pardeep22@gmail.com](mailto:garg.pardeep22@gmail.com)

P. Tiwari  
Ingenious E-Brain Solutions, Gurugram, Haryana, India  
e-mail: [aksqpt7@gmail.com](mailto:aksqpt7@gmail.com)

## 1 Introduction

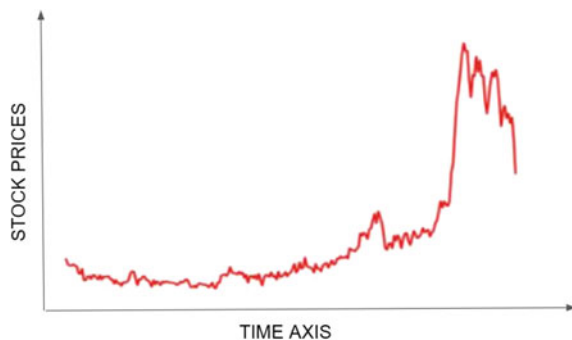
A financial market is a place where people trade financial securities and derivatives, like stocks, bonds, commodities and options features respectively. It is known that financial markets are very uncertain and volatile. Investors remain more and more concerned with their investment decisions as they want to get maximum returns on their investments. Financial markets give higher returns, but there is an equal amount of risk involved. Investors have to think very carefully regarding their decisions as their one mistake can lose their money. This is where forecasting comes as an aid, giving investors an edge in their decision-making process.

When the stock prices are talked about, lots of factors like the past performance of the company, news related to it, the sentiment of the people toward the company, etc., come into play. It is the aggregate of all these factors which in turn affect the prices of a given stock, which is like considering some variables and looking at their dominance over each other, and later some operations are performed at them to find a result. It is the diversity of these various parameters which make accurate forecasting of the stock prices difficult.

Many algorithms are being used nowadays to build a prediction model, and each of these has its pros and cons. Some of these are better than the others, and some have their limitations. The usage of artificial neural networks (ANN) has made the job much more efficient and accurate. Neural networks tend to provide better results when compared to traditional machine learning models, but in the case of the stock market, the dependency on sentiment analysis and other underlying factors like closing price, opening price, volume, etc., is high which is highlighted and demonstrated in this paper.

The graph in Fig. 1 presents a trend line for the stock prices of a company (Reliance Industries Limited) in a given trading year. Over a period of time, various spikes and dips can be observed in this figure. The motivation of this paper is to study the impact of the news in the market and the sentiments revolving around it, in context to these high and lows and their impact on the forecasting of the stock prices. The paper analyzes how the stock prices are influenced by the buyer's sentiment in a market place and how can then further use this to improve our forecasting.

**Fig. 1** Trend line for stock prices of RIL



In this paper, the possibilities and effectiveness of sentiment analysis in stock price prediction have been explored. The paper presents a parallel study of the two algorithms, ARIMAX and LSTM type of recurrent neural network (RNN). The performance of both the models is evaluated before and after the sentiment analysis was added to each model. Traditionally, the algorithms like auto-regressive (AR), auto-regressive integrated moving average (ARIMA), support vector machines (SVMs), etc., have been used from time to time in financial institutes. The algorithm ARIMAX can be described in simple words as ARIMA model with one or more exogenous variables, i.e., independent variables which are directly affecting the forecasting. LSTM which is an RNN can simply be defined as a feedback connection which learns from its recent past to avoid long-term dependency problems and predict the value of a given variable or feature.

The remainder of the paper is organized as follows. Literature survey is discussed in Sect. 2. The methodology is presented in Sect. 3 in which the approach used in the paper is discussed along with the detailed description of the dataset, LSTM algorithm, ARIMAX algorithm, sentiment analysis and performance metrics used in the paper. Results obtained are discussed in Sect. 4. And the conclusion of the paper is discussed in Sect. 5.

## 2 Literature Survey

Anita Yadav, C. K. Jha and Aditi Sharan in [1] used long short-term memory (LSTM) for time series prediction in Indian stock market. The data was collected from the Indian stock market, and then the model was evolved over it; after that, it was revamped by contrasting the stateless and stateful models and tuning for hidden layers. They used historical data for companies—TCS, Reliance, ICICI and Maruti. The model was tested using standard deviation values, spread in the box and whisker plot diagram which suggested that stateless LSTM was more stable as compared to stateful LSTM because of more stability.

Suhartono in [2] used the feedforward neural network (FFNN) model to forecast Indonesian inflation and then differentiated the result with ARIMA and ARIMAX models. The author showed that FFNN gives outstanding results in forecasting inflation in Indonesia. The best FFNN model in training data tends to yield overfitting on testing. Also, they suggested that the forecasting accuracy of ARIMAX was similar to FFNN with input based on ARIMAX.

Dev Shah, Wesley Campbell and Farhana Zulkernine in [3] selected two models, namely LSTM and deep neural network (DNN) for forecasting daily and weekly movements of Indian BSE Sensex for the company Tech Mahindra (NSE-TECHM). In the initial phase, both LSTM and DNN performed well in daily predictions after comparing the results trained for accuracy and error. Finally, LSTM outperformed DNN in weekly analysis predictions after the inclusion of more attributes. LSTM was more accurate in finding primary trends and making future forecasting on volatile stock datasets.

In [4], the authors have applied a sliding window approach to predict future values on a short-term basis using the daily closing price. For this approach, the authors had selected two different sectors, IT and Pharma. Companies were selected with the help of NIFTY-IT index and NIFTY-Pharma index. Training data consists of stock prices of Infosys over a period of time, and the test data had stock prices for Infosys, TCS and CIPLA. Three different deep learning models, namely RNN, LSTM and convolutional neural network (CNN), were used, and the performance of each model was quantified using percentage error. The authors proposed CNN methodology as the best due to sudden changes occurring in the stock market.

Kirti Pawar, Raj Srujan Jalem and Vivek Tiwari in [5] proposed implementation of RNN along with LSTM in comparison with traditional machine learning algorithms for stock market prediction. The historical data of Apple, Google and Tesla was considered from Yahoo Finance. The result shows that RNN-LSTM gives more accurate results than traditional machine learning algorithms, such as regression, SVM, random forest, FFNN and backpropagation.

In [6], Ghosh, Neufeld and Sahoo employed random forest and LSTM models as training methodologies to analyze forecasting of stocks of S&P 500 for intraday trading. They observed that LSTM outperforms random forests with higher accuracy. Social network media analytics is showcasing the promise for the prediction of financial markets [7]. Measurements of collective emotional states derived from large network scale are correlated to stock transaction data over a time period.

Forecasting on time series data for stock prices is now a trending research area. In previous years, many neural network models had been suggested to clarify the problem of economic data and to achieve precise forecasting results. The models integrated with ANN provide better results rather than using a single model. This has been concluded in [8] by analyzing the performance between BSE100 stock market index and NIFTY MIDCAP50 stock market index by different neural network models, and forecasting accuracy is analyzed and measured using performance metrics like mean absolute error (MAE), mean absolute percentage error (MAPE), percentage mean absolute deviation (PMAD), mean squared error (MSE) and root mean squared error (RMSE).

In [9], the authors have compared single-layer perceptron (SLP), multi-layer perceptron (MLP), radial basis function (RBF) and SVM and old statistical techniques including statistical methods and applications (SMA) and ARIMA to forecast the performance of market of Karachi Stock Exchange (KSE) at the end of the day using different attributes as input including oil rates, gold and silver rates, Foreign Exchange (FEX), news and social media feed and predicting the market as positive or negative. MLP performs better in comparison to other techniques.

Ping-Feng and Chih-Sheng in [10] proposed a combination approach that uses the distinctive strength of ARIMA which is one of the broadly used models in time series forecasting and SVM used in successfully solving the nonlinear regression in forecasting stock prices. The presented hybrid model greatly improves the prediction performance of a single ARIMA or single SVM model in predicting stock prices.

### 3 Methodology

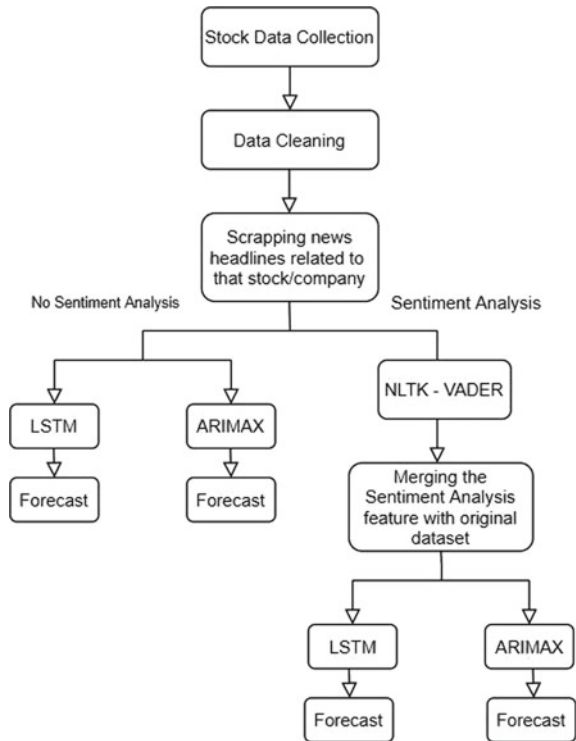
The block diagram of our approach is shown in Fig. 2.

The steps of the approach used are discussed as follows.

The approach starts from the data collection point where the stock data has been collected for the given company from the open sources like Yahoo Finance. The collected data has been then cleaned for null values and checked for abnormalities like negative prices, etc. which accounts for the cleaning part. The next data collection has been made from the various stock news Web sites to scrap the news headlines for a particular company using Python tools. And then the LSTM and ARIMAX algorithms are applied to forecast the prices for a given stock on both sentiment analyzed data and non-sentiment analyzed data.

The existing statistical forecasting models are not able to maintain a constant efficiency in forecasting the stock market prices largely because these do not account for ever-changing news and leaks in the market place which affects how the prices go up and down. This paper deals with the stock prices of an individual company and impact of news around it instead of an index because particular news might be beneficial to one company while may have a negative impact on the other's stock price and hence not truly reflecting the purpose of this paper. Moreover, the investors

Fig. 2 Block diagram of our approach





invest in a particular stock instead of an index, and hence, it is advantageous to analyze a single company.

The time series data is used to predict the closing price of the stock. The LSTM [1, 5] and ARIMAX [2] algorithms used in this paper first determine the stock price based on the features like open, high, low, volume using a dataset of a calendar fiscal year.

In the next step, the sentiment analysis is embedded in both models. The news for the particular company is Web scrape using *Scrapy* and *Beautifulsoup* from the openly available news platforms for a year. The text data then collected is passed through neuro-linguistic programming (NLP) techniques mainly Valence Aware Dictionary Sentiment Reasoning (VADER) for sentiment analysis of the news data collected. The detailed description of our approach is discussed in the next section.

### 3.1 Dataset

The dataset used in this paper is that of Reliance Industry Limited (RIL) scrapped from the Yahoo Finance Web site [11–13]. The reason for choosing this dataset is that Reliance Industries (RIL) play a major role in the Indian financial markets. The reason being it is a part of both the Nifty50 and Sensex, the flagship indices of Bombay Stock Exchange (BSE) and National Stock Exchange (NSE), respectively. So when there is a change in the trend of it, the rest constituents of both the indices have a tendency to follow it and hence driving the financial markets along with these showing either an uptrend or a downtrend. The effect sometimes can be observed on other industries too because if suddenly RIL has undergone a downtrend meaning the investors are not interested in buying and selling its shares which they already have, its price falls leading to declining in the points of indices. Once the points of an index start to decline, all the constituents of that particular index face the outcomes whether it is positive or negative. Therefore, it can be said that sentiments play an important role and especially in the case of RIL as it is considered to be a blue-chip stock and is considered to be a benchmark for many. So what happens with RIL can sometimes happen with stocks of other companies also even if these are from a different industry.

The data size used has been kept comparatively low to get a corresponding correct and more detailed per day news headlines data for the given company. The dataset considered is for a trading year which is around 253 trading days [14, 15]. The dataset originally had six features, namely *Open*, *High*, *Low*, *Close*, *Adj Close*, *Volume*, the *sentiment* feature has been added later. The training to testing data ratio has been kept as 75 to 25% in this paper.

The sample of the dataset used before the sentiment analysis was incorporated is shown in Table 1.

The text data (news) for sentiment analysis has been scrapped from the financial news Web site portal using *BeautifulSoup* and *Scrapy*. The sample of the financial

**Table 1** Sample of the dataset used

Date	Open	High	Low	Close	Adj. close	Volume
2019-03-07	0.2080	0.215	0.1682	0.2150	0.2150	308,690
2019-03-08	0.1801	0.200	0.1800	0.1860	0.1980	81,169
2019-03-11	0.1900	0.210	0.1700	0.1950	0.1950	112,475
2019-03-12	0.1955	0.210	0.1943	0.1943	0.1943	80,953
2019-03-13	0.1800	0.199	0.1800	0.1990	0.1990	28,003

Date = the date of the day on which the following trades were done  
 Open = the price of a single stock at the opening of the market  
 High = the highest price traded for the day  
 Low = the lowest price traded for the day  
 Close = the price of the single stock at the closing of the market  
 Adj Close = the adjusted closed price reflects the closing price of a given stock with reflection of other attributes like corporate actions etc.  
 Volume = it refers to the amount of shares traded on that particular day

```

↳ <bound method DataFrame.info of
0    RIL leased 4,000 acre land from Navi Mumbai SE...
1                                     none
2    Reliance Jio Long Term Recharge Packs: Prices,...
3    Choksey on why Reliance could appreciate 100% ...
4    Oil regulator hikes tariff of pipeline transpo...
..                                     ...
248  Reliance Industries worst hit, valuation drops...
249                                     none
250  Reliance Retail Ventures Limited (RRVL), a sub...
251  eliance Retail acquires Shri Kannan Department...
252  Reliance Industries Ltd - 500325 - Resolution ...
    
```

**Fig. 3** Sample of the financial news scrapped

news scrapped using these techniques in order of date-wise trading days is shown in Fig. 3.

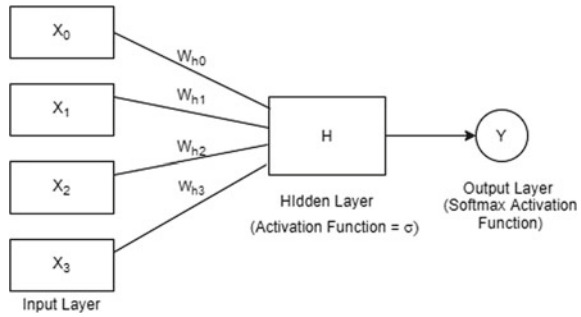
### 3.2 LSTM (Long Short-Term Memory)

The simple FFNN is the simplest form of an artificial neural network. The data in FFNN flows only in the forward direction from ‘X’ (input layer), though ‘H’ (hidden layer— $\sigma$  activation function) to the ‘Y’ (output layer—softmax activation function). The FFNN with a single neuron in the hidden layer is depicted in Fig. 4.

The following equations explain how the value of the hidden layer is obtained and the output layer using a softmax activation function [16].

$$H = \sigma(x_0 * w_{h0} + x_1 * w_{h1} + x_2 * w_{h2} + x_3 * w_{h3}) \tag{1}$$

**Fig. 4** Feedforward neural network with a single neuron in a hidden layer



$$Y = \text{SoftMax}(H * w_y) \tag{2}$$

where

- $\sigma$  = activation function,
- $(x_0, x_1, x_2, x_3)$  = input layer
- and  $(w_{h0}, w_{h1}, w_{h2}, w_{h3}, w_y)$  = weights of input and hidden layers.

Activation function Rectified Linear Unit (ReLU) performs the job of converting the sum of the weighted inputs from the node to the activation of that node. ReLU activation function returns 0 if it receives negative input, and for the positive value, it returns that value.

$$F(x) = \max(0, x) \tag{3}$$

where  $x$  is the positive value and  $F(x)$  is the activation function ReLU.

The problem with FFNN is that it is not efficient with the time series data. Time series predictions are made based on past trends and learning, and FFNN does not consider that. This problem is addressed with the RNN. RNN has loops in its network, thus allowing the information to remain in the network.

An RNN network is shown in Fig. 5, where  $X$  is the input,  $A$  is RNN and  $H$  is the output value.

The RNN is an FFNN, where the nodes/layers are arranged sequentially due to the nature of the dataset (time series). Unlike FFNN the value of a hidden layer is not only calculated from the input layer but also using the previous time step value and weights ( $w$ ). The expanded form of RNN is presented in Fig. 6.

RNN is capable of learning from the previous information, but its shortcoming is highlighted when a little more context than the recent past is required to solve the problem, i.e., the amount of data required to predict the next step is more. The LSTM was designed for solving these issues of long-term reliance.

LSTM [1, 5] is an RNN network with an added feature of ‘Memory’ cell. These are capable of learning long-term reliance. The LSTM cell has a different design, and update is maintained by the gates.

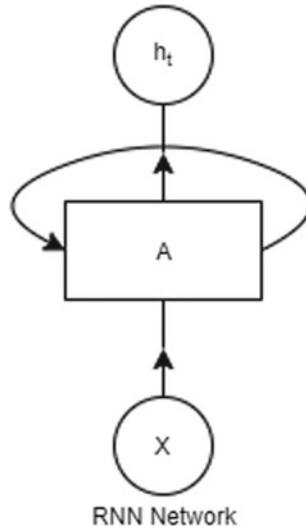


Fig. 5 An RNN network

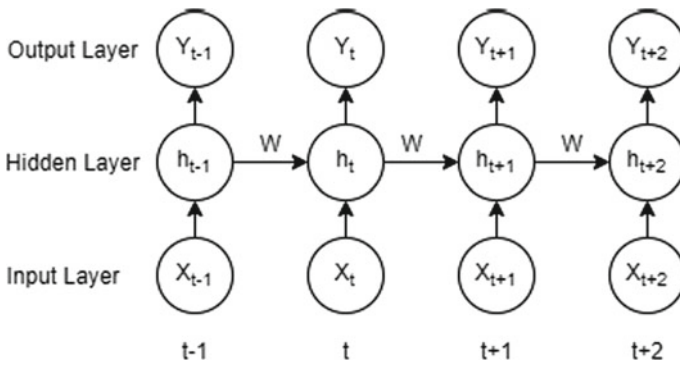


Fig. 6 Expanded form of RNN

The LSTM cell is shown in Fig. 7, where  $X_t$  is the input and  $H_t$  is the output value.

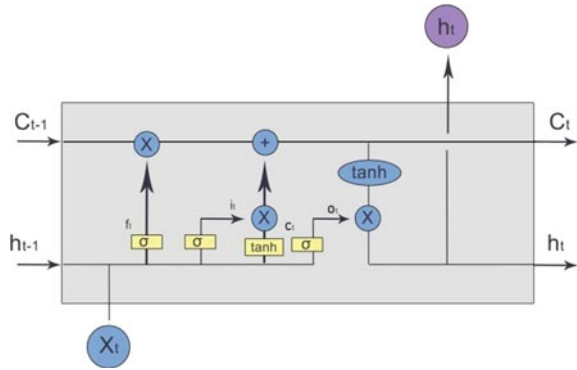
The following mathematical equations describe the implementation of LSTM gates.

$$C_t = C_t + (I_t * C'_t) \tag{4}$$

where  $C_t$  = (current memory state at time step 't' and gets passed to next time step) and  $I_t$ : input gate.

So to get  $H_t$ , Tan h can be applied:

Fig. 7 LSTM cell with gates



$$H_t = \text{Tanh}(C_t) \tag{5}$$

The LSTM used in this paper has four layers with the following parameters.  
 Layer 1 units = 60, activation function = relu, dropout = 0.2  
 Layer 2 units = 60, activation function = relu, dropout = 0.2  
 Layer 3 units = 80, activation function = relu, dropout = 0.2  
 Layer 4 units = 120, activation function = relu, dropout = 0.2.

### 3.3 ARIMAX (Autoregressive Integrated Moving Average Exogenous Variable)

ARIMA [2] is a method that combines terms of AR (autoregressive) and MA (moving average). ARIMA requires the time series data to be static; hence, it is differentiated at least once. The following equations depict ARIMA modeling.

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \dots + \beta_0 Y_0 + \varepsilon_t \tag{6}$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \dots + \beta_0 Y_0 + \varepsilon_{t-1} \tag{7}$$

where

- $Y_t$  = function of lags of  $Y_t$
- $\beta_t$  = coefficient of the lag t that model estimates
- $\varepsilon_t$  = the errors from the equation.

In this paper, the impact of the variables like sentiment analysis, closing price, etc., is addressed on the prediction. ARIMAX provides us with the options to add variables whose value is identified by the factors outside the model of the study. The following equation describes ARIMAX modeling [17].

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d (X_t - m_t) = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (8)$$

where,

$X_t$  = data time series

$\phi_i$  = autoregressive parameter

$\theta_i$  = moving average parameter

$\varepsilon_t$  = normally distributed error term (mean = 0)

$L$  = lag of operators.

### 3.4 Sentiment Analysis

The stock prices are highly volatile because of the involvement of a massive pool of small and big traders. These investors most of the times rely on news and leaks about a company to predict whether it will do good or not. The news about a company's involvement in a scandalous deal, new conflicting tariffs or government policies or top management malpractices on one hand can cause a stock price to fall while news like new big investments, good public image or strong cash flow attracts more people to invest and hence yielding higher stock prices.

The data for this sentiment analysis has been scrapped from a financial news Web site using *BeautifulSoup* and *Scrapy* in the form of news headlines of text data. In this paper, Natural Language Toolkit (NLTK) has been used to process this text data to get meaningful analytical figures. NLTK uses Valence Aware Dictionary and Sentiment Reasoner (VADER) for sentiment analysis which uses a dictionary and rule-based model sentiment analysis tool. VADER uses a dictionary of words; these words are mapped onto their categorical negative, positive, neutral, compound values. VADER also describes the score of a given text whether it will be positive or negative.

To avoid this issue of the score for a text where positive score describes good news while a negative score describes negative news and 0 is neutral, the news has been divided into the following three categories.

- 0: Negative scores (less than zero).
- 1: Zero score (equal to zero).
- 2: Positive scores (more than zero).

The sample of the dataset when sentiment analysis feature was incorporated to the dataset is depicted in Table 2.

**Table 2** Sample dataset with sentiment analysis features incorporated

Date	Open	High	Low	Close	Adj. close	Volume	Sentiment
2020-03-02	0.6550	0.7000	0.5450	0.6800	0.6800	135,539	0
2020-03-03	0.7000	0.7000	0.6000	0.6500	0.6500	58,679	1
2020-03-04	0.6999	0.6999	0.6350	0.6675	0.6675	5050	2
2020-03-05	0.6300	0.6300	0.5001	0.6050	0.6050	127,909	1
2020-03-06	0.6100	0.6100	0.4401	0.4900	0.4900	312,306	2

### 3.5 Evaluation Parameters

#### 3.5.1 Root Mean Square Error (RMSE)

This paper uses RMSE as the metric to assess the effectiveness of the given model in forecasting the price. It is the square root of the variance of the difference between the observed value and the predicted value. It is an absolute fit of a model. The value of RMSE should be as lower as possible for a model to be considered better. The mathematical formula [2] to find the value of RMSE is described in Eq. (9), where  $y$  is the value of the feature to be predicted.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_{\text{pred}, i} - y_i)^2}{n}} \quad (9)$$

The evaluation parameter RMSE is based on mean squared error (MSE) which is used by many ML algorithms for computing error, but the problem with this performance metric is that it is not scaled to the original error since errors are squared; hence, RMSE has been used in this paper as an evaluation parameter to match the original scale.

#### 3.5.2 Mean Absolute Deviation (MAD)

MAD is the metric to assess the performance of the given model in forecasting the price. MAD is an error statistic. It can be described as the mean difference between the actual and predicted value at each data point. The lower is the value of the MAD, lower the data is spread out and hence the prediction is better.

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (10)$$

MAD has been used alongside the RMSE as RMSE value is driven by ‘one big error.’ It means that one big error in the forecast would affect the RMSE value to a greater extent; MAD, on the other hand, shows the deviation between actual and

predicted prices per period in absolute terms resulting in a much lesser change in the evaluation parameter if a big difference occurs for a particular forecast.

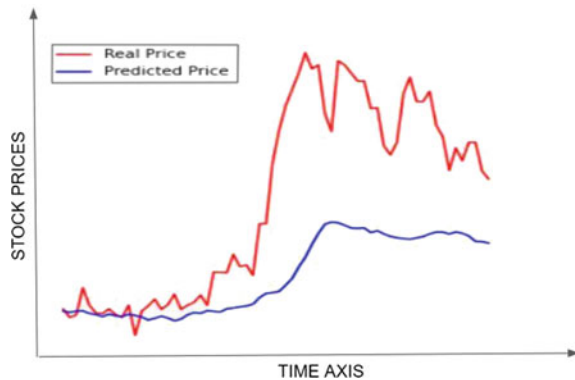
### 4 Result and Discussion

The initial LSTM model worked with given data yielded the avg. RMSE of about 0.48 and avg. MAD of about 0.33. The features like opening price, volume, high, low were included for the prediction model. The epoch (number of times the model was trained) showed a truly random behavior. The result obtained using the LSTM model regarding the behavior of actual price and predicted price is depicted in Fig. 8, where the X-axis represents the time while the Y-axis represents the stock prices.

The tabular comparison of the plotted graph for LSTM model which shows the difference in the actual price and the predicted price is depicted in Table 3.

After the feature of the sentiment analysis was added to the model/dataset, approximately 30% improvement has been observed in the prediction of the stock prices in case of RMSE. The average RMSE has been noted as around 0.34 for this case. Approximately 18% improvement has been observed in the prediction of the stock prices in case of MAD. The average MAD has been noted as around 0.27 for this case.

**Fig. 8** Result of LSTM model



**Table 3** Actual and predicted price values using LSTM

Actual	Predicted
0.17320	0.181097
0.15200	0.184764
0.15870	0.180734
0.22515	0.174252
0.18200	0.168685



The actual values of RMSE and MAD obtained without and with sentiment analysis using the LSTM model are presented in Table 4.

The result obtained using the LSTM model after incorporation of sentiment analysis regarding the behavior of actual price and predicted price is shown in Fig. 9, where the X-axis represents the time while the Y-axis represents the stock prices.

The tabular comparison of the plotted graph for LSTM model with sentiment analysis shows the difference in the actual price and the predicted price in Table 5.

When ARIMAX was used, it was also initially trained on features like opening price, volume, high, low. The initial ARIMAX model worked with given data yielded the avg. RMSE of about 25.16 and avg. MAD of about 19.11. The result obtained using the ARIMAX model regarding the behavior of the actual price and predicted price is depicted in Fig. 10.

The tabular representation shows a comparison between the actual price and the predicted price corresponding to the curves of the ARIMAX model in Table 6.

When the sentiment analysis was incorporated into the model, an average of 27% improvement in the RMSE value was observed. The average RMSE has been noted as around 18.3, while in parameter MAD, an improvement of around 15% approximately has been observed. The average MAD has been noted as around 16.2 for this case.

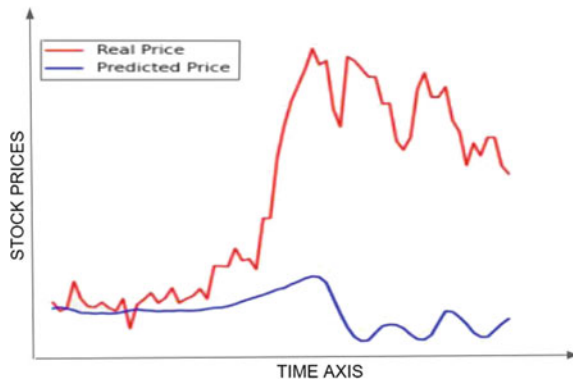
The actual values of RMSE and MAD obtained without and with sentiment analysis using the ARIMAX model are presented in Table 7.

The result obtained showing the comparative analysis of the actual and predictive price for the ARIMAX model with the sentiment analysis is presented in Fig. 11.

**Table 4** RMSE and MAD values using LSTM

Model	Value of RMSE	Value of MAD
LSTM	0.48	0.33
LSTM with sentiment analysis	0.34	0.27

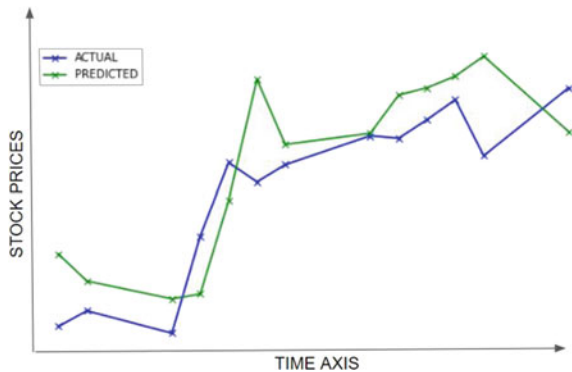
**Fig. 9** Result of LSTM model after incorporation of sentiment analysis



**Table 5** Actual and predicted price values using LSTM with sentiment analysis

Actual	Predicted
0.17320	0.171943
0.15200	0.181883
0.15870	0.196561
0.22515	0.207061
0.18200	0.198105

**Fig. 10** Result of ARIMAX model



**Table 6** Actual and predicted price values using ARIMAX

Actual	Predicted
0.146275	0.158542
0.147085	0.147160
0.145920	0.149244
0.150975	0.159283

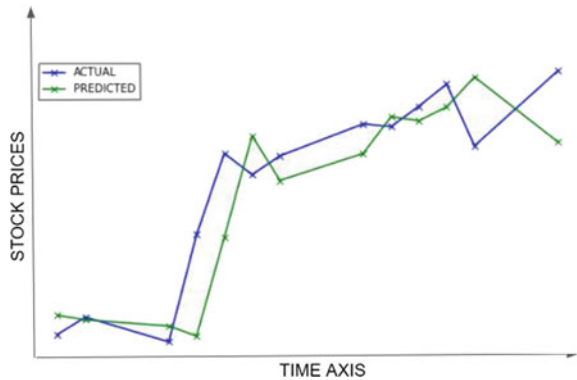
**Table 7** RMSE and MAD values using ARIMAX

Model	Value of RMSE	Value of MAD
ARIMAX	25.16	19.11
ARIMAX with Sentiment Analysis	18.3	16.2

The following table shows the comparison of the actual price and the predicted price when the sentiment analysis feature is added to the ARIMAX model in Table 8.

For the validation of the result, random news has been extracted from the source [13] to test whether the approach is accurate. The following news hit the market on 2019-11-30—‘RIL shares hit a record high, near 10 lakh crore milestone in the market cap’ which accounts to solidify the trust in the company shares hence is considered positive news of the sentiment (classified 3 according to our approach).

**Fig. 11** Result of ARIMAX model after incorporation of sentiment analysis



**Table 8** Actual and predicted price values using LSTM with sentiment analysis

Actual	Predicted
0.146275	0.157621
0.147085	0.129456
0.145920	0.149210
0.150975	0.156983

**Table 9** Actual and predicted price values using LSTM

Actual	Predicted
0.17320	0.169583

**Table 10** Actual and predicted price values using LSTM

Actual	Predicted
0.17320	0.171038

The following forecast has been computed by the LSTM algorithm with the corresponding actual price as mentioned in Table 9.

When the positive sentiment analysis (classified 3 according to our approach) has been accounted in the data, the LSTM algorithm computed following forecast which obtained better result as shown in Table 10.

## 5 Conclusion

In this paper, it is highlighted how the Indian stock market rather any stock market is highly dependent on the sentiments and news in the market. It does not matter how better an algorithm is, without considering the market news its accuracy will always be flawed. At the end of the day, the trading is done by the investors who are in

turn human beings only and are vulnerable to this news and leaks spreading around. It has been observed that both the algorithms considered in the paper showed an improvement in upward of 30% in RMSE when sentiment analysis was considered in the dataset and thereby emphasizing the impact of the sentiment analysis in the stock market to be a major factor. Hence, it is concluded that the application of the sentiment analysis feature reduced the RMSE by 30%, and in turn, the model accuracy was increased. It has been observed that both the algorithms considered in the paper achieved an improvement in range of 15–30% when sentiment analysis was considered in the dataset and thereby emphasizing the impact of sentiment analysis in the stock market to be a major factor. In the stock market trading, a small difference even of this scale can change the game for a trader and even a company. Hence, it is concluded that application of the sentiment analysis feature increased the accuracy of the fit approximately between 15 and 30% for a given dataset. In future, this work can be extended by working on improving the efficiency more by considering the indexes from around the world and how they tend to affect the price of blue-chip stocks like RIL and also to project a live prediction.

## References

1. Yadav A, Jha C, Sharan A (2020) Optimizing LSTM for time series prediction in Indian stock market. *Procedia Comput Sci* 167(2020):2091–2100
2. Suhartono S (2005) Neural network, ARIMA and ARIMAX models for forecasting Indonesian inflation. *J Widya Manag Acc* 5(3):311–322
3. Shah D, Campbell W, Zulkernine FH (2018) A comparative study of LSTM and DNN for stock market forecasting. In: 2018 IEEE international conference on big data (big data), Seattle, WA, USA, pp 4148–4155
4. Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP (2017) Stock price prediction using LSTM, RNN and CNN-sliding window model. In: 2017 international conference on advances in computing, communications and informatics (ICACCI), Udupi, pp 1643–1647
5. Pawar K, Jalem RS, Tiwari V (2019) Stock market price prediction using LSTM RNN. In: Proceedings of ICETEAS, pp 493–503
6. Ghosh P, Neufeld A, Sahoo JK (2020) Forecasting directional movements of stock prices for intraday trading using LSTM and random forests, April 2020
7. Zhang G, Xu L, Xue Y (2017) Model and forecast stock market behavior integrating investor sentiment analysis and transaction data. *Cluster Comput* 20:789–803
8. Kumar DA, Murugan S (2013) Performance analysis of Indian stock market index using neural network time series model. In: Proceedings of the 2013 international conference on pattern recognition, informatics and mobile engineering, pp 72–78
9. Mehak U, Adil SH, Raza K, Ali SSA (2016) Stock market prediction using machine learning techniques. In: 2016 3rd international conference on computer and information sciences (ICCOINS)
10. Nelson DMQ, Pereira ACM, Oliveira RAD (2017) Stock market's price movement prediction with LSTM neural networks. In: 2017 international joint conference on neural networks (IJCNN), Anchorage, AK, USA, pp 1419–1426
11. <https://in.finance.yahoo.com/quote/RELIANCE.NS?p=RELIANCE.NS&.tsrc=fin-srch>. Accessed 30 Apr 2020

12. <http://www.moneycontrol.com/company-article/relianceindustries/news/RI>. Accessed 30 Apr 2020
13. <https://trendlyne.com/latest-news/1127/RELIANCE/reliance-industries-ltd/>. Accessed 30 Apr 2020
14. Suma V (2019) Towards sustainable industrialization using big data and internet of things. *J ISMAC Acc* 1(01):24–37
15. Karthiban MK, Raj JS (2019) Big data analytics for developing secure internet of everything. *J ISMAC Acc* 1(02):129–136
16. [https://keras.io/api/layers/recurrent\\_layers/lstm/](https://keras.io/api/layers/recurrent_layers/lstm/)
17. <https://pyflux.readthedocs.io/en/latest/arimax.html>

# Design of Medical Image Cryptosystem Triggered by Fusional Chaotic Map



Manivannan Doraipandian and Sujarani Rajendran

**Abstract** Image is considered as important data in the medical field because usage of medical images for diagnosing the disease keeps on increasing in the modern digital medical field. These images need to be encrypted at the time of transmitting over an insecure network for maintaining integrity and confidentiality. This work aims to propose a medical image cryptosystem with less computational time based on the fusional model of the modified one-dimensional chaotic map. First, the proposed fusional chaotic model generates the chaotic sequence and then the generated chaotic series is used for encrypting the intensity of medical images. Simulations and security assessment are evaluated by applying statistical and differential attacks. Robustness against exhaustive and noise attack results indicates the strength of the developed cryptosystem. Comparison analysis illustrates that the proposed cipher has efficient and enhanced security than state of the art. Hence, the developed cryptosystem has absolutely opted of medical image applications.

**Keywords** Image encryption · Chaos · Tent map · Henon map · Sine map

## 1 Introduction

In this digital era, digital images are shared between entities which is at an all-time high. Thus, the combination of medical images and Internet gives a tremendous improvement in the medical field, not only for diagnosing the disease but also for encouraging the telemedicine and e-health applications [1]. Remote medical

---

M. Doraipandian

School of Computing, SASTRA Deemed University, Thanjavur, TamilNadu, India

e-mail: [dmv@cse.sastra.edu](mailto:dmv@cse.sastra.edu)

S. Rajendran (✉)

Department of Computer Science and Engineering, SASTRA Deemed University, Kumbakonam, Tamilnadu, India

e-mail: [rsujarani@src.sastra.edu](mailto:rsujarani@src.sastra.edu)

consulting requires the communication of medical image like MRI, CT, and X-ray images online. As consequence, integrity and confidentiality of these images have to be maintained at the time of transmission [2]. Hence, encryption plays an important role in transferring these medical images securely. Conventional cryptosystems like AES, DES, and IDEA have been proved as inappropriate for image encryption because of the image's properties like huge pixels, high redundancy, and correlation [3]. Numerous researchers identified that the chaotic cryptography is one of the recent technologies which is completely opted for image cryptosystem. Different directions and techniques of chaos-based image cryptosystem are discussed in the state of the arts [4–6]. Some of the chaos-based medical image cryptosystems is briefly given below.

Fu et al. [7] proposed an efficient medical image cryptosystem by utilizing Arnold cat map for bit permutation and the 1D logistic map used to generate the chaotic series used for substitution process with a minimum number of rounds. Ravichandran et al. [8] developed a new two fusional 1D chaotic maps for extensively confusing the pixels, and XOR operation is utilized for diffusing the image. Different attack analyses visualized the efficiency and security of the cryptosystem. Sathishkumar et al. [9] utilized 1D Bernoulli map and logistic map for seed key generation and to generate chaotic series for executing diffusion operation, and in this, cryptosystem images are read in the form of zigzag order and divided into  $8 \times 8$  blocks and then confusion and diffusion are applied on each block for encrypting the medical image. Cao et al. [10] presented a secure medical image protocol by combining three parts: decomposition of bit plans, generation of chaotic sequence, and permutation of pixels. Experiments and analysis results of that cryptosystem demonstrated the strength of that algorithm.

Most of the medical image cryptosystem discussed above are utilized 1D map for generating the chaotic sequence applied for confusion and diffusion stage. Even though 1D map has its own merits due to its lower-key size and limited life of randomness, it may cause threats to the security of the images [11]. As a consequence, higher-dimensional map became popular among researchers and different cryptosystem has been developed using a multi-dimensional map. Two-dimensional (2D) chaotic maps are found to possess good chaotic properties sufficient computational complexity [12]. Thus, 2D chaotic maps are employed in different image cryptosystem. Based on inspiring on the above state of the arts, a fusional chaotic map has been developed by combining three one-dimensional map like Henon, Tent, and Sine map. The proposed Tent–Henon–Sine (THS) map produces random series for a long period of time and has sufficient keyspace which will enhance the security of the image, a and dual confusion process has executed for greatly decrease in the correlation among pixels of the image. Security analysis has been executed by employing different attacks, and the results indicate the efficiency and security of the developed cryptosystem.

The paper framed as follows. Section 2 describes the basics of the developed cryptosystem. A detailed description of the proposed design is given in Sect. 3. Simulation results are shown in Sect. 4. Different security analysis and comparison are evaluated in Sect. 5. Section 6 concludes the proposed work.

## 2 Preliminary Concept

### 2.1 Tent Map

Tent map is one of the classical one-dimensional maps that is often introduced in the early stages of the chaotic map literature [13]. It generates a chaotic series in the form of a tent, so it is named as a tent map. Despite its minimalistic shape, it has many interesting features. It goes by the equation defined in Eq. (1). Another 1D map utilized for creating the THS map is sine map which is defined in Eq. (2).

$$x_{n+1} = \begin{cases} \mu x_n/2 & x_n < 0.5 \\ \mu(1 - x_n)/2 & 0.5 \leq x_n \end{cases} \quad (1)$$

$$x_{n+1} = \alpha \sin(\pi x_n)/4 \quad (2)$$

The positive integer value ( $\mu$ ) is a constant which must strictly lie within 0 and 4 and  $\alpha$  should be greater than 4. Thus, a chaotic series is achieved by executing the above equation iteratively.

### 2.2 Henon Map

In order to conquer the setbacks of the one-dimensional chaotic maps and to promote more randomness and clandestineness, chaotic series of two dimensions was introduced [14]. The two-dimensional chaotic map that is incorporated here is the Henon Map. The equation of the Henon map is in Eq. (3).

$$\begin{cases} x_{d+1} = y_d + 1 - ax_d^2 \\ y_{d+1} = bx_d \end{cases} \quad (3)$$

This map relies on the four parameters;  $x$ ,  $y$ ,  $a$ ,  $b$ . A classical Henon map has its constant value for the parameter  $a$  and  $b$  as 1.4 and 0.3, respectively.

### 2.3 Proposed THS Map

The THS map is obtained by fusing the above-mentioned maps, i.e., Tent, Henon, and Sine map. The equation obtained as a result of fusing these three maps is given in Eq. (4).



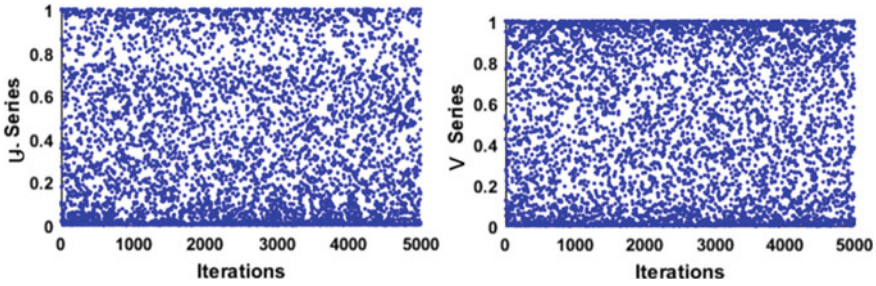


Fig. 1 Trajectories of chaotic series

$$\begin{cases} \begin{cases} u_{i+1} = 1 - ((\frac{\alpha}{4} \times \sin(3.14 \times (u_i)))) + v_i \\ v_{i+1} = \beta \times (\sin(u_i)) \end{cases} & u_i < 0.5 \\ \begin{cases} u_{i+1} = 1 - ((\frac{\alpha}{4} \times \sin(3.14 \times (1 - u_i)))) + v_i \\ v_{i+1} = \beta \times (\sin(1 - u_i)) \end{cases} & u_i \geq 0.5 \end{cases} \quad (4)$$

The chaotic series generated as a result of fusion possesses initial and system parameters should be in the range of  $0 \leq u, v \leq 1$  and  $\alpha, \beta \geq 4$ . The chaotic series generated by the THS map is highly random in contrast with the chaotic series generated by individual maps. The trajectories of the chaotic series are shown in Fig. 1 which defines the randomness of the THS map.

### 3 Image Cryptosystem Promoted by 2D THS Map

The proceeded cryptosystem in connection with THS map has a combination of confusion and diffusion process. Scrambling the pixels randomly depends on the chaotic series which is executed in confusion, and changing the pixel rate to a key image is performed in diffusion. The key image is filled chaotically with two chaotic series which increase the complexity of the algorithm. The block view of the suggested system is presented in Fig. 2.

From Fig. 1, it can be identified that the two chaotic series have good random properties. The proposed architecture that can be visualized in Fig. 2 is the grouping of confusion and diffusion process. Three ways of confusion are executed: first, the image is vertically divided into two equal parts, left and right parts of the image are shuffled separately and then combine both parts to shuffle the entire image. After, the shuffled image is taken as an input for the diffusion process. For both processes, the generated chaotic series are utilized. In confusion process, index of sorted series is utilized, and in diffusion process, integer format chaotic series is utilized to execute arithmetic operation between image and key series. The detailed explanation is given in the following steps.

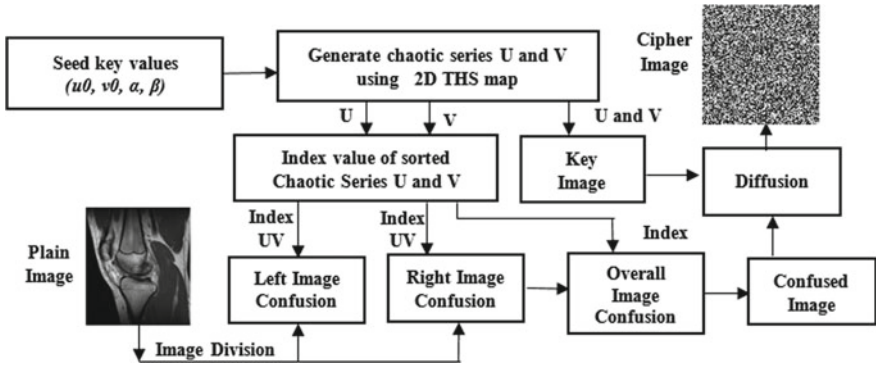


Fig. 2 Overall architecture of the proposed medical image cryptosystem

**Step 1:** Initialize the four keys  $\{x_0, y_0, \alpha, \beta\}$  with concerned values taken as a seed key of 2D THS map for generating chaotic series.

**Step 2:** Input mage (IM) and two key series  $U = \{u_1, u_2, u_3, \dots, u_{size}\}$  and  $V = \{v_1, v_2, v_3, \dots, v_{size}\}$  are originated depends on the seed values. Where  $size = (L \times B)/2$  and L and B represent the length and breadth of the image.

**Step 3:** Confuse IM with Procedure 1 to obtain the shuffled image SIM.

**Step 4:** Diffuse SIM with Procedure 2 to obtain the final encrypted image EIM.

### 3.1 Procedure 1: Image Cryptosystem—Confusion Part

**Input:** Input image IM and chaotic key series.

$$U = \{u_1, u_2, u_3, \dots, u_{size}\} \quad V = \{v_1, v_2, v_3, \dots, v_{size}\}$$

**Output:** Shuffled image SIM.

**Step 1:** Vertically divide the image into two parts as given in Eq. (5)

$$\begin{aligned} \text{left}_{x,y} &= \{IM_{1,1}, IM_{1,2} \dots, IM_{x,y}\} \\ x &= 1, 2, \dots, B/2 \quad \text{and} \quad y = 1, 2, \dots, L \end{aligned} \tag{5}$$

$$\begin{aligned} \text{Right}_{x,y} &= \{IM_{1,1}, IM_{1,2}, \dots, IM_{x,y}\} \\ x &= 1, 2, \dots, B/2 \quad \text{and} \quad y = 1, 2, \dots, L \end{aligned}$$

**Step 2:** Select B chaotic series  $U1 = u_1, u_2, \dots, u_B$  and  $V1 = v_1, v_2, \dots, v_{B/2}$

**Step 3:** In ascending order, chaotic series are sorted and store the new index value as given in Eq. (6).

$$[\text{sort } u1, \text{index } u1] = \text{sort}(U1) \quad (6)$$

$$[\text{sort } v1, \text{index } v1] = \text{sort}(V1)$$

**Step 4:** Shuffle the pixels of the left part using the following code.

```

For i = 1: L
  For j = 1: B/2
    Temp = left(i, j)
    left(i, j) = left(indexu1(i), indexv1(j))
    left(indexu1(i), indexv1(i)) = left(i, j)
  End
End

```

**Step 5:** Select  $B + 1$  to  $B + 256$  chaotic series  $U2 = u_{B+1}, u_{B+2}, \dots, u_{B+256}$  and  $V2 = v_{B+1}, v_{B+2}, \dots, v_{B+128}$

**Step 6:** In ascending order, sort the chaotic series and store the new index value as given in Eq. (7).

$$[\text{sort } u2, \text{index } u2] = \text{sort}(U2) \quad (7)$$

$$[\text{sort } v2, \text{index } v2] = \text{sort}(V2)$$

**Step 7:** Shuffle the pixels of the right part using the following code.

```

For i = 1: L
  For j = (B/2 + 1):B
    Temp = right(i, j)
    right(i, j) = right(indexu2(i), indexv2(j))
    right(indexu2(i), indexv2(i)) = left(i, j)
  End
End

```

**Step 8:** Combine the two parts and shuffle the entire image by using the following code.

$$\text{SIM} = \text{Merge}(\text{left}, \text{right})$$

**Step 9:** Select chaotic series from  $k = B + 256$  to  $k + 256$  and sort and store the index as in Eq. (8)

$$U3 = u_{k+1}, u_{k+2}, \dots, u_{k+256} \text{ and } V3 = v_{k+1}, v_{B+2}, \dots, v_{B+256}$$

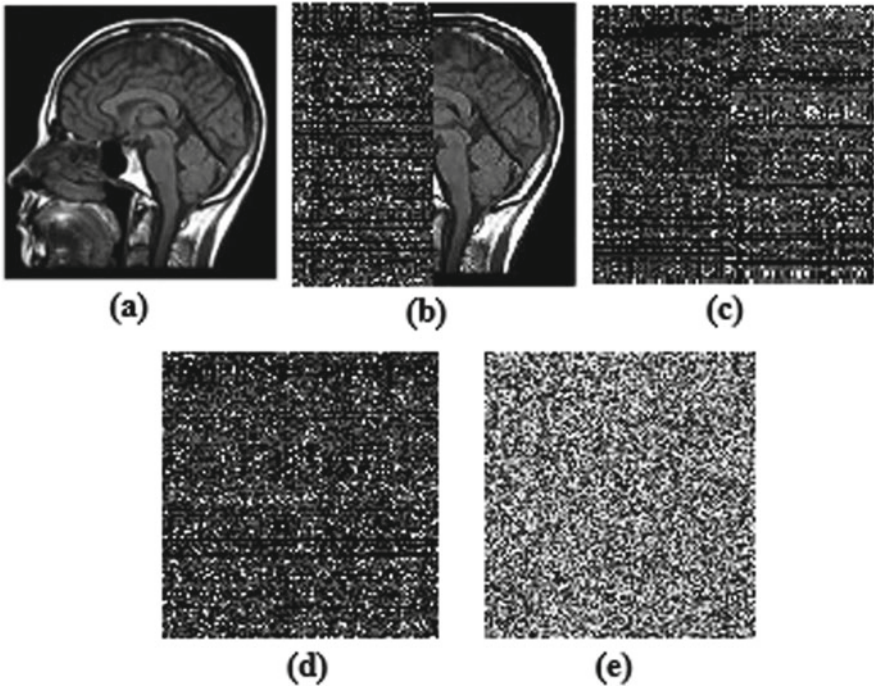
$$[\text{sort } u3, \text{index } u3] = \text{sort}(U3) \tag{8}$$

$$[\text{sort } v3, \text{index } v3] = \text{sort}(V3)$$

**Step 10:** Shuffle the entire image SIM by using the following code.

```
For  $i = 1:L$   
  For  $j = 1:B$   
     $Temp = SIM(i, j)$   
     $SIM(i, j) = SIM(\text{index}u2(i), \text{index}v2(j))$   
     $SIM(\text{index}u2(i), \text{index}v2(i)) = SIM(i, j)$   
  End  
End
```

Finally, confused image SIM is obtained. The output of each shuffling process is shown in Fig. 3.



**Fig. 3** Experimental result: **a** original image **b** left; **c** right; **d** whole image and **e** diffused image

### 3.2 Procedure 2: Image Cryptosystem—Diffusion Process

**Input:** Confused image SIM, Key streams U and V.

**Output:** Final encrypted image EIM.

**Step 1:** Transform the decimal for chaotic series and store it as a two-dimensional array that is taken as a key image with the size equal to the image. The creation of key1 and key2 is shown in Eqs. (9) and (10).

$$\begin{aligned} \text{Key1}_{i,j} &= ((U \times 10^{14}) \bmod 256) \text{ if } ((j \bmod 2) == 0) \\ \text{Key1}_{i,j} &= ((V \times 10^{14}) \bmod 256) \text{ if } ((j \bmod 2) \neq 0) \end{aligned} \quad (9)$$

$$\begin{aligned} \text{Key2}_{i,j} &= ((V \times 10^{14}) \bmod 256) \text{ if } ((j \bmod 2) == 0) \\ \text{Key2}_{i,j} &= ((U \times 10^{14}) \bmod 256) \text{ if } ((j \bmod 2) \neq 0) \end{aligned} \quad (10)$$

**Step 2:** Pixels of the confused image are diffused by executing XOR between key images and confused image as expressed in Eq. (11).

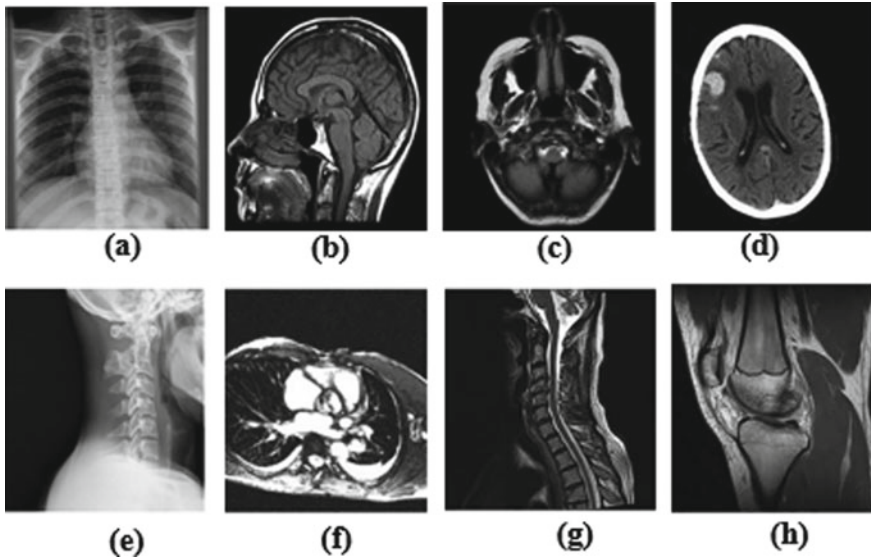
$$\text{EIM} = \text{bitxor}(\text{key2}(\text{bitxor}(\text{Key1}, \text{SIM}))) \quad (11)$$

Finally, encrypted image EIM is obtained and ready for secure communication.

Decryption will be the backtrack of the encryption process by using the same keys which are shared between the end users.

## 4 Experiment Result and Analysis

The developed cryptosystem is executed in the platform of MATLAB 2016a. A different set of images has been taken for evaluation purpose from S. Barre: Medical Imaging: Samples shown in Fig. 4 with size  $256 \times 256$ . For considering the time efficiency and compatibility .jpg format images ARE used instead of DICOM. The chosen key for the experimental purpose is  $u_0 = 0.23456543467876$ ,  $v_0 = 0.54678392028475$ ,  $\alpha = 12$  and  $\beta = 13$ . Any value can be taken for  $\alpha$  and  $\beta$  which should be greater than 4. First, the chaotic series is generated, and after that, the confusion process is executed to get the confused image, and finally, the diffusion process is executed to obtain the final encrypted image. The result of the developed cryptosystem is shown in Fig. 3.



**Fig. 4** Sample medical images taken for experimental purpose

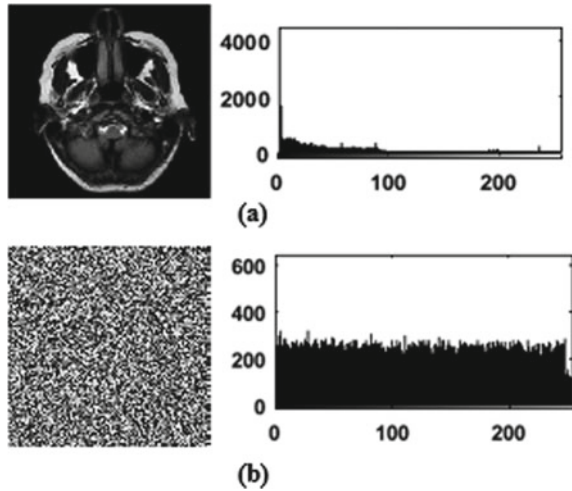
## 5 Performance and Security Analysis

A good cryptosystem should be able to resist different attacks like differential and statistical attacks. The level of security of the cryptosystem can be proved by different analysis like correlation coefficient, histogram, and entropy analysis. These analyses are used to find out how much random and uniform distribution of pixels are achieved by applying the developed cryptosystem. Cipher attack analysis is used to find out the level of robustness. These analyses are executed, and the results are discussed in the following subsections.

### 5.1 Histogram Analysis

This analysis is mainly used to evaluate the uniform distribution of pixels in an encrypted image [15]. Figure 5 shows the graphical form of pixel distribution in an original and encrypted image. On analysing the diagram, it can be cleared that the pixel in cipher image is uniformly distributed, so the proposed architecture strongly resists the statistical attacks.

**Fig. 5** Histogram analysis:  
**a** original image;  
**b** encrypted image

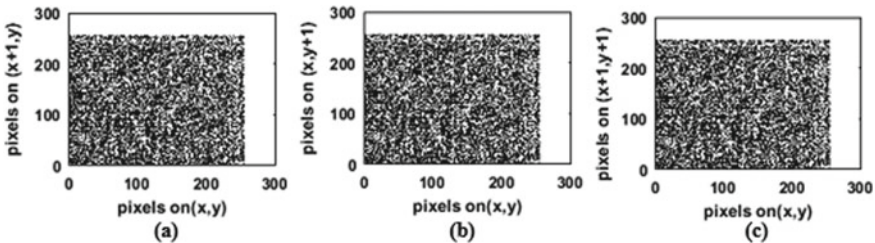


### 5.2 Correlation Coefficient Analysis

The correlation between the pixels in the original image is always high particularly for medical images [16]. A good cryptosystem should reduce such correlation among the pixels. The efficient way to find the security level based on correlation is to execute the correlation coefficient analysis between original and encrypted image. Equation (12) represents the computation, and Fig. 6 shows the distribution of pixels in all direction of the cipher image. From Fig. 6, it can be recognized that the proposed scheme efficiently reduced the correlation among pixels so it can withstand statistical attack.

$$C_{pq} = \frac{E[(p - E(p))(q - E(q))]}{\sigma_p \sigma_q} \tag{12}$$

The expected value of  $p$  and  $q$  is given by  $E(p)$  and  $E(q)$ , and the standard deviation of  $p$  and  $q$  is given by  $\sigma_p$  and  $\sigma_q$ .  $p$  and  $q$  represent the intensity of the image.



**Fig. 6** Correlation coefficient analysis: different direction of cipher image correlation of **a** horizontal; **b** vertical; **c** diagonal



**Table 1** Correlation coefficient and entropy results and comparison

Test images	Proposed scheme				Enayatifar et al. [17]			
	HC	VC	DC	Ent	HC	VC	DC	Ent
4a	0.0081	-6.3e <sup>-04</sup>	0.0095	7.9978	0.0127	0.0068	0.0028	7.9987
4b	-0.007	0.006	-0.009	7.9950	0.0046	0.0059	0.0041	7.9987
4c	0.0038	-0.008	-0.001	7.9915	0.0083	0.0147	0.0016	7.9981
4d	-1.5e <sup>-04</sup>	-0.010	0.0076	7.9916	0.0142	0.0079	0.0057	7.9990
4e	0.0018	0.0054	-0.009	7.9968	0.0158	0.0115	0.0086	7.9987
4f	-7.9e <sup>-04</sup>	0.0145	0.0097	7.9959	0.0029	0.0031	0.0013	7.9988
4g	2.12e <sup>-04</sup>	-0.003	0.0161	7.9933	0.0084	0.0062	0.0021	7.9982
4h	0.0089	-0.007	-0.007	7.9967	0.0061	0.0097	0.0019	7.9988

The correlation result of the standard test images and their comparison is shown in Table 1. In table, HC, VC, and DC represent horizontal, vertical, and diagonal correlation, and Ent indicates Entropy. From the comparison, it can be identified that the proposed system highly reduced the correlation among pixels than the state of the art.

### 5.3 Information Entropy

Degree of uncertainties is the most important features of randomness, and it can be calculated by entropy analysis [18]. The general entropy value for the random is exactly 8. So if any cryptosystem achieved the result of entropy as nearest to 8, then it can withstand against linear attacks. Entropy result of the encrypted image is calculated by employing Eq. (13).

$$E(I) = \sum_{k=0}^M P(I_k) \log_2 \frac{1}{P(I_k)} \tag{13}$$

$I_k$  represents the  $k$ th pixel value of  $M$  size medical image.  $P(I_k)$  means the probability of  $(I_k)$ . The entropy result of standard medical image and its comparison is shown in Table 1. From the result, it can be seen that all the entropy of the



encrypted image is nearest to 8; hence, the proposed cryptosystem can fight against linear attacks.

## 5.4 Key Space Analysis

The size of the keys is a key role to justify the security level of the cryptosystem [19]. Sufficient size keys resist brute-force attack. The proposed cryptosystem has two system parameter keys and two initial value keys. The four keys  $\alpha$ ,  $\beta$ ,  $u_0$  and  $v_0$  are decimal numbers. As per IEEE decimal format standard, the precision is about  $10^{-15}$ , so the total key size will be  $>2^{128}$  which is sufficient to fight against brute-force attacks.

## 5.5 Cipher Image Attack Analysis

On the transmission of encrypted image, sometimes hackers distort the image by simply cropping and adding noise to the transmitted image. So even the authenticated receiver who has correct keys, he is not able to get exact original image after decryption. Because a tiny change in the encrypted image can fully collapse the result of decryption. Hence, a good cryptosystem should be able to withstand cipher image attacks. So, our cryptosystem has been tested by applying cropped attack and noise attack.

### 5.5.1 Cropped Attack Analysis

Robustness of the cryptosystem can be identified by analysing the result of crop attack [15]. The familiar logic for applying crop attack is crop the encrypted image intentionally with different size. Thus, the encrypted image is cropped, and the resultant image is shown in Fig. 7. Table 2 shows the quality of the decrypted image by calculating the quality metrics MSE, PSNR, and correlation between original and decrypted image. The mathematical function is used to calculate the MSE and PSNR which are described in Eq. (14). This result illustrates that our cryptosystem can withstand cipher image attack.

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (O(i, j) - D/E(i, j))^2$$

$$\text{PSNR} = 20 \log_{10} \frac{I_{\max}}{\sqrt{\text{MSE}}} \quad (14)$$

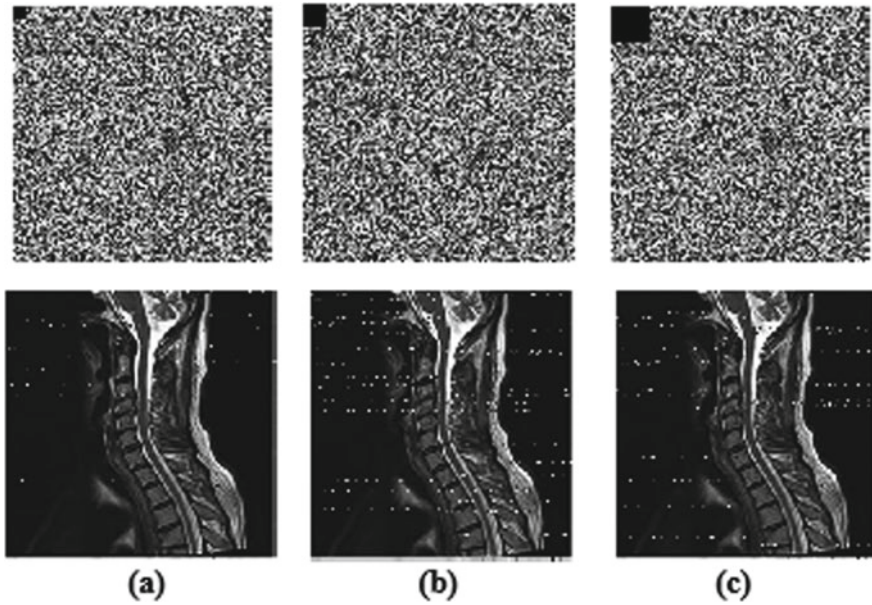


Fig. 7 Crop attack analysis: data loss of a 0.5%; b 1.0%; c 2.5%

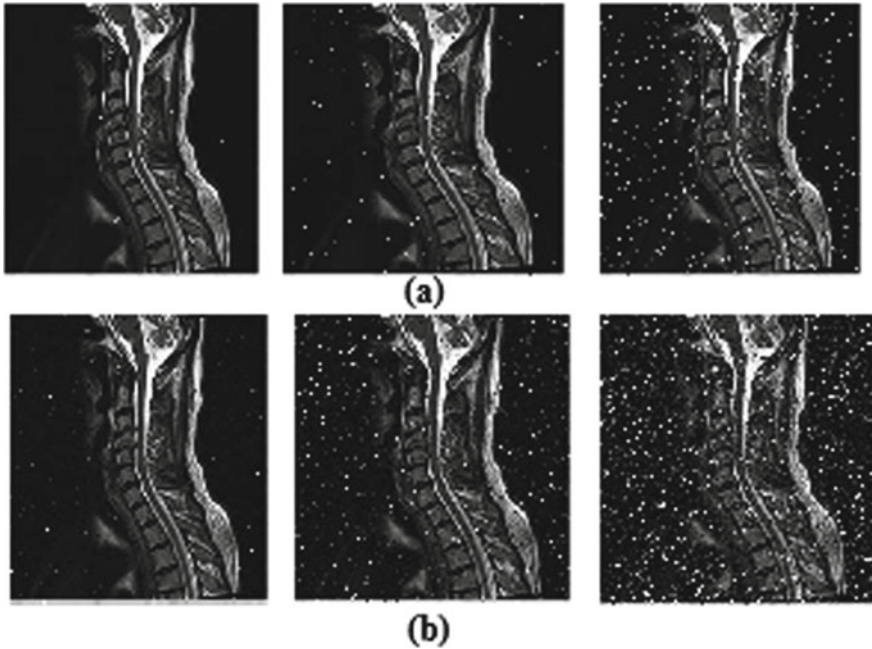
Table 2 Quality of decrypted image against crop and noise attack analysis

Cipher image of Fig. 4g	Affected ratio	MSE	PSNR	Correlation
Crop attack	0.5%	49.50	31.18	0.9901
	1.0%	144.37	26.53	0.9718
	2.5%	364.92	22.51	0.9317
Salt-and-pepper noise	0.001	18.70	35.41	0.9963
	0.01	166.98	25.90	0.9676
	0.02	819.03	18.99	0.8591
Gaussian noise	0.00001	154.95	26	0.9697
	0.0001	507.61	21	0.9066
	0.001	1714	15	0.7348

O and D indicate the original and decrypted image. M and N represent the image size.

### 5.5.2 Noise Attack Analysis

Sometimes, during transmission noises are added to the images, salt-and-pepper noise and Gaussian noise are the popular noise embed with the encrypted image [20].



**Fig. 8** Noise attack analysis: **a** salt-and-pepper noise of 0.001, 0.01, 0.02; **b** Gaussian noise of 0.00001, 0.0001, 0.001

To prove the robustness, this noise is added to the encrypted image with different densities and their corresponding decrypted image is given in Fig. 8 and the quality metrics values of the decrypted image are given in Table 2. On seeing Fig. 8 and Table 2, it can be accepted that the developed cryptosystem withstand up to 0.001 and 0.00001 amount of salt-and-pepper and Gaussian noise attacks.

## 6 Conclusion

In this paper, a new image cryptosystem has been developed by utilizing the fusion THS map. The range of initial values and system parameters of THS map increased the security level. In the confusion stage, image pixels are scrambled to reduce the correlation among the pixels, and in diffusion stage, the complexity of the cryptosystem is increased by execution XOR between the confused image and chaotic key image. Statistical, key size, cipher image attack analyses are executed to illustrate the level of security of the proposed cryptosystem. The results conclude that the proposed one satisfied the expected level of security for secure transmission and storage of medical images in real-time online medical application.

## References

1. Karthiban MK, Raj J (2019) Big data analytics for developing secure internet of everything. *J ISMAC* 1
2. Bindhu V (2019) Biomedical image analysis using semantic segmentation. *J Innov Image Process* 1:91–101
3. Liu H, Wen F, Kadir A (2019) Construction of a new 2D Chebyshev-Sine map and its application to color image encryption. *Multimed Tools Appl* 78:15997–16010
4. Mollaefar M, Sharif A, Nazari M (2017) A novel encryption scheme for colored image based on high level chaotic maps. *Multimed Tools Appl*. 76:607–629
5. Hu T, Liu Y, Ouyang C (2017) An image encryption scheme combining chaos with cycle operation for DNA sequences. *Nonlinear Dyn* 87:51–66
6. Rajendran S, Doraipandian M (2018) Biometric template security triggered by two dimensional logistic sine map. *Procedia Comput Sci* 143:794–803
7. Fu C, Meng Wh, Zhan Yf, Zhu Zl, Lau FCM, Tse CK, Ma Hf (2013) An efficient and secure medical image protection scheme based on chaotic maps. *Comput Biol Med* 43:1000–1010
8. Ravichandran D, Praveenkumar P, Balaguru Rayappan JB, Amirtharajan R (2016) Chaos based crossover and mutation for securing DICOM image. *Comput Biol Med* 72:170–184
9. Sathishkumar GA, Bhoopathybagan K, Sriraam N, Venkatachalam SP, Vignesh R (2011) A novel image encryption algorithm using two chaotic maps for medical application. *Commun Comput Inf Sci* 133:290–299
10. Cao W, Zhou Y, Chen CLP, Xia L (2017) Medical image encryption using edge maps. *Signal Process* 132:96–109
11. Rajendran S, Krithivasan K, Doraipandian M (2020) Fast pre-processing hex chaos triggered color image cryptosystem. *Multimed Tools Appl* 12447–12469
12. Huang X (2012) Image encryption algorithm using chaotic Chebyshev generator. *Nonlinear Dyn* 67:2411–2417
13. Borujeni SE, Eshghi M (2013) Chaotic image encryption system using phase-magnitude transformation and pixel substitution. *Telecommun Syst* 52:525–537
14. Mokhtar MA, Sadek NM, Mohamed AG (2017) Design of image encryption algorithm based on different chaotic mapping. *Natl Radio Sci Conf NRSC, Proc*, pp 197–204
15. Chai X, Fu X, Gan Z, Lu Y, Chen Y (2019) A color image cryptosystem based on dynamic DNA encryption and chaos. *Signal Process* 155:44–62
16. Al-Najjar H, Alharthi S, Atrey PK (2016) Secure image sharing method over unsecured channels. *Multimed Tools Appl* 75:2249–2274
17. Nematzadeh H, Enayatifar R, Motameni H, Guimarães FG, Coelho VN (2018) Medical image encryption using a hybrid model of modified genetic algorithm and coupled map lattices. *Opt Lasers Eng* 110:24–32
18. Hua Z, Yi S, Zhou Y (2018) Medical image encryption using high-speed scrambling and pixel adaptive diffusion. *Signal Process* 144:134–144
19. Zhang X, Zhao Z, Wang J (2014) Signal processing: image communication chaotic image encryption based on circular substitution box and key stream buffer. *Signal Process Image Commun* 29:902–913
20. Babaei A, Motameni H, Enayatifar R (2020) A new permutation-diffusion-based image encryption technique using cellular automata and DNA sequence. *Optik (Stuttg)* 203

# User Engagement Recognition Using Transfer Learning and Multi-task Classification



Hemant Upadhyay, Yogesh Kamat, Shubham Phansekar, and Varsha Hole

**Abstract** Digital learning and the virtual classroom had an enormous influence during this new era of modernization, which has brought a revolution on the interested student to acquire at their own comfort as well as the desired pace of learning. But the important success factor of traditional classroom pedagogy, i.e., real-time content delivery feedback, is missing. Eventually, engagement becomes crucial to strengthen and improve user interaction. It constrained us to resolve this gap issue before, because of the lack of publicly accessible datasets. But, now since DAiSEE is the first multi-mark video grouped dataset which comprises student's recordings for perceiving the client's full of feeling conditions of boredom, confusion, engagement and frustration in wild, that makes it a benchmarked dataset for solving such kind of problems. In this paper, the model to employ on our convolutional neural networks is a custom Xception Model, a depth-wise separable convolution pre-trained on the billion of images along with multi-task classification layers on it to recognize the affective states of user efficiently. The proposed Xception network slightly outperforms the frame-level classification benchmarked by DAiSEE's model.

**Keywords** Convolutional neural network · Transfer learning · Engagement · Xception · Multi-task classification

---

H. Upadhyay (✉) · Y. Kamat · S. Phansekar · V. Hole  
Sardar Patel Institute of Technology, Munshi Nagar, Andheri West, Mumbai, Maharashtra  
400058, India

e-mail: [hemant.upadhyay@spit.ac.in](mailto:hemant.upadhyay@spit.ac.in)

Y. Kamat

e-mail: [yogesh.kamat@spit.ac.in](mailto:yogesh.kamat@spit.ac.in)

S. Phansekar

e-mail: [shubham.phansekar@spit.ac.in](mailto:shubham.phansekar@spit.ac.in)

V. Hole

e-mail: [hole@spit.ac.in](mailto:hole@spit.ac.in)

## 1 Introduction

Nowadays, computer-based learning is becoming popular in academia and can involve multiple ways, varying Massive Open Online Courses (MOOCs) [1], Learning Management System (LMS) [2] or Virtual Learning Environment (VLE) [2]. Through this, students will study at all times and almost everywhere, also helps to deliver innovative ways of teaching students, reforming the traditional studies approach and drawing students from around the world. NPTEL, Udemy, Coursera, edX, etc., are the few best-known Web sites [2]. The Online Learning Alliance reported that at least one online course was pursued by 6.7 million students (about 32% of all schools). Nevertheless, in the study of more than 2000 universities, 69% of senior academic coordinators thought that virtual learning [3] was necessary for the prosperity of their institutions, with the majority they agree its better than conventional learning [2].

Under the conventional teaching method, teachers take different measures to evaluate the progress of their graduates, enthusiasm and engagement, via tests, attendance, observation, etc. However, there is no face-to-face contact that is accessible in computer-based learning, and the amount of student engagement in activities is difficult to define [4].

In short, these MOOCs do not provide instantaneous feedback to students (or instructors), as compared to conventional classroom learning. Strangely, the MOOCs have a dropout rate of more than 90% with a completion rate of about 45% for the first task. The significant factors why the platforms are abandoned which are mentioned in an online survey include poor course design with lack of appropriate reviews, reading exhaustion, lack of relevancy of topics. That is why review systems ought to be strengthened to render these services more social [3].

Knowing the engagement of the users at various levels of e-learning may help to develop intuitive interfaces that enable students to properly integrate their content, minimize dropout rates and personalize the learning experience [1]. The degree to which users participate positively by thought, communication and engaging with the content of a course, the other fellow mates and their instructors are usually termed as student engagement. Research interest is important when developing smart instructional applications in diverse learning settings, like interactive apps, free vast online courses and smart tutoring programs. For example, if students are frustrated, the program should interfere and re-engage them. To enable the learning method to change its learning atmosphere and provide the students with the right responses and must first evaluate the engagement automatically, but it becomes difficult since there is no publicly available huge dataset and building own custom data is expensive [3].

This paper proposes a deep convolutional neural network to boost engagement identification in the wild that overcomes the data sparsity problem. By using transfer training, i.e., re-train the model on a pre-build, readily available trained model (Xception) [5], change a few layers accordingly and re-train on custom data. Here, the data which are being used is DAiSEE [1] (first benchmarked dataset for engagement detection in wild). The remaining paper is structured as the literature survey and

related study are covered in Sect. 2. In Sect. 3, datasets, transfer learning (Xception Model) are presented and proposed modeling methodology. The performance of the proposed model is included, and discussed results in Sects. 4 and 5 include recap of our work and research.

## 2 Background and Related Work

In the background works, research relevant to engagement detection is presented. Xiang et al. proposed a system to identify and examine the effects of divided attention on mobile phones during the learning process as they are likely to multi-task while using MOOCs unlike in the traditional classroom learning process. This research was conducted on 18 users and showed that the impact is vigorously deceptive on learning outcomes and photoplethysmography (PPG) [6] was implicitly captured to analyze it.

Whitehill et al.'s [7] interesting work examined approaches for automatic recognition of facial expressions of students and found that teachers constantly assess their students in their degree of engagement, with facial facets as head posture and basic facial actions, such as brow-raising, eye closure and high lip raising, playing a key role. They also discovered that judgments of 10-second clip engagement can be approximated with an average of 10 s in single framework judgments. The paper reveals that after an evaluation test, correlation is estimated fairly (and statistically significantly) between humans and structures.

Since in MIL, labeling of engagement at frequent intervals in user videos is expensive and noisy. The dataset annotation was done by using crowdsourcing, and labeling was done based on the intensity of engagement, i.e., highly engaged, engaged, barely engaged, disengaged.

Before their research, there was no publicly, huge and labeled dataset available as many of them used their custom-made datasets. They created the DAiSEE dataset, a benchmarked dataset for further this kind of problems. For analyzing they used EmotionNet [1] and before passing the video as input to the model, the input is forwarded via FaceNet [1] to detect and extract faces. Also, they have identified and mentioned a few challenges of their publicly available dataset.

Prabin Sharma et al. made a system that distinguishes the eye and head posture and consolidates them with the face emotions to identify three degrees of engagement. They utilized two calculations significant to the understudy's front face, i.e., Haar cascade algorithms [8], to identify the position of eye and convolution neural network that was trained in eye images to see whether the student was defying a web camera and played out a twofold evaluating in these two classifications. Another was prepared with grayscale pictures to arrange facial feelings, i.e., "angry," "disgust," "fear," "happy," "sad," "surprise," "neutral," i.e., on FER13 [9] datasets, called mini-Xception model [8]. This investigation was tried on 15 understudies, and the outcomes show that the framework effectively recognizes each time the students were "very engaged," "nominally engaged" and "not engaged at all," what's more,

the outcomes additionally show that students with the best scores likewise have a higher engagement.

### 3 Data Set and Transfer Learning

#### 3.1 *DAiSEE Dataset*

DAiSEE [1] is the first multi-label video classification dataset comprising 9068 video snippets captured from 112 users belonging to the age group of 18–30, all of whom were currently enrolled students having race as Asian, with 32 females and 80 males. There were a total of 12583 video snippets, each 10 s long as Whitehill et al. [7] observed that 10-second labeling tasks are more than enough for identifying the user’s affective states “in the wild.” The dataset was classified into four levels for each of the affective states, as studies show that the six basic expressions, i.e., anger, disgust, fear, joy, sadness, and surprise, are not reliable in prolonged learning situations, as they are prone to rapid changes. Also, crowd annotators label the video as neutral when they are unsure, to avoid this they follow this labeling strategy to make a robust dataset. After data cleaning, they end up with a dataset that has 9068 video snippets, and each numbered uniquely to help differentiate between the settings of video snippets.

#### 3.2 *Importance of Transfer Learning [10]*

##### **Introduction to Transfer Learning**

Traditional machine learning algorithm works smoothly when the model has been trained and tested on some dataset which has similar feature space and appropriation. At the point when these changes, it is expected to remake the model without any preparation utilizing your predefined custom information, which is very costly or difficult to remember the recently prepared information. Transfer learning between task spaces would be attractive. Transfer learning [10] is an ML approach in which a model produced for an errand is reutilize for another assignment. In basic machine learning approach, it tries to learn each task from scratch, while transfer learning attempts to move the knowledge from some past assignments to an objective undertaking when the last has less high-quality information [10].

##### **Common approaches in Transfer Learning**

In transfer learning, the accompanying three principles explore issues featured beneath:



1. What to transfer: Describes which portion of information can be moved, here and there it might have a place with an individual space or common between various areas.
2. How to transfer: Focuses on various learning algorithms that are required to move the information.
3. When to transfer: Ask at what time transferring of information ought to be finished.

Sometimes, the source space and target area are not pertinent, and results of the transfer become unsuccessful. Once in a while, it might debase the exhibition of learning in the objective area which is alluded to as negative transfer [11].

Some basic ways to deal with the new task to be performed while profiting by recently learned task:

1. Feature extraction [11] does not adjust the previous model layers and permits new layers to profit by complex features gained from past layers. Be that as it may, these features are not particular for the new layers and can regularly be improved by fine-tuning.
2. Fine-tuning [11] changes the parameters of a current CNN to prepare other layers. The output layer is reached out with arbitrarily initialized weights for the new layers, and a little learning rate is utilized to tune all parameters from their unique qualities to limit the loss on the new layers. In some cases, some layers of the models are frozen (e.g., the convolutional layers) to forestall over-fitting.
3. Adding new nodes/layers to models is an approach to save the previous model parameters while learning new features. Importance of transfer learning Oztel et al. [12] found that while training a CNN model from scratch may do not execute effectively. This circumstance is clarified with certain reasons:
  - (1) Lack of input data: For any classification task, the network appears to be huge which requires more data. If transfer learning [11] approach is used, early layers of any network will do feature extraction and thus training can finish rapidly and effectively. This is the intensity of transfer learning.
  - (2) Imbalanced input data: Unsuccessful training can be clarified with lacking transfer data and imbalanced information. Since an equivalent number of pictures has a place with each class was utilized, imbalanced input data cannot be used for this study.

Thus, when the transfer learning and learning from scratch approaches were analyzed for engagement recognition tasks, the transfer learning approach outperformed over-preparing without any preparation approach. This can be disclosed that inferable from the benefit of transfer learning, early layers of the network can extricate fundamental features quickly and effectively.

## 4 Deep Convolutional Neural Network (DCNN) Learning

Segregating emotions and recognizing what sort of emotion is a facial emotion recognition problem. Subsequently, a state-of-the-art DCNN model that performs facial emotion recognition problems ought to likewise perform well in engagement detection. In this paper, the custom Xception network has been embraced and showed engagement detection performance on the DAiSEE dataset.

### 4.1 Pre-processing

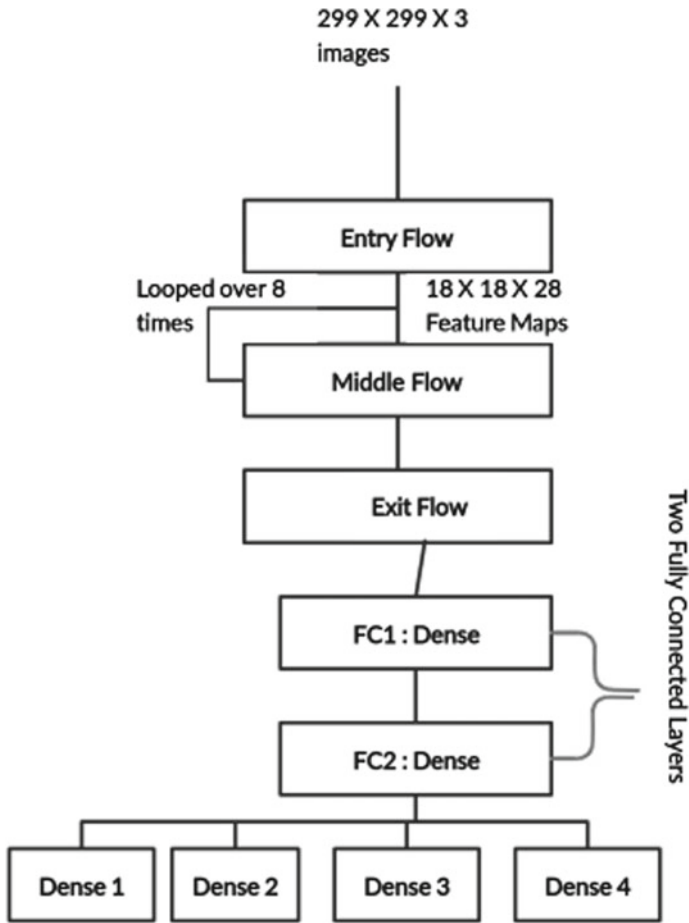
For frame extraction from the video snippets of train, test and valid, ffmpeg method having frame size as 0.7 fps has been used. Since it did not have a lot of incredible figuring assets, the Tensorflow Data API(tf.data) [13] is utilized, as it makes it conceivable to deal with a lot of information, read from various data formats and perform complex changes. To reshape the image in the size of  $299 \times 299 \times 3$  because that is the size of input which Xception Model takes in, bi-linear interpolations method is used and set “anti-alias” is true since when downsampling an image with anti-aliasing the sampling filter kernel is scaled to properly anti-alias the input image signal. [13] The total count of the number of frames for train, test, validate was 37,506, 12,488, 10,003. Along with, only test data frames are also extracted with default size for model generalization and its total count was 536,416.

### 4.2 Model Architecture

The input to our engagement model is an RGB scale image at  $299 \times 299 \times 3$  resolution. The output is four classes: engagement, frustration, boredom and confusion. Our custom Xception model has appeared in Fig. 1. The model is separated into three stages: first the data initially flows to entry phase, after that passes through the middle flow which is looped eight times, lastly through the exit flow after that, fine-tuned it by including four dense classification layers having loss function as Sparse Categorical Cross-entropy [13] that will perform multi-task classification of our dataset.

## 5 Experimental Results

Each affective state is classified using the given models:



**Fig. 1** Custom Xception Model having four individual dense classification layers used for identifying different levels of classes of our output

With no fully connected layers [9], no fine-tuning [9]

S. No.	Classification	Accuracy (%)
1	Boredom	42.72
2	Engagement	47.9
3	Confusion	67.24
4	Frustration	44.29

With no fully connected layers, fine-tuning

S. No.	Classification	Accuracy (%)
1	Boredom	41.94
2	Engagement	50
3	Confusion	67.55
4	Frustration	42.29

With fully connected layers, no fine-tuning

S. No.	Classification	Accuracy (%)
1	Boredom	46.12
2	Engagement	44.21
3	Confusion	67.33
4	Frustration	44.92

With fully connected layers, fine-tuning

S. No.	Classification	Accuracy (%)
1	Boredom	43.82
2	Engagement	43.93
3	Confusion	67.42
4	Frustration	43.49

Figure 2 shows the baseline accuracy results from our studies. From that, came to see that model with fine-tuning and no fully connected layers generally perform better than all other models, Also, there has been slightly outperformed the frame-level classification benchmark set given by DAiSEE [1]. Additionally, tried to make this custom model more generalized by setting default rate during ffmpeg frame conversion while converting the test dataset and tested it with the model and came to know that choosing 0.7 as fps does not affect the performance of the model. Table 1 displays the Top-1 accuracy obtained on the model with fully connected layers and no fine-tuning.

## 6 Compliance of Ethical Standards

### 6.1 What Will Be Done with Your Paper

This study was funded by no individual/institution.

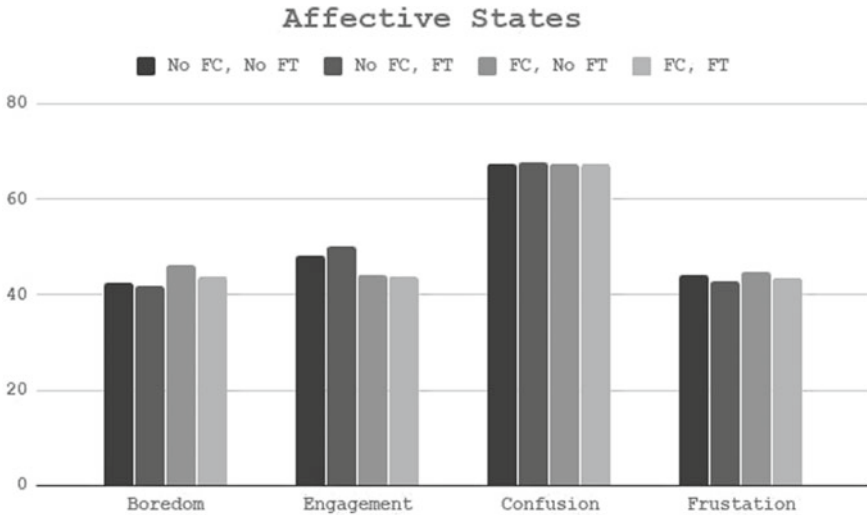


Fig. 2 Baseline accuracy results of proposed study

**Table 1** With fully connected layers, no fine-tuning on default rate while framing conversion

S. No.	Classification	Accuracy (%)
1	Boredom	44.66
2	Engagement	45.17
3	Confusion	67.22
4	Frustration	44.55

**Table 2** With fully connected layers, fine-tuning on default rate while framing conversion

S. No.	Classification	Accuracy (%)
1	Boredom	43.40
2	Engagement	44.49
3	Confusion	67.40
4	Frustration	43.94

## 6.2 Conflict of Interest

Author Hemant Upadhyay has received research grant on DAiSEE dataset by filling the form and accepting norms given by them on their site.

## 7 Conclusion

In this study, a custom Xception model is devised, which is a depth-wise separable convolution trained on ImageNet dataset along with DAiSEE dataset, the first multi-label video classification dataset for recognizing user's affective states with multi-task classification layers for classifying four different classes of our dataset, i.e., engagement, boredom, frustration and confusion. Since, having not much powerful computing resource, tried to optimized this engagement detection task efficiently, by using Tensorflow DataAPI, ffmpeg and many more. These models have been assessed as for the exhaustive scope of pattern models to exhibit its adequacy and demonstrated that it prompts significant improvement against a benchmarked model.

## References

1. Gupta A, D'Cunha A, Awasthi K, Balasubramanian V (2018) DAiSEE: towards user engagement recognition in the wild, arXiv preprint: [arXiv:1609.01885](https://arxiv.org/abs/1609.01885)
2. Dixson MD (2015) Measuring student engagement in the online course: The Online Student Engagement Scale (OSE). *Online Learn* 19(4):n4
3. Kaur A, Mustafa A, Mehta L, Dhall A (2018) Prediction and localization of student engagement in the wild. In: 2018 digital image computing: techniques and applications (DICTA). IEEE, pp 1–8
4. Hussain M et al (2018) Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Comput Intell Neurosci* 2018
5. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*
6. Xiao X, Wang J (2017) Understanding and detecting divided attention in mobile MOOC learning. In: *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp 2411–2415
7. Whitehill J, Serpell Z, Lin Y-C, Foster A, Movellan JR (2014) The faces of engagement: automatic recognition of student engagement from facial expressions. *IEEE Trans Affect Comput* 5(1):86–98
8. Sharma P, Joshi S, Gautam S, Filipe V, Reis MJCS (2019) Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. arXiv preprint [arXiv:1909.12913](https://arxiv.org/abs/1909.12913)
9. Nezami OM, Dras M, Hamey L, Richards D, Wan S, Paris C (2018) Automatic recognition of student engagement using deep learning and facial expression. arXiv preprint [arXiv:1808.02324](https://arxiv.org/abs/1808.02324)
10. Pan SJ, Yang Q (2009) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22(10):1345–1359
11. Li Z, Hoiem D (2017) Learning without forgetting. *IEEE Trans Pattern Anal Mach Intell* 40(12):2935–2947
12. Oztel I, Yolcu G, Oz C (2019) Performance comparison of transfer learning and training from scratch approaches for deep facial expression recognition. In: 2019 4th international conference on computer science and engineering (UBMK). IEEE
13. Tf.data: Build Tensorflow input pipelines: Tensorflow Core. <https://www.tensorflow.org/guide/data>

# QoS Aware Multi Mapping Technology in SD-WAN



Viswanathan Varsha and C. N. Sminesh

**Abstract** SDN separates the control plane from the data plane and thus provides dynamic and flexible management of the network. Using a single controller in the control plane will create a single point of failure. To overcome this problem, multiple controllers are used in the control plane. Multi-controller architecture has two kinds of mapping between switch and controller, single mapping, and multi mapping. The single mapping maps each switch to a single controller. When one controller is down, then another suitable controller in the slave role needs a role change, and this will increase the recovery time. While in the proposed multi mapping technology, each switch is mapped to two controllers. When one controller fails or overloads, then the other controller takes over the switch. Since the other controller is also in an equal role, recovery time is better. QoS parameters for both single mapping and multi mapping are compared using SWITCH topology and Bandcon topology from Internet topology zoo. For switch topology, multi mapping has reduced packet loss by 38%, reduced end-to-end delay by 21%, and increased throughput by 36% in comparison with single mapping. For Bandcon topology, multi mapping has reduced packet loss by 43%, reduced end-to-end delay by 36%, and increased throughput by 45% in comparison with the single mapping.

**Keywords** Software-defined networks · Multi mapping · Single mapping · Quality of service · Modified density peak clustering algorithm · End-to-end delay · Packet loss · Throughput

---

V. Varsha (✉) · C. N. Sminesh  
Government Engineering College, Thrissur, India  
e-mail: [varsha.vaishakam@gmail.com](mailto:varsha.vaishakam@gmail.com)

C. N. Sminesh  
e-mail: [smineshcn@gmail.com](mailto:smineshcn@gmail.com)

## 1 Introduction

A traditional network consists of the vertically integrated control plane and data plane. The control plane has a forward table that has the information about where to send packets. Hence, the forwarding table is used for decision-making. The data plane takes information from the control plane and sends packets according to that from one switch to another. In a traditional network, both of these planes exist on the networking device itself. This is not suitable for day by day growing network communication. This led to the emergence of the software-defined network. This helps to overcome the limitations of the traditional network. In SDN, the data plane and control plane are decoupled. A programming interface is used between the controller and switches. The controller and switches interact with each other using the southbound interface, and the application interacts with the controller using the northbound interface. Open-flow is one of the first software-defined networking standards. It is a communication protocol that helps the controller to interact with the data plane devices. In SDN, the control plane has global network information and global application information. The network can be reprogrammed without handling the data plane elements.

A single centralized controller architecture is simple and provides easy network management. This is suitable for small and medium-sized networks. The limitation of using a single controller is a single point of failure, and this destroys the entire network. Multiple controller architectures are used to overcome the problems of single controller architecture.

Instead of using a single controller, multiple controllers can be used to overcome a single point of failure. SDN with multiple controllers are suitable for large networks. Multiple controllers will handle the whole network. Each controller will have a domain to control. Even if one controller fails, the other controllers are there to handle the network. Thus, a single point of failure can be avoided.

Sridharan et al. [1] proposed a multi mapping method for the switch-controller mapping. In multi-controller-based SDN, the mapping between switch and controller can be done in two ways—single mapping and multi mapping. In single mapping, each switch is mapped to a single controller. In multi mapping, each switch is mapped to multiple controllers.

In single mapping, when controller failure is detected, the switch needs to find another controller to which it can send its flow setup requests. A suitable controller needs to be found from all the controllers available in the network. The other controllers are in the slave role to this switch which needs a role change to master or equal role. This will increase the recovery time.

While in multi mapping, both controllers are in an equal role to the switch. When one controller fails or overloads, the other controller controls the switch. Since the other controller is also in an equal role, the recovery time is better. For simplicity, each switch is mapped to two controllers. Flow requests from a switch are alternatively handled by these two controllers. For the first half of the simulation time, the best controller will handle the flow setup request from the switch. For the next half of



the simulation time, the second-best controller will be used. The result for the best and second-best controllers for every switch is taken from the previous work done by Smineesh et al. [2].

The performance metrics chosen are packet loss, throughput, and end-to-end delay. Single mapping is compared with multi mapping using these metrics. SWITCH and Bandcon topologies are taken from the Internet topology zoo for simulation. Mininet along with MiniEdit is used to simulate the topologies. POX controller is a python-based controller, and its default port is 6633. D-ITG is used to generate packets and monitor the QoS parameters for single mapping and multi mapping.

The rest of the paper is arranged as follows. Section 2 discusses the existing literature. Section 3 details the proposed strategy using multi mapping, while Sect. 4 describes the simulation, and Sect. 5 describes the performance analysis of the proposed system. The paper is concluded in Sect. 6.

## 2 Literature Survey

SDN is a network paradigm that decouples the data plane from the control plane [3, 4]. This separation provides more flexibility to the network. Traditional networks are more complex; SDN provides better performance than traditional networks. The control plane decides, where to forward the packets. The Data plane consists of switches that simply forward the packets. In SDN, with a single controller, each switch sends its flow setup request to this single controller and controller responses without a packet-out message. A single controller needs to respond to every single request. SDN with a single controller in the control plane is more vulnerable, and there is a high chance of a single point of failure.

This can be overcome by using multiple controllers in the control plane. Multiple controllers in the control plane protect the network in case of a single point of failure [5]. Even if one of the controllers fails, the other controller will take responsibility. Each controller will be having several switches assigned to it. Each switch sends its flow setup request to its assigned controller. Multi-controller-based SDN provides more stability to the network. The multiple controllers introduce new challenges to SDN.

In software-defined wide area networks (SD-WANs), finding an optimal number of controllers and their placement is termed as the controller placement problem. Wang et al. [6] surveyed the controller placement problem in software-defined networking. Smineesh et al. [7] proposed an optimal controller placement strategy using the exemplar-based clustering approach for software-defined networking. In SD-WANs, computationally simple solutions for controller placement can be developed using non-parametric clustering algorithms. DP clustering is a density-based non-iterative clustering algorithm. Smineesh et al. [2] proposed the modified DP which yields the minimum utility value for the controller placement in all the selected network topologies. This result can be used to map every switch to its best and

**Table 1** The best second-best controller for SWITCH topology

Switches	Best controller	Second-best controller
1, 2, 4, 5, 6, 7, 8, 9, 11, 14, 15, 16, 17, 18, 19, 20, 17, 28, 29	C1	C2
0, 3, 10, 12, 13, 21, 23, 24, 25, 26	C2	C1

**Table 2** The best second-best controller for Bandcon topology

Switches	Best controller	Second-best controller
0, 2, 3, 12, 13	C1	C2
4, 5, 6, 7, 8, 9, 14, 15, 16	C2	C1
1, 10, 11, 17, 18, 19, 20, 21	C3	C2

second-best controller. From this result, switch-controller mapping can be done successfully.

The SWITCH topology from Internet topology zoo consists of 30 switches, and each switch having one host. According to [2], SWITCH topology uses two controllers in the network. The result for the best and second-best controller for every switch is taken from [2] and is shown in Table 1.

The Bandcon topology from Internet topology zoo consists of 22 switches, and each switch having one host. According to [2], Bandcon topology uses three controllers in the network. The result for the best and second-best controller for every switch is taken from [2] and is shown in Table 2.

Sridharan et al. [8] proposed multi mapping in SDN. The mapping between switch and controller can be done in two ways. The first one is single mapping, where every switch is mapped to a single controller. The second one is multi mapping, where each switch is mapped to multiple controllers. This provides protection even in the case of single controller failure. For this purpose, the best and second-best controller for every switch needs to be found. Another challenge of SDN with multiple controllers is balancing the load among controllers. Wang et al. [9] proposed a switch migration-based decision-making scheme for balancing the load. Migration efficiency is used as a matrix to improve switch migration in SMDM. Migration efficiency is used to find the migration target controller. Migration efficiency consists of both migration cost and load variation. Switch migration can be divided into subproblems. SMDM is used migration efficiency to find the migration cost. In the SMDM, the primary goal is to select the target controller and the switch to migrate. Cui et al. [10] proposed a load balancing scheme based on response time. Whenever a switch gets a new flow request, this needs a rule installation. This creates a response delay in the network. In this scheme, the load balancing strategy is based on the response time. Load balancing includes the selection of an overloaded controller, selection of migrating switch, and selection of migration target.

Another method for load balancing is proposed by Xu et al. [11]. The static mapping between switches and controllers makes it difficult to adapt to different traffic variations in the control plane. Some controllers may be overutilized, and others may be underutilized. Thus, dynamic mapping is needed between switches and controllers. This is done by migrating switches from the overloaded controller to light loaded controllers. This will provide load balancing in the control plane. In Balcon, a minimum number of switches is migrated. Balcon is a heuristic solution used to provide load balancing in SDN through switch migration. The best switch migration is based on the communication pattern of switches. That is, switches with a strong connection are assigned to the same controller. BalCon and BalConPlus provide load balancing by migrating the minimum number of switches. It migrates switches with a strong connection.

Another challenge in SDN is fault tolerance. Isong et al. [12] proposed a controller fault tolerance framework for small to medium-sized networks. Zhang et al. [13] proposed a resource-saving replication for controllers in multi-controller SDN against network failures. Aly and Al-anazi [14] proposed an Enhanced Controller Fault Tolerant (ECFT) model for software-defined networking.

Sridharan et al. [1] proposed a load balancing in SDN with multi mapping. Usually, each switch will have a primary controller and a backup in case of failure. The backup controller is activated only when the primary controller fails. This needs the role change of controllers. There is a chance that the packet will be lost. To avoid this problem, each switch is assigned to multiple primary controllers. The load of the switch is divided between multiple controllers. A minimum fraction of the load must be given to assigned controllers. QoS aware mapping is used to map each switch into multiple controllers to achieve minimum QoS for every controller.

### 3 Proposed System

Two kinds of switch-controller mapping are possible in distributed controllers—single mapping and multi mapping. In single mapping, each switch is assigned to a single controller. Whereas in multi mapping, each controller is assigned to multiple controllers. The flow requests of a switch will be distributed between multiple controllers. Here, for simplicity, each switch is mapped to two controllers. These two controllers will have an equal role for a switch. In the case of master–slave architecture, one controller has a master role, and the other controller has a slave role. In case of failure of the master controller, the slave controller will take care of the switch. The single mapping follows master slave architecture. In single mapping, when controller failure is detected, the switch needs to find another controller to which it can send its flow setup requests. A suitable controller needs to be found from all the controllers available in the network. The other controllers are in the slave role to this switch which needs a role change to master or equal role. This will increase the recovery time. While in multi mapping, both controllers are in an equal role to the switch. When one controller fails or overloads, the other controller

---

**Algorithm 1** QoS aware multi mapping

---

**Input :** Network topologies SWITCH, Bandcon;Switch list  $S = \{s_1, s_2, \dots, s_N\}$ **Output:** Set of controller pairs for each switch, Packet loss, Throughput, Delay**Procedure:** Multi Mapping()

```

1: for < each switch  $i$  in  $S$  > do
2:   Select the best and second best controller for every switch using modified density
   peak algorithm.
3:   switch  $i$  send flow request to best controller
4: end for
5: for < each half of simulation time > do
6:   Send flow request from switch to the other controller.
7: end for
8: for < each flow > do
9:   Compute evaluation metrics such as Packet loss, Delay, and Throughput
10: end for

```

---

**Fig. 1** Algorithm-QoS aware mapping

takes over the switch. Since the other controller is also in equal role, recovery time is better. For simplicity, each switch is mapped to two controllers such as the best and second-best controller (Fig. 1).

The input for the multi mapping is network topologies such as SWITCH and Bandcon from the Internet topology zoo. The output is the values for QoS parameters. At first, every switch is mapped to its best and second-best controller. The result for the best and second-best controller is taken from the paper proposed by Sminesh et al. [2]. Initially, each switch sends its flow setup requests to the best controller. For the next half of the simulation, the time switch sends its flow setup requests to the other controller. Then, flows are created using D-ITG. For different flows, QoS parameters are evaluated.

The architecture of the proposed system can be divided into sub-modules. The integration of all these sub-modules will provide a QoS aware multi mapping technology for SD-WAN. The proposed system consists of three sub-modules; network topology creation, multi mapping, controller switching. In the first module, SWITCH and Bandcon network topologies are simulated. The second module will map each switch into its best and second-best controller. In the controller switching module, the switch will alternatively send flow setup requests to its mapped controllers.

### 3.1 Create Network Topology

The network topologies are created from the standard dataset Internet topology zoo. There are many topologies available in the dataset, but only a few of them contain the required metrics for the modified density peak clustering algorithm, geographical

location, and bandwidth. Those topologies are selected and simulated using mininet and its GUI tool miniEdit. As per Internet topology switches, hosts and links are added to the custom topology. For simulation, SWITCH and Bandcon topologies are selected.

### 3.2 *Multi Mapping*

From the modified density peak clustering algorithm, the best and second-best controllers for every switch are taken. Both of these controllers will be acts as primary controllers and have an equal role to the switch. Multiple POX controllers can be established on multiple ports. The default port is 6633. One controller is established on port 6633, and the other one is established on port 6634.

### 3.3 *Controller Switching Module*

Flow requests from a switch are alternatively handled by these two controllers. For the first half of the simulation time, the best controller will handle the flow setup request from the switch. For the next half of the simulation time, the second-best controller will handle the flow setup request from the switch. Since both controllers are in an equal role, there will be no need for role change requests or role change messages.

## 4 **Simulation and Results**

Python is used for implementing simulation using mininet. An open-flow controller [15] is a type of SDN controller that uses the open-flow protocol. The controller placement in each network topology obtained from the modified density peak clustering algorithm is simulated using mininet. SWITCH topology and Bandcon topology are used for simulation. Mininet and MiniEdit are used to simulate these topologies. SWITCH topology consists of 30 switches and 2 controllers. Every switch is mapped to both controllers as shown in Table 1. Bandcon topology has 22 switches and 3 controllers. Every switch is mapped to its best and second-best controller as shown in Table 2.

The performance of networks in both single mapping scenarios and multi mapping scenarios is analyzed from the perspective of delay, throughput, and packet loss.

The comparison between a single mapping scenario and multi mapping scenario of network topology can be carried out by sending a various number of flows. Then measure performance metrics directly from the D-ITG log file. Simulation results

show that deploying multi-controller mapping instead of a single controller mapping is better in the perspective of the performance metrics chosen.

#### 4.1 Performance Metrics

The different metrics which are used to evaluate the proposed system is the end-to-end delay, throughput, and packet loss.

- End-to-end delay: End-to-end delay is one of the important performance metrics used in telecommunications or computer networks. It is the time taken for a network to send a bit from one node to another.
- Packet loss: Packet loss indicates the number of packets lost during transmission. It can be calculated by the ratio of the number of packets lost and the total number of packets sent.
- Throughput: Network throughput is the rate (in bps—bits per sec or pps—packets per second) at which packets/bits are successfully delivered over a network channel. The data these messages belong to may be delivered over a physical or logical link or it can pass through a certain network node.

### 5 Result Analysis

#### 5.1 End-to-End Delay

The observed values of end-to-end delay in the SWITCH network topology in single mapping and multi mapping scenarios are shown (Table 3).

The observed values of end-to-end delay in the Bandcon network topology in single mapping and multi mapping scenarios are shown (Table 4).

From the tables can be observed that the end-to-end delay in a network topology with multi mapping is lesser than that with a single mapping (Figs. 2 and 3).

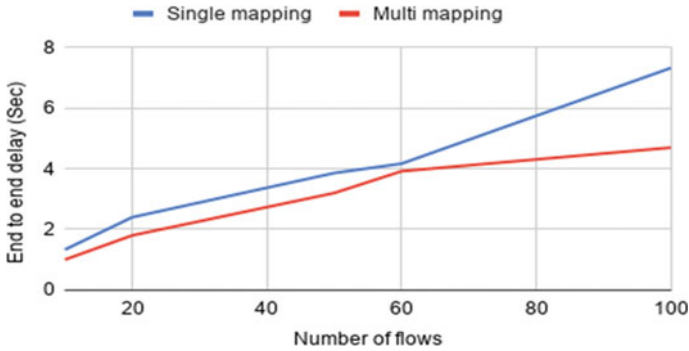
For switch topology, multi mapping has reduced end-to-end delay by 5% in comparison with single mapping. For Bandcon topology, multi mapping has reduced end-to-end delay by 21% in comparison with single mapping.

**Table 3** Delay in SWITCH topology

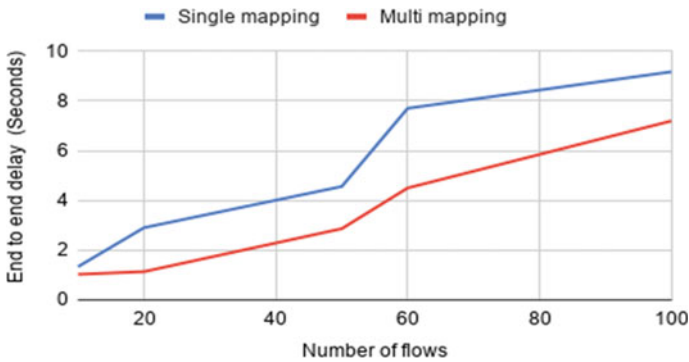
No. of flows	Single mapping (s)	Multi mapping (s)
10	1.33	1
20	2.4	1.8
50	3.86	3.2
60	4.17	3.92
100	7.33	4.7

**Table 4** Delay in Bandcon topology

No. of flows	Single mapping (s)	Multi mapping (s)
10	1.33	1.02
20	2.9	1.13
50	4.59	2.86
60	7.71	7.71
100	9.18	9.18



**Fig. 2** Delay in SWITCH topology



**Fig. 3** Delay in Bandcon topology

### 5.2 Throughput

The computed throughput of the SWITCH network topology with single mapping and multi mapping scenarios is shown in Table 5.

The computed throughput of the Bandcon network topology with single mapping and multi mapping scenarios is shown in Table 6.

**Table 5** Throughput in SWITCH topology

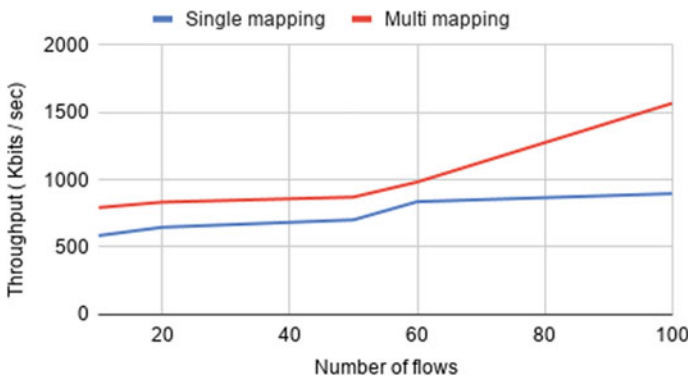
No. of flows	Single mapping (kbits/s)	Multi mapping (kbits/s)
10	585	793
20	647	833.6
50	702	871.2
60	837	983
100	897	1568

**Table 6** Throughput in Bandcon topology

No. of flows	Single mapping (kbits/s)	Multi mapping (kbits/s)
10	459	750
20	516	819
50	582	829
60	683	874
100	686	927

From the above graphs, it is clear that deploying multiple mapping in any topology will provide a higher data transmission rate than that of using a single mapping (Figs. 4 and 5).

For switch topology, multi mapping has increased throughput by 17% in comparison with single mapping. For Bandcon topology, multi mapping has increased throughput by 27% in comparison with single mapping.



**Fig. 4** Throughput in SWITCH topology



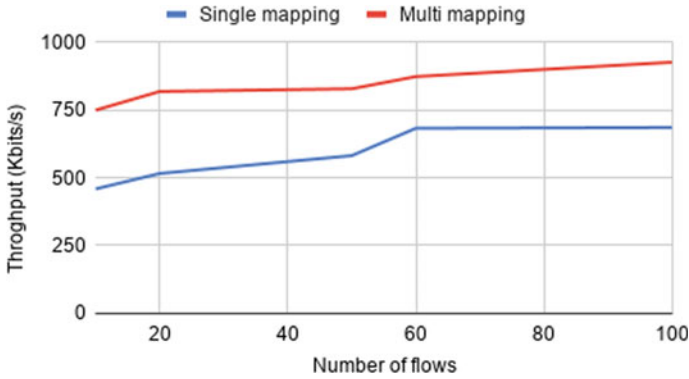


Fig. 5 Throughput in Bandcon topology

### 5.3 Packet Loss

The packet loss in the SWITCH network topology in single mapping and multi mapping scenarios is shown (Table 7).

The packet loss in the Bandcon network topology in single mapping and multi mapping scenarios is shown (Table 8).

Packet loss also increases with the number of flows. In both network topologies, packet loss in the multi mapping scenario is lesser than that of a single mapping scenario (Figs. 6 and 7).

Table 7 Packet loss in SWITCH topology

No. of flows	Single mapping (%)	Multi mapping (%)
10	4.9	1.5
20	4.45	2.75
50	15.32	11.94
60	18.71	12.53
100	19.02	12.99

Table 8 Packet loss in Bandcon topology

No. of flows	Single mapping (%)	Multi mapping (%)
10	5.19	1.61
20	5.4	2.95
50	19	10.1
60	25.3	16.02
100	28.02	23.11

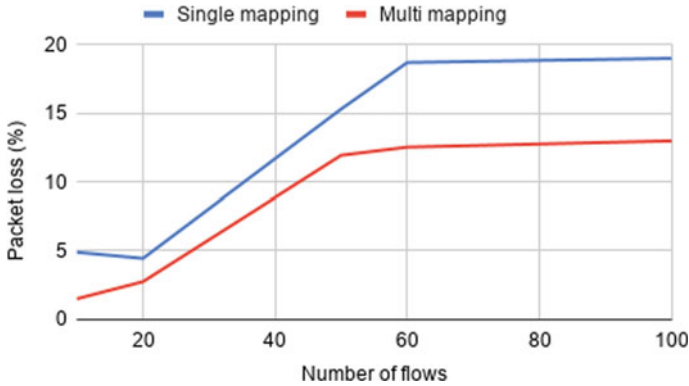


Fig. 6 Packet loss in SWITCH topology

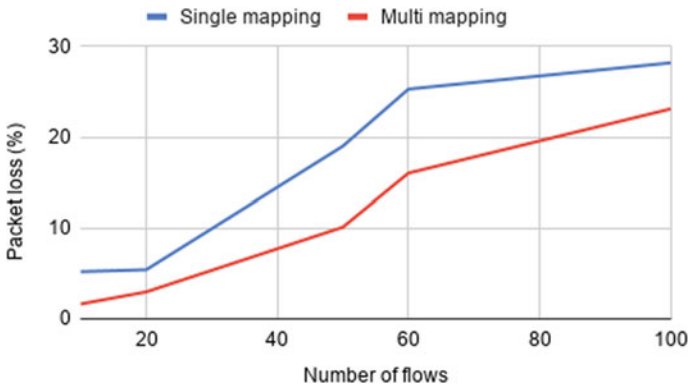


Fig. 7 Packet loss in Bandcon topology

For switch topology, multi mapping has reduced packet loss by 22% in comparison with single mapping. For Bandcon topology, multi mapping has reduced packet loss by 18% in comparison with single mapping.

From the simulation results, it is observed that the value of QoS parameters such as end-to-end delay, throughput, and packet loss is improved for multi mapping in comparison with the single mapping.

## 6 Conclusion

The proposed algorithm of the multi mapping technique to improve the QoS in SD-WAN has been developed. The simulation for both single mapping and multi mapping is done with the help of the tool mininet. SWITCH and Bandcon topology

is taken from Internet topology zoo. Topologies are simulated using MiniEdit. The performance metrics used are end-to-end delay, throughput, and packet loss. The QoS parameters for both single mapping and multi mapping are compared using SWITCH topology and Bandcon topology. For switch topology, multi mapping has reduced packet loss by 38%, reduced end-to-end delay by 21%, and increased throughput by 36% in comparison with single mapping. For Bandcon topology, multi mapping has reduced packet loss by 43%, reduced end-to-end delay by 36%, and increased throughput by 45% in comparison with single mapping.

## References

1. Sridharan V, Mohan PM, Gurusamy M (2019) QoC-aware control traffic engineering in software defined networks. *IEEE Trans Netw Serv Manag* 17(1)
2. Sminesh CN, Kanaga EGM, Roy A (2019) Optimal multi-controller placement strategy in SD-WAN using modified density peak clustering. *IET Commun* 13(20):3509–3518
3. Akyildiz IF, Lee A, Wang P, Luo M, Chou W (2014) A roadmap for traffic engineering in SDN-OpenFlow networks. *Comput Netw* 71:1–30
4. Kreutz D, Ramos FM, Verissimo PE, Rothenberg CE, Azodolmolky S, Uhlig S (2014) Software-defined networking: a comprehensive survey. *Proc IEEE* 103(1):14–76
5. Zhang Y, Cui L, Wang W, Zhang Y (2018) A survey on software defined networking with multiple controllers. *J Netw Comput Appl* 103:101–118
6. Wang G, Zhao Y, Huang J, Wang W (2017) The controller placement problem in software defined networking: a survey. *IEEE Netw* 31(5):21–27
7. Sreejish AG, Sminesh CN (2018) An optimal controller placement strategy using exemplar-based clustering approach for software defined networking. In: *Emerging trends in engineering, science and technology for society, energy and environment*. CRC Press, pp 925–930
8. Sridharan V, Gurusamy M, Truong-Huu T (2017) Multi-controller traffic engineering in software defined networks. In: *2017 IEEE 42nd conference on local computer networks (LCN)*. IEEE, pp 137–145
9. Wang CA, Hu B, Chen S, Li D, Liu B (2017) A switch migration-based decision-making scheme for balancing load in SDN. *IEEE Access* 5:4537–4544
10. Cui J, Lu Q, Zhong H, Tian M, Liu L (2018) A load-balancing mechanism for distributed SDN control plane using response time. *IEEE Trans Netw Serv Manage* 15(4):1197–1206
11. Xu Y, Cello M, Wang IC, Walid A, Wilfong G, Wen CHP, Marchese M, Chao HJ (2019) Dynamic switch migration in distributed software-defined networks to achieve controller load balance. *IEEE J Sel Areas Commun* 37(3):515–529
12. Isong B, Mathebula I, Dladlu N (2018) SDN-SDWSN controller fault tolerance framework for small to medium sized networks. In *2018 19th IEEE/ACIS International conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD)*. IEEE, pp 43–51
13. Zhang L, Wang Y, Zhong X, Li W, Guo S (2018) Resource-saving replication for controllers in multi controller SDN against network failures. In: *NOMS 2018—2018 IEEE/IFIP network operations and management symposium*. IEEE, pp 1–7
14. Aly WHF, Al-anazi AMA (2018) Enhanced controller fault tolerant (ECFT) model for software defined networking. In: *2018 fifth international conference on software defined systems (SDS)*. IEEE, pp 217–222
15. Shalimov A, Zuikov D, Zimarina D, Pashkov V, Smeliansky R (2013) Advanced study of SDN/OpenFlow controllers. In: *Proceedings of the 9th central & Eastern European software engineering conference in Russia*, pp 1–6

# Auto-Completion of Queries



Vidya S. Dandagi and Nandini Sidnal

**Abstract** Search engines are exceedingly dependent on the query auto-completion. Query auto-completion is an ongoing activity that puts forwards a group of words for every click dynamically. Query suggestions help in formulating the query and improving the quality of the search. Graphs are data structures that are universal and extensively used in computer science and related fields. The graph machine learning approach is growing rapidly with applications such as friendship recommendation, social network, and information retrieval. Node2vec algorithm is used to study the feature illustration of nodes in a graph. It is derived by word embedding algorithm Word2vec. A supervised Recurrent Neural Network using Long short-term memory (LSTM) is employed to compute the accuracy. This model confirms 89% accuracy for query auto-completion. Greater the accuracy better the model.

**Keywords** Query Auto-Completion (QAC) · Node2vec · Long short-term memory (LSTM) · Knowledge graph (KG)

## 1 Introduction

The present search engine bestows a large amount of data for a particular query, but there are several issues with the present search system such as poor precision and recall. Semantic search accuracy is improved by considering the user's intent and the meaning of the terms depending on the context [1]. Query auto-completion (QAC)

---

V. S. Dandagi (✉)

Department of Master of Computer Application, KLE's Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, Karnataka, India  
e-mail: [vsdandagi@gmail.com](mailto:vsdandagi@gmail.com)

N. Sidnal

Department of Computer Science and Engineering, KLE's Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, Karnataka, India  
e-mail: [sidnal.nandini@gmail.com](mailto:sidnal.nandini@gmail.com)

is the major feature in the present search engine [2]. The important goal of a QAC system is to decrease the user’s effort and to overcome the difficulties like writing extended texts, typo errors, and identifying the most appropriate query terms [3]. The semantic web (SW) is the expansion of the present World Wide Web (WWW). It allows the search on the web to be carried out based on the meaning. Semantic web uses W3C standards. Semantic web is an intelligent web that allows users to use the shared meaningful knowledge representation on the web. Artificial intelligence applications are created using the relationship between individuals and classes of objects representing the domain knowledge with the help of ontology [4].

Ontology can give good acknowledgment of concept, and the knowledge of understanding the relationship between entities. Ontology can be used for indexing, ranking process, and also query reformulation and expansion purposes. Ontologies are used to describe links between different types of semantic knowledge and are used in devising data searching strategies [6]. The major advantage of using domain ontology is its capability to define a semantic model of the data combined with the linked domain knowledge [7]. The application of ontology in information retrieval has given better results. Resource description framework (RDF) and SPARQL are the frequently used terminologies with the semantic web. RDF is the regular way of the representation of data on the SW. Triple stores are the storage for RDF data that constitutes the triples. The triples are subject, predicate, and object. Predicate means the link between subject and object. Nodes in the graph are the resources, the properties depict the labels of the nodes, statements between two resources are the edges in the graph. To query the RDF data, the language used is SPARQL [8].

Figure 1 depicts the structure of the RDF statement that constitutes of subject-predicate and object. The knowledge-based system consists of a knowledge base and a reasoning engine, which constitute the knowledge graph (KG). Figure 2 shows the architecture of a knowledge graph. The KG combines and understands information, generates an ontology, and uses a reasoner to obtain a novel knowledge.

Knowledge graph represents the facts, entities, relationships, and semantic descriptions. Entities are objects linked to real-world and abstract ideas. Relationship deduces the relationship among entities. Semantic description encloses the types

Fig. 1 An RDF statement [8]

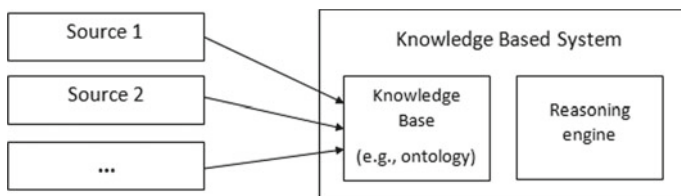
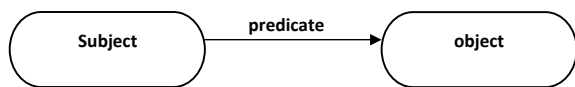


Fig. 2 Architecture of knowledge graph [9]

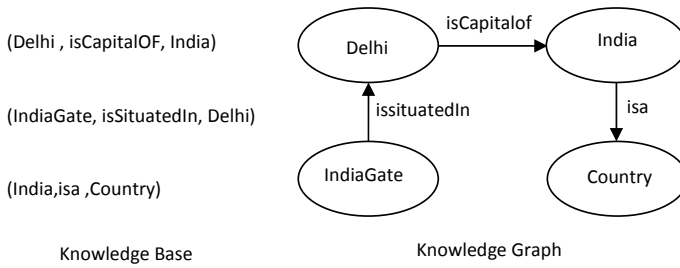


Fig. 3 Example of knowledge base and knowledge graph

and properties of entities with a definite meaning. The knowledge graph visualizes entities as nodes and relationships as edges. Knowledge graph that creeps the whole web is called as a self-sufficient semantic web [9].

Figure 3 shows the example of a knowledge base with three facts, represented by a knowledge graph.

Graphs play an important role in modern machine learning. Embedding nodes in graphs is a way to convert nodes, edges, and their features into vector space. Node2Vec builds upon the successful foundation of Word2Vec. Node2Vec converts the graph into a sequence of words using random walk.

In this proposed model, the triples generated are more semantically represented in the KG. The triples are created using the domain ontology and inferred using a reasoner. Terms in the query are depicted as, subject and predicate and object. (s, p, o), when the subject and predicate is known the object is to be auto-completed. LSTM model is used to learn long-term dependencies of the terms in the queries.

In this paper, a brief review of related work is put forth in Sect. 2, the proposed architecture is discussed in Sect. 3, and in Sect. 4 discussions and results are presented, and Sect. 5 conclusion is presented.

## 2 Literature Review

The main aim of the knowledge graph is to place the entities and relations into a semantic space skip-gram model, and the soft logic reasoning method is used to study the semantic relatedness. TransE and TransH and TransR are simple embedding techniques. TransE and TransR map entities and relations in the same semantic space.

The main aim of knowledge graph with concepts is to place the concepts of entities and relations of a KG in semantically low-dimensional space. Skip-gram model is used to study the semantic relatedness of concepts and entities that helps in embedding. Entity embeddings when combined with concept information give better results to entity classification than textual information [10].

To predict the triples in a KG, a multi-layer sequential Recurrent Neural Network is proposed. This model is capable of predicting the whole triples given only one

triple. The model takes the triples that are incomplete as input and predicts their tail entities. A beam search method is implemented with a large window size for prediction [11].

To improve the accuracy of link prediction, a new way named entity link prediction for knowledge graph (ELPKG) is projected. Entity link prediction algorithm is used to recognize the relationship between entities. The method implemented is known as the probabilistic soft logic-based reasoning method. This process adds paths and semantic-based features to understand the relationship among entities. To calculate the similarity of tuples, the euclidean distance is calculated. The two tuples are similar if the distance between them is the shortest. This ELPKG addresses the missing relations in the Knowledge Graph [12].

Multi-label deep neural network model highlights on relation prediction. The purpose of this model is to predict the relationship among the entities in a knowledge graph. Link prediction among entities is significant for creating huge ontologies and for knowledge graph completion. If the relation is predicted accurately, it can be augmented to a given ontology [13].

TransE, TransH, and TransR are easy and productive embedding techniques. The entity and relation embeddings are created by taking into consideration the translation from head and tail entity. In TransE model, relationships occur by explaining them on the low-dimensional embeddings of entities. A cascading embedding framework is introduced in which the extraction of semantic, graph features are considered. These features are extracted from knowledge embedding and graph embedding. They are the inputs for the cascade learning model. This helps in handling the challenge of unbalanced relations [14].

Knowledge graphs are formed by semantic web language known as resource description framework. A new method for embedding knowledge graph is through real-valued tensors. These tensor-based embeddings recover relationship from the knowledge graph. When the average degree of an entity in a graph is high, the tensor decomposition model performs well. This achieves more improvement in predicting new facts across the knowledge graph [15].

In classification using link prediction (CULP) model, the link predictor helps in the classification process and captures the classes, labeled and unlabeled data. The CULP algorithm uses a novel formation called label embedded graph and link predictor to search the class of unlabeled data. A link predictor known as compatibility score is used as predictors for CULP. Classification using link prediction utilizes a graph called labeled edge graph that enables the use of link predictor [16].

A novel method that combines the convolutional neural network and the bidirectional long short-term memory is put forth. Combination of accurate triples and corrupted triples are grouped to form the training information. A path ranking algorithm is adopted to get the relation paths for each training instances that are relevant to the relation  $r$ . Random walks are performed on the whole graph to know which relation links the source entity with its target. To understand the semantic co-relation between two entities and the path between them is done by creating a vector representation. The attention process takes into consideration the track between the entities. This model produces multistep reasoning over path representation in an embedding

area. For large graphs extracting the features from the paths is carried out by the path encoder [17].

A simple bilinear model for knowledge graph completion is proposed. Knowing the present links between the entities, link prediction helps in predicting new links for a knowledge graph. The tensor factorization method is useful for link prediction problems. The first tensor factorization method is Canonical Polyadic, it computes two embedding vectors for each entity. Simple is the improvement of Canonical Polyadic in which the two entities are learned dependently, one embedding vector for each relation and two vectors for each entity. One vector is used when the entity is head, and the other is used when the entity is the tail [18].

Knowledge graph completion includes time data as one of the features in the embedding approach. Long short-term memory (LSTM) model is used to learn the knowledge graph inferred facts in which time is augmented [19].

Knowledge graphs are generated by adding and deleting the triples which are not static. Knowledge graph created dynamically fail to consider the embeddings. Knowledge embedding and contextual element embedding are two ways of representation for dynamic knowledge graph embedding. Contexts of entities and relationships are encoded based on the attentive graph convolutional network model, which selects the most important information from the context for a given entity and relation. This model is stronger and has a possibility of scaling for online learning of embeddings [20].

In this proposed work [11], stacking multiple RNN cells improves the performance. The control of the layer number for predicting the entity does not give better performance. In [22], ConvE is applied to the triples on a multi-layer convolutional network that improves the performance [17]. Path encoder is successful in taking out features from paths in huge graphs. Large KGs have entities of multiple types, whereas in this model a single type of entity representation is used. Semantic features between the entities are not considered. Simple studies the two embeddings of each entity dependently [18]. Simple performs compared to tensor factorization techniques. By changing the weights knowledge can be included in the embeddings. The embeddings studied through simple are interpretable. As the size of the embeddings increases the complexity of simple increases [19]. In temporal KG, the relations between the entities is held only for a specific time interval. This proposed method is strong for some real-world challenges of KG that is sparsity and heterogeneity of temporal expressions.

An ontology is created for a tourism domain. Since they contain concepts and relations. Entities are represented as nodes and relations as edges in a knowledge graph. Given a query, the model can auto-complete, by taking into consideration the past search history of the user. The goal of the work is to learn the long-term dependencies using long-short term memory (LSTM).



### 3 Approach

Node2vec is a framework that consists of an algorithm to learn the feature of nodes in the network. It uses two parameters  $p$  and  $q$  that bias the random walk. Node embeddings are derived by word embedding algorithm Word2vec. Learning embeddings is done by analyzing the relationship between the entities. An embedding is a relatively low-dimensional space into which you can translate high dimensional vectors [5]. Node2vec uses a sampling strategy that takes in four arguments:

- a. Number of walks: From each node in the graph, the number of random walks is generated.
- b. Walk Length: Number of nodes in each random walks,
- c.  $p$ : return hyperparameter, and
- d.  $q$ : InOut hyperparameter.

For random walk generation, the algorithm will transit to each node in the graph and generate random walk and walk length.

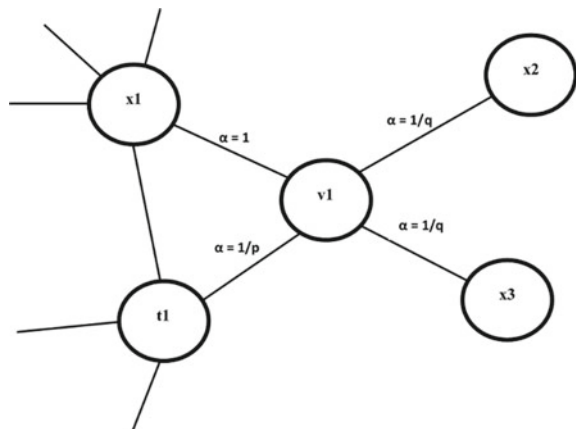
Figure 4 depicts the random walk procedure in Node2vec. The probability of movement from  $\langle v1 \rangle$  to any one of its neighbors is  $\langle \text{edgeweight} \rangle * \langle \alpha \rangle$  (normalized), where  $\langle \alpha \rangle$  depends on the hyper parameters.  $p$  reins the probability to return to  $\langle t1 \rangle$  through the  $\langle v1 \rangle$ .  $q$  reins the probability to move and investigate undiscovered parts of the graph.

#### Recurrent Neural Network:

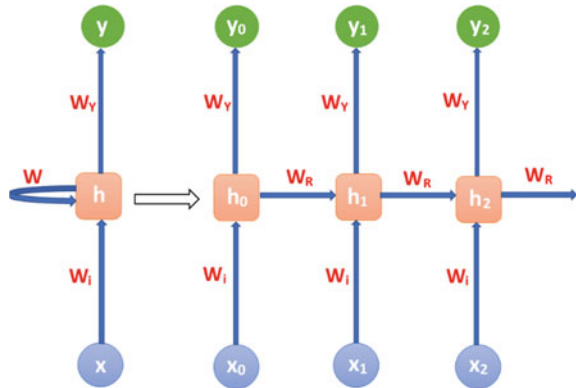
They are the class of neural networks. To process the sequence of inputs, RNNs use internal hidden states. In RNN, the data from the previous texts are represented as a low-dimensional vector. Figure 5 depicts the architecture of the RNN.

A portion of the neural network,  $h$  comes across at the input  $x$  and outputs the value  $y$ . A loop allows information to be passed from one step of the network to the next. The chain-like nature shows that RNNs are related to sequences of lists. RNNs are unable to connect the information.

Fig. 4 Random walks procedure in Node2vec [5]

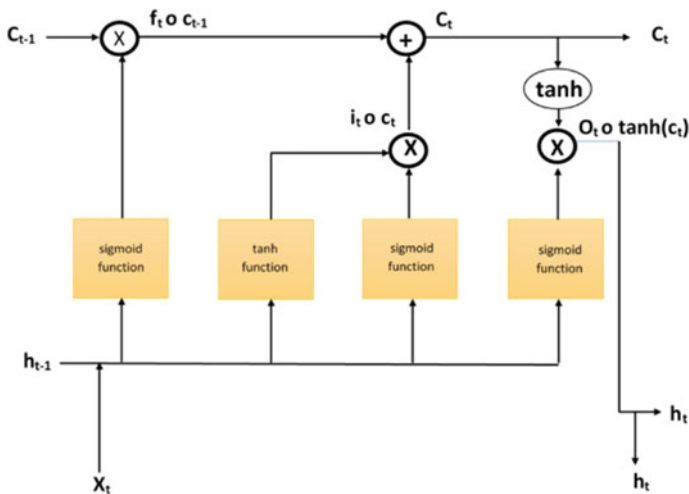


**Fig. 5** Architecture of Recurrent Neural network [11]

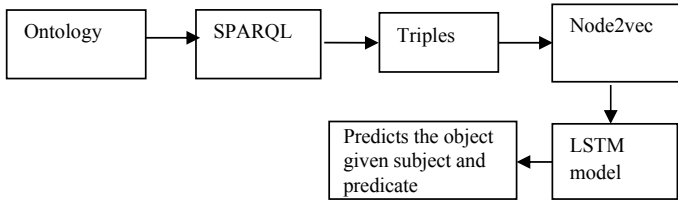


Long short-term memory (LSTM) is a particular kind of RNN that can learn long term dependencies. The proposed work uses LSTM with sequential data (Fig. 6).

The key of LSTMs is the cell state that is the horizontal line on the top of the diagram. Cell state is the key element, behaves like a conveyor belt of LSTM that is represented as a horizontal line. Adding and removing data to the cell state is controlled by the gates. There are three gates, input gate, forget gate, and input gate. The forget gate controls the data in the cell state. This decision either 0 or 1 is made by the sigmoid layer. 1 means to hold while a 0 represents removal. To update the cell state, there is the input gate. The sigmoid output decides which information is to be retained by the tanh output. The output gate controls the hidden state, it contains information about the previous inputs. The hidden state is used for predictions.



**Fig. 6** LSTM cell [21]



**Fig. 7** Proposed diagram

**Table 1** Dataset format

Session id	Subject	Predicate	Object
123	Delhi	Is capital of	India
123	India	Is a	Country
124	India Gate	Is situated	Delhi

### 3.1 Proposed Architecture

In Fig. 7 shows the flow diagram. Ontology is created with the help of protege. It consists of classes that represent the concepts in a domain, a taxonomy, property values, relations between classes restrictions on properties, slots, and individuals. Using the query language triples (s.p.o) are created. This is converted into graph, and embedding is performed using Node2vec. LSTM model is used for prediction.

### 3.2 Dataset

This dataset is divided as 80% training and the remaining as test dataset. The dataset format is as follows in Table 1

### 3.3 Evaluation Metrics

The evaluation metrics used are the hits@N, mean reciprocal rank(MRR).

hits@N is defined as the proportion of correct entities in the top  $k$  ranks.

$$\text{hits@}N = \frac{1}{|T1|} \sum_{(s,r,o) \in T1} \text{indicator}(\text{rank}(s, r, o) \leq N)$$

where T1 is the test set, |T1| is the number of triples, and indicator is the function defined as  $\text{indicator}(x1) = \begin{cases} x1 = 1 & \text{True} \\ x1 = 0 & \text{False} \end{cases}$

Reciprocal rank is a statistical method for assessing any process that gives a record of outcomes based on the probability of correctness, and it is the multiplicative inverse of the rank of the first correct answer. MRR is the mean of the reciprocal rank of outcomes for the same queries  $Q$ .  $MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$  rank<sub>i</sub> refers to the rank position of the relevant query from the  $i$ th query. [12].

### 4 Results and Discussions

Queries relating to the tourism domain are represented as a graph and embedding using Node2vec. There are 5 layers of embedding, dropout, LSTM, dense layer for activation by Relu and sigmoid. LSTM model is to predict the next word given the subject and predicate. One hot code is used to represent the subject and predicate. The model is trained for 250 epochs, with the categorical cross-entropy as the loss function. The accuracy of the model is 89.4%, it infers that the relevant terms given by the random walk help in auto-completing. MRR score for prediction of the object is 81.9.

Figure 8 shows the graph of epoch versus accuracy. As the number of epochs increases, the accuracy of the model improves.

Figure 9 shows that increase in the epoch number decreases the loss. The loss function gives an insight into the training process and also the direction in which the network learns. The above figure gives a good learning rate.

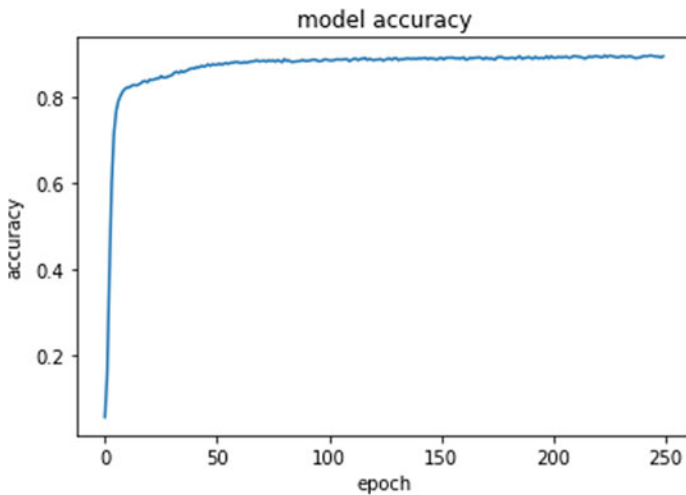
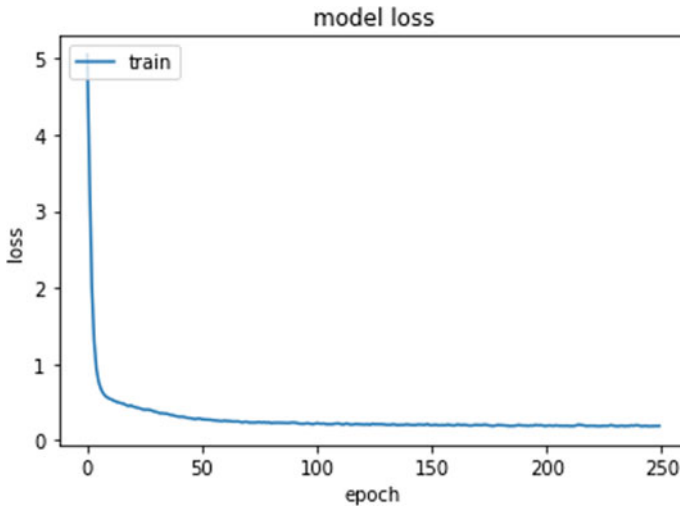


Fig. 8 Model accuracy



**Fig. 9** Epoch versus loss

## 5 Conclusion

In this paper, a Recurrent Neural network long short-term memory is presented for auto-completing the query. Ontology for a tourism domain is represented as subject, predicate, and object. Entities are represented as nodes and relations as edges. Node2vec is a framework that studies the features of the nodes. Node2vec algorithm generates node embeddings. Performing a random walk over the entire graph generates a path. Higher the accuracy more relevant are the predictions.

## References

1. Chauhan R, Goudar R, Sharma R, Chauhan A (2013) Domain ontology based semantic search for efficient information retrieval through automatic query expansion. In: International conference on intelligent systems and signal processing (ISSP). IEEE, Gujarat, pp 397–402
2. Jadhav PA, Chatur PN, Wagh KP (2016) Integrating performance of web search engine with Machine Learning approach. In: 2nd international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE, Chennai, pp 519–524
3. Wang Y, Ouyang H, Deng H, Chang Y (2017) Learning online trends for interactive query auto-completion. *IEEE Trans Knowl Data Eng* 29(11):2442–2454
4. Khamparia A, Pandey B, Pardesi V (2014) Performance analysis on agriculture ontology using SPARQL query system. In: International conference on data mining and intelligent computing (ICDMIC). IEEE, New Delhi, pp 1–5
5. Grover A, Leskovec J (2016) Node2vec: scalable feature learning for networks. In: Proceedings of the 22nd international conference on knowledge discovery and data mining. ACM SIGKDD, New York, pp 855–864

6. Azizan A, Bakar ZA, Noah SA (2014) Analysis of retrieval result on ontology-based query reformulation. In: International conference on computer, communications, and control technology (I4CT). IEEE, Langkawi, pp 244–248
7. Munir K, Sheraz Anjum M (2018) The use of ontologies for effective knowledge modelling and information retrieval. *Appl Comput Inf* 14(2):116–126
8. Chawla T, Singh GE, Pilli S, Govil MC (2016) Research issues in RDF management systems. In: International conference on emerging trends in communication technologies. IEEE, Dehradun, pp 1–5
9. Ehrlinger L, WöB W (2016) Towards a definition of knowledge graphs. In: *Semantics (Posters, Demos, Success)*. Leipzig, vol 48, pp 1–4
10. Guan N, Song D, Liao L (2019) Knowledge graph embedding with concepts. In: *Knowledge-based systems*. Semantic, vol 164, pp 38–44 (2019)
11. Guo L et al (2018) DSKG: a deep sequential model for knowledge graph completion. In: *China conference on knowledge graph and semantic computing*. Springer, Singapore (2018)
12. Ma JT et al (2019) ELPKG: a high-accuracy link prediction approach for knowledge graph completion. *Symmetry* 11(9):1096
13. Onuki Y et al (2019) Relation prediction in knowledge graph by multi-label deep neural network. *Appl Netw Sci* 4(1):20
14. Li D, Madden A (2019) Cascade embedding model for knowledge graph inference and retrieval. *Inf Process Manage* 56(6):102093
15. Padia A et al (2019) Knowledge graph fact prediction via knowledge-enriched tensor factorization. *J Web Seman* 59:100497
16. Fadaee SA, Haeri MA (2019) Classification using link prediction. *Neurocomputing* 359:395–407
17. Jagvaral B et al (2020) Path-based reasoning approach for knowledge graph completion using CNN-BiLSTM with attention mechanism. *Expert Syst Appl* 142:112960
18. Kazemi SM, Poole D (2018) Simple embedding for link prediction in knowledge graphs. In: *Advances in neural information processing systems*, Canada
19. García-Durán A, Dumančić S, Niepert M (2018) Learning sequence encoders for temporal knowledge graph completion. In: *Proceedings of conference on empirical methods in natural language processing*, association for computational linguistics, pp 4816–4821. Brussels, Belgium
20. Wu T et al (2019) Efficiently embedding dynamic knowledge graphs. <http://arXiv.org/1910.06708v1> [cs.DB]
21. [https://en.wikipedia.org/wiki/Long\\_short-term\\_memory](https://en.wikipedia.org/wiki/Long_short-term_memory)
22. Dettmers T et al (2018) Convolutional 2D knowledge graph embeddings. In: *Thirty-Second AAAI conference on artificial intelligence*



**Vidya** has been an educator for over 10 years. She is working as Assistant Professor in Master of Computer Application. She is a research scholar, working on Semantic web, Query completion and Data Analytics. She has guided many post graduate projects that include web development applications and android application.

She has publications in International journal of Web Semantic Technology, IEEE and Test Engineering and Management journal.



**Nandini** has been a researcher and educator for over twenty five years. Nandini has held a number of academic positions (Academic Manager, Head of the Department) at universities and education institutions both in India and Australia. She has held administrative positions at VTU, Belgaum Campus, KLESCET and now at MIT, Sydney campus.

Her research interests lie in the area of Cognitive Computing using intelligent agents, Data Analytics and Semantic web. She has more than 30 research publications in peer reviewed journals and has presented some of them in International conferences. She has published couple of books on Cognitive Computing for Bidding in E-Auctions, Scholar's Press and a C Programming Laboratory Hand Book for Beginners by Wiley publications. She has received best paper awards in International conferences.

Nandini has developed and practiced innovative teaching practices such as active and experiential learning. She had developed Intelligent Multi-agent framework for Adaptive e-learning as part of master thesis work. Further, she was a lead member to implemented virtual learning and assessment for specific master courses of VTU in 2005. She is now an active member of a sponsored project at MIT on Game based learning and is practicing feedforward learning for industrial based project units. She is also developing a portfolio platform for research based projects. Currently working on applying block chain in supply chain management.

# Design of CMOS Active Inductors for RFIC Applications: A Review



Zishani Mishra, T. Prashanth, N. Sanjay, Jagrati Gupta, and Amit Jain

**Abstract** This paper represents the relative differences between the previously proposed active inductor designs. The parameters such as quality factor, power consumptions, and inductance tuning have been considered while comparing the performances of these inductors, operating at radio frequency. Inductors are widely used in integrated circuits for their tunability property. RF applications of inductors include the use in filters, low-noise amplifiers, and VCOs. Over the years different designs of active inductors have been proposed to eliminate the problems associated with conventional on-chip inductors. Different designs that propose techniques to reduce distortions, to improve the quality factor and tunability have been considered in this paper.

**Keywords** Active inductor (AI) · RFIC · Distortion · Quality factor · Gyrator-C · Inductance

## 1 Introduction

High functioning RF integrated circuits have been in the upfront due to its rising need in wireless communication applications [1]. A lot of research is being carried out on designing efficient RF circuits such as low-noise amplifiers [2], voltage-controlled oscillators [3, 4], filters [5], variable phase shifter. Inductors are widely used in RF filtering, matching networks, voltage control oscillators, RF chokes, etc., [6–10]. The traditional passive inductor limits the operative bandwidth which increases the size and is used as discrete off-chip components [7]. On-chip inductors such as the bond wire inductors and spiral inductors served as passive components on the chip. These types of inductors were extensively used but came along with several challenges [11–13] such as mutual coupling, electric and magnetic penetration into substrate, and

---

Z. Mishra · T. Prashanth · N. Sanjay · J. Gupta · A. Jain (✉)  
Department of Electronics and Communication Engineering, CMR Institute of Technology,  
Bengaluru, India  
e-mail: [amit.j@cmrit.ac.in](mailto:amit.j@cmrit.ac.in)



various losses due to the skin and proximity effects. Moreover, only fixed inductance value as possible from these passive inductors [7].

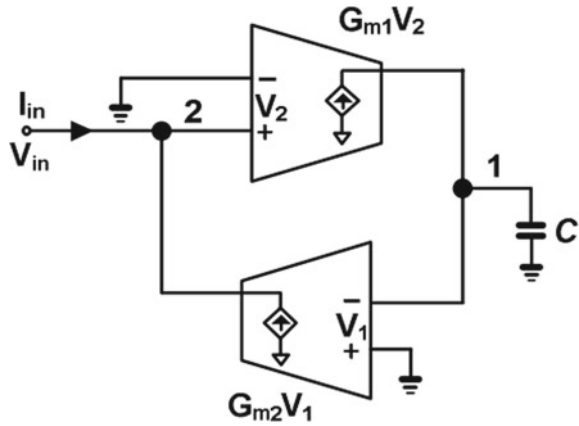
Researchers studied extensively on the possibilities of eliminating the challenges associated with passive inductors, and in the process, many designs of active inductors (AI) were proposed. The use of the gyrator-C-based CMOS AI is prominent in recent studies for its ability to tune the circuits over a larger frequency range along with its higher quality factor and reduced area requirements [6–10, 14]. Different types of gyrator-C topologies have been designed over the years such as lossless and lossy single-ended gyrator-C [6]. Different techniques for the enhancement of the quality factor, noise factor, inductance tunability, and linearity have been discussed in the literature [15–20].

The work proposed in [8] presents a CMOS AI operated at the RF—band and was based on double—feedback transconductor topology. The proposed design shows a very high Q-factor of greater than 1000 and was used to implement a bandpass filter to check its effectiveness. The linearity of the CMOS AI is one major aspect that most researchers need to work upon. The major distortions are caused due to the transconductance and output conductance of the MOSFET. One of the methods to improve the linearity is by reducing the gate length [9]. The drain current of the MOSFET contributes significantly to the non-linear performance as it compresses the transconductance by generating DC in excess. In [10], a feed-forward current source is designed to clone the current that is generated in the MOSFET in such a way that it acts as compensation current. By doing so, the undesired non-linear component can be eliminated, hence making the system linear. Many designs of active inductors used in filters and power dividers are studied [15]. Takahide and Toshihiro presented low-distortion designs of active inductors based on the compensation current mechanism [16]. The more compact design of active inductor is discussed in [9], where the AI designed was capable of operating at high radio frequencies with lowered supply voltages. Uyanik and Tarim [17] designed a wide tunability active inductor which is operated at low power and provides very high-quality factor. The differential active inductor uses the cascode-grounded technique to lower the noise and improve the tunability. Another significant work has been presented in [20], which shows an improved noise factor with a wide frequency band.

## 2 Background

Two transconductors with transconductances  $G_{m1}$  and  $G_{m2}$  connected in the feedback loop as shown in Fig. 1 which form the basic structure of the Gyrator-C topology to replicate the inductor effect. A capacitor denoted as "C" is attached to port 1 which is working like a load impedance. The capability of the gyrator to invert the current–voltage characteristic plays a major role in its usage as an inductor. The topology can convert the intrinsic (parasitic) capacitance of a transconductor to an inductance [14].

**Fig. 1** Lossless single-ended gyrator-C topology [14]



$$I_{in} + G_{m2}V_1 = 0 \tag{1}$$

where  $V_1 = -\frac{G_{m1}V_2}{sC}$

Using this value of  $V_1$  in (1) got

$$I_{in} = \frac{V_2}{sC / G_{m1}G_{m2}} \tag{2}$$

The admittance of the gyrator-C network at port 2 is calculated as

$$Y = \frac{I_{in}}{V_2} = \frac{1}{s(C / G_{m1}G_{m2})} \tag{3}$$

From Eq. (3) can be found out

$$\frac{C}{G_{m1}G_{m2}} = \frac{QV}{I^2} \tag{4}$$

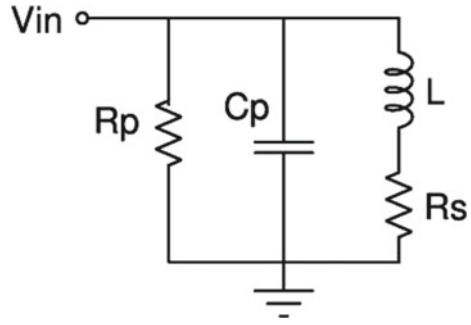
By putting  $Q = It$  in (4) got

$$\frac{C}{G_{m1}G_{m2}} = \frac{V}{I/t} = L \tag{5}$$

Hence, by this mechanism, an inductor is duplicated at port 2, which is a single-ended lossless inductor having an inductance of

$$L = \frac{C}{G_{m1}G_{m2}} \tag{6}$$

**Fig. 2** Equivalent circuit of the gyrator [20]



The input impedance of the gyrator topology shown in Fig. 1 is inversely proportional to its load impedance and is calculated as

$$Z_{in} = \frac{sC}{G_{m1}G_{m2}} \tag{7}$$

The MOSFETs used for the implementation of the transconductors are non-ideal, having parasitic components. Considering the non-ideal effects of MOSFET, an equivalent gyrator circuit has been shown in Fig. 2 [20]. The value of these parasitic components depends on the design of the active inductor. The inductive bandwidth of AI in terms of the parasitic components can be calculated as

$$\frac{R_S}{2\pi L} \leq f_{ind\_bandwidth} \leq \frac{1}{2\pi\sqrt{LC_P}} \tag{8}$$

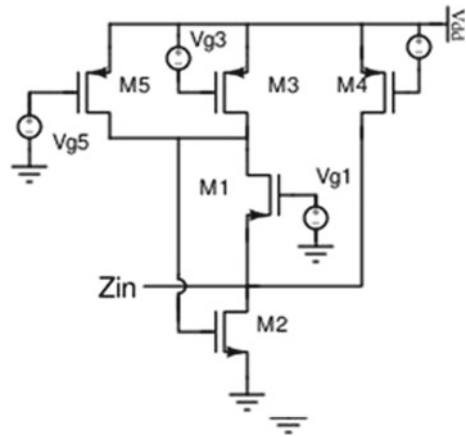
### 3 Design of Active Inductors

The most significant and important designs from the literature are elaborated in this section. The different design techniques of AI are discussed, and their performances are analyzed.

#### 3.1 Active Inductor with Distortion Reduction Circuit

Figure 3 shows the active inductor design that is implemented [16], using the distortion reduction method to eliminate the non-linear component. In this design, a current source  $M_5$  is connected in parallel with a bias current source. Since the source terminal of  $M_2$  and the gate terminal of  $M_5$  are connected to ground, the summation of  $V_{gs}$

**Fig. 3** Active inductor design with inherent property to reduce the distortion

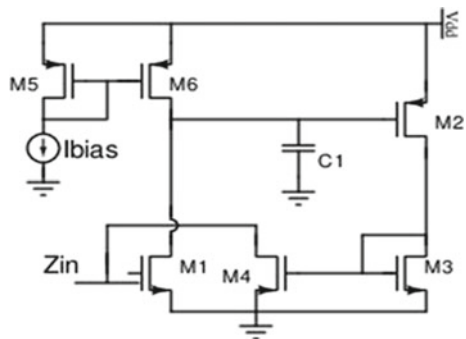


of  $M_5$  and  $M_2$  remains constant. Here, the drain current of  $M_5$  acts as the compensation current. The use of the  $M_5$  transistor is an effective way to improve linearity; however, it increases the series resistance of AI significantly.

### 3.2 A Low Voltage High-Q CMOS Active Inductor

More compact design is proposed by Uyanik and Tarim [17] as shown in Fig. 4. As observed in figure,  $M_1$  is in common-source configuration and realizes the negative transconductance. The  $M_2$ – $M_4$  combination provides the positive transconductance. The MOSFETs  $M_5$  and  $M_6$  are utilized for biasing purpose. Current mirror  $M_3$ – $M_4$  is used to invert the negative transconductance of  $M_2$ . The design uses less number of MOS transistors and hence reduces the size of the active inductor extensively. One advantage of this design is that none of the transistors suffers from body effect. Also, the noise is contributed only by two of the transistors which significantly reduces the

**Fig. 4** Low voltage high-Q CMOS AI



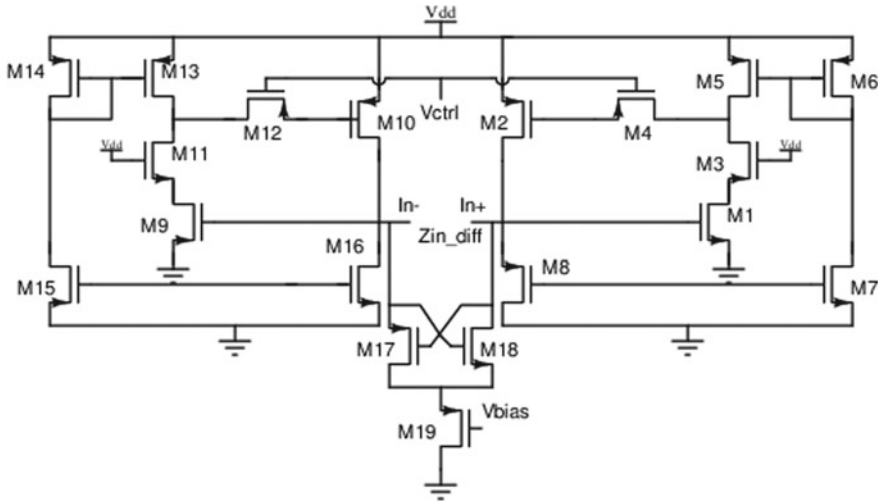


Fig. 5 Differential active inductor design

overall noise of the circuit [17]. This design was simulated with a supply voltage of 1.2 V using 0.13  $\mu\text{m}$  CMOS technology.

### 3.3 Differential Active Inductor Design

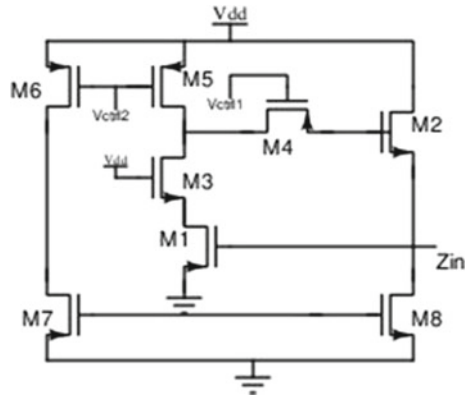
A differential AI has been shown in Fig. 5 [18]. By applying the differential input voltage between the gates of  $M_1$  and  $M_9$ , the  $G_{m1}$  and  $G_{m9}$  transconductances convert the voltage to a drain current charging the capacitances  $C_{gs2}$  and  $C_{gs10}$  of  $M_2$  and  $M_{10}$  transistors, respectively, [18]. The circuit is based on the common-grounded AI topology [7]. To obtain higher inductance and high-quality factor, a cascode circuit topology is implemented using  $M_3$  and  $M_1$  as noticed in figure. The equivalent resistance is decreased by reducing the conductance of transistor  $M_1$  which improves the inductance and the quality factor.

To further improve the performance, a resistance  $R_f$  is added in feedback which significantly increases the inductance of the designed AI [18]. The design was implemented using the 0.13  $\mu\text{m}$  CMOS technology.

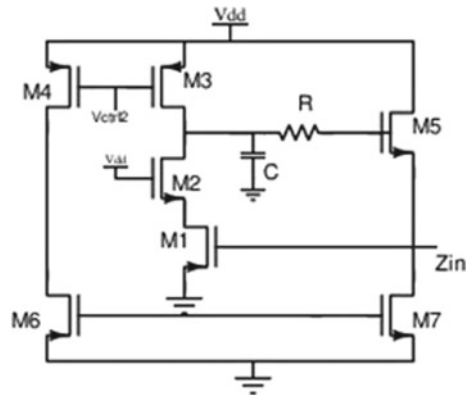
### 3.4 Adjustable Active Inductor

The circuit [19] consists of transistor  $M_1$ ,  $M_2$  realizing the negative and positive transconductances, and transistors  $M_5$ ,  $M_6$ ,  $M_7$ , and  $M_8$  are used to implement current mirrors as shown in Fig. 6. A cascode stage is constructed by transistor  $M_3$  which

**Fig. 6** Adjustable active inductor



**Fig. 7** A low power, high Q-factor grounded active inductor



enhances the inductance and the quality factor of the designed active inductor. Feedback resistance is also implemented with a PMOS transistor  $M_4$  to further improve the overall performance. Two tuning techniques are used to achieve a wide tuning range for the designed circuit. In first technique, a control voltage  $V_{ctrl1}$  is applied to the gate of transistor  $M_4$ , whereas in the second technique the transconductances of transistors  $M_1$ – $M_3$  are varied by varying the current source, implemented by transistor  $M_5$ – $M_8$ . The designed circuit is simulated using  $0.13 \mu\text{m}$  CMOS technology in cadence environment.

### 3.5 A Low Power, High $Q$ -Factor Grounded Active Inductor

The circuit [20] shown in Fig. 7 utilizes the common drain and common-source configuration of the gyrator-c approach. These two configurations are used to implement the negative and positive transconductances, respectively.  $M_1$  and  $M_5$  transistors form the basic topology of the circuit. It is observed that the cascode structure is adopted by integrating transistor  $M_2$  in a feedback loop. The feedback resistance improves the design further, as it creates an additional inductance.  $M_3$  and  $M_7$  in the proposed design act as current sources, and two voltage sources are used for biasing purpose. To check the functionality of the designed AI, simulations were performed in cadence environment.

## 4 Simulation Results and Discussion

The different active inductor designs discussed in the previous section were simulated in the cadence environment. The results were studied and analyzed thoroughly to draw a comparison for different performance parameters. The frequency tuning, inductance tuning range, quality factor, and noise factor were among the parameters considered for the comparison. The various active inductor designs showed improved performances in terms of certain parameters and could be accordingly used based on the requirement and applications. Table 1 shows the comparisons studied through different active inductor designs elaborated in the previous sections. It is observed that a maximum inductance tuning was obtained in [18] but at the cost of higher power consumption. The work presented in [17] shows a very high-quality factor of 3900 at the operating frequency of 5.75 GHz but it has lower inductance tuning. The AI proposed by Bharath et al. [20] presents an optimized performance in terms of inductance tuning, maximum quality factor, and power consumption. The work proposed by Hammadi et al. [19] also shows the very good result for most of the parameters.

**Table 1** Performance comparison with the existing work in the literature

References	Technology (um)	Vdd (V)	Power (mW)	Inductance (nH)	$Q_{\max}$	Frequency (GHz)
[16]	0.18	–	–	19–21	14.1@1.12 GHz	0.07–1.11
[17]	0.13	1.2	1	38–144	3900@5.75 GHz	0.3–7.32
[18]	0.13	1	2.59	79.8–495	2600@1.47 GHz	0.784–3.19
[19]	0.13	1	0.87–1.32	10.97–44.44	2565@2 GHz	1.76–6.5
[20]	0.18	1.8	0.99	25.7–180	341@2.51 GHz	0.79–2.69

## 5 Conclusion

In this paper, different active inductors designs have been analyzed, and their performances are compared in terms of inductance value, maximum achievable quality factor, power consumptions, etc. Different design techniques to improve different parameters have been thoroughly investigated, and hence, based on the requirement and applications, the designs can be integrated as on-chip inductors in RFICs. Among the works discussed, the maximum quality factor is observed as 3900 [17] whereas the maximum bandwidth is obtained as 2.4 GHz from the work proposed by Hammadi et al. [18]. The work presented by Hammadi et al. also shows the maximum inductance tuning range, from 79.8 to 495. The most optimized performance is observed in [20]. Finally, this paper summarizes the simulation results.

**Acknowledgements** Authors thankfully acknowledges the Center of Excellence Integrated Circuit, Department of Electronics and Communication Engineering, CMR Institute of Technology, Bengaluru, for the support.

## References

1. Prameela B, Daniel AE (2019) A novel high Q active inductor design for RFIC applications. In: Third international conference on computing and network communications, procedia computer science, vol 171, pp 2626–2634
2. Reddy V, Sammeta P, Kumari R, Quadir NA, Jain A (2019) A 0.5 V LNA design for 2.4 GHz wirelessbody area network applications. In: 3rd international conference on electronics, communication and aerospace technology (ICECA). <https://doi.org/10.1109/iceca.2019.8821857>
3. Li J, Yu J, Li X, Hu C (2019) A Wide Tuning Fully-Swing Differential VCO with improved active inductors for SERDES application. In: 3rd international conference on electronic information technology and computer engineering (EITCE). <https://doi.org/10.1109/eitce47263.2019.9094845>
4. Wang H, Wang H (2019) A two-terminal active inductor with minimum apparent power for the auxiliary circuit. *IEEE Trans Power Electron* 34:1013–1016
5. Hammadi AB, Haddad F, Mhiri M, Saad S, Besbes K (2018) RF and microwave reconfigurable bandpass filter design using optimized active inductor circuit. *Int J RF Microwave Comp Aided Eng*. <https://doi.org/10.1002/mmce.21550>
6. Patel DP, Oza S (2011) CMOS active inductor: a technical review. *Int J Appl Eng Res* 13:9680–9685
7. CMOS Active Inductors (2008) CMOS active inductors and transformers. Springer, Boston, MA
8. Yodprasit, Ngarmnil J (2000) Q-enhancing technique for RF CMOS active inductor. In: IEEE international symposium on circuits and systems (ISCAS), Geneva, Switzerland, vol 5, pp 589–592. <https://doi.org/10.1109/iscas.2000.857503>
9. Kang S, Choi B, Kim B (2003) Linearity analysis of CMOS for RF application. *IEEE Trans Microw Theory Tech* 51:972–977
10. Ler C, A'ain AKB, Kordesch AV (2009) CMOS active inductor linearity improvement using feed-forward current source technique. *IEEE Trans Microw Theory Techniques* 57:1915–1924. <https://doi.org/10.1109/tmtt.2009.2025426>



11. Yecil A, Yuce E, Minaei S (2020) MOSFET-C-based grounded active inductors with electronically tunable properties. *Int J RF Microwave Comput Aided Eng*. <https://doi.org/10.1002/mmce.22274>
12. Noferesti F, Zarei H, Pour MJE, Bijari A (2019) A fully differential CMOS active inductor with high quality factor and high tunability. In: 27th Iranian conference on electrical engineering. <https://doi.org/10.1109/iraniancee.2019.8786615>
13. Zhang Z et al (2019) A novel active inductor with high self-resonance frequency high Q factor and independent adjustment of inductance. In: The 8th IEEE international symposium on next-generation electronics. <https://doi.org/10.1109/isne.2019.8896629>
14. Faruqe O, Saikat MMM, Bulbul MAK, Amin MT (2017) Comparative analysis and simulation of active inductors for RF applications in 90 nm CMOS. In: 3rd international conference on electrical information and communication technology (EICT), Khulna, pp 1–6. <https://doi.org/10.1109/eict.2017.8275233>
15. Bhattacharya R, Basu A, Koul SK (2015) A highly linear CMOS active inductor and its application in filters and power dividers. *IEEE Microw Wirel Compon Lett* 25:715–717. <https://doi.org/10.1109/lmwc.2015.2479718>
16. Sato T, Ito T (2013) Design of low distortion active inductor and its applications. *Analog Integr Circ Sig Process* 75:245–255
17. Uyanik HU, Tarim N (2007) Compact low voltage high-Q CMOS active inductor suitable for RF applications. *Analog Integr Circuits Signal Process* 51:191–194
18. Hammadi AB, Mhiri M, Haddad F et al (2016) An enhanced design of RF integrated differential active inductor. *BioNanoSci* 6(2016):185–192
19. Hammadi AB, Mhiri M, Haddad F, Saad S, Besbes K (2017) Study of wide adjustable active inductor circuits: design approaches and reconfiguration methods. *ICEMIS*. <https://doi.org/10.1109/ICEMIS.2017.8273063>
20. Bharath L, Anila D, Ajay CN, Shrivani B, Jain A: A wide-band, low-power grounded active inductor with high Q factor for RF applications. In: Bindhu V, Chen J, Tavares J (eds) *International conference on communication, computing and electronics systems*. Lecture notes in electrical engineering, vol 637. Springer, Singapore

# Optimized Web Service Composition Using Evolutionary Computation Techniques



S. Subbulakshmi, K. Ramar, Anvy Elsa Saji, and Geethu Chandran

**Abstract** In service computing, Quality of Service (QoS)-aware web service composition is considered as one of the influential traits. To embrace this, an optimal method for predicting QoS values of web service is implemented where credibility evaluation is computed by accumulating reputation and trustworthiness. An automatic approach for weight calculation is invoked to calculate the weight of QoS attributes; it improves WS QoS values. QoS value is optimized by using Genetic Algorithm. Services with high QoS values are taken as candidate services for service composition. Instead of just selecting services randomly for service composition, cuckoo-based algorithm is used to identify optimal web service combination. Cuckoo algorithm realizes promising combinations by replacing the best service in lieu of worst service and by calculating the fitness score of each composition. A comparative study proved that it can provide the best service to end-users, as cuckoo selects only service composition with high fitness score.

**Keywords** Web service · Genetic algorithm · Cuckoo optimization algorithm · Levy flight · Quality of service · Credibility · Web service composition

---

S. Subbulakshmi (✉) · A. E. Saji · G. Chandran  
Department of Computer Science and Applications, Amrita Vishwa Vidyapeetham, Amritapuri,  
India

e-mail: [subbulakshmis@am.amrita.edu](mailto:subbulakshmis@am.amrita.edu)

A. E. Saji

e-mail: [anvylsasaji@gmail.com](mailto:anvylsasaji@gmail.com)

G. Chandran

e-mail: [geethuchandran999@gmail.com](mailto:geethuchandran999@gmail.com)

K. Ramar

Muthayammal Engineering College, Namakkal, TamilNadu, India

e-mail: [dean.cse@mec.edu.in](mailto:dean.cse@mec.edu.in)

# 1 Introduction

Web services are independent, free-standing and loosely coupled solicitations available over the web. Nowadays, the demand for web services (WS) turns out to be more and more salient due to leveraging number of WS [1]. Services are defined as solicitation of facts accessible via standard web protocol. Web service composition is an elucidated agent performing orchestration or choreography of existing WS which supports reusing of services to develop new corporate WS.

Composition of WS fascinates the interest of researchers. While collaborating with services provided by the web, it encounters intricate problems identifying promising services most optimal to the defined problem. Service composition process can be carried out in two ways—static or dynamic [2]. Before establishing composition, an intellectual process model with a set of tasks and their flow of processing is authorized. It is initiated with a query phrase which is examined with existing contextual data to assess the set of all factual atomic WS required to perform business tasks.

In static WS composition, accumulation of services is done at design time, and it involves stakeholder's mediation. To achieve a specific requirement, WS is executed one by one where the composition is done manually. It is not flexible, time-consuming and a tough mission. The rapid growth of digital technologies and leveraging number of WS demands dynamic WS composition. It simplifies the solution process as it is performed at run time without user mediation. It is generated by choosing atomic services automatically, by invoking machine-independent agent module which realizes actual requirements and its flow of operations.

For choosing a superlative WS from among a varied set of related services, the Quality of Service (QoS) is needed to recognize [3]. QoS is elucidated for calculation of overall service performance. Availability, throughput, transmission time, assets of service are considered to measure QoS [4]. The actual objective of QoS is to pursue the most favourable composite service concerning QoS requirements. Credibility is considered as a major factor to assess promising quality service. Essentially two facets of QoS values are appraised, i.e., service provider's QoS and user's QoS. The familiarity of the user is esteemed as user trusts and tracks registers of service suppliers as the reputation of service.

Service quality reputation is redefined by assessing the dissimilarity between all QoS properties of service given by providers and set of realized QoS values. Trust value for each service is realized with the user's experience. Reputation, trust value and weightage of individual services are considered for calculating WS credibility. Process of calculating weight for service is done based on the aspect ratio of objective weights.

Optimization is considered a major part in selecting and orchestrating WS composition. Evolutionary algorithms proposed are genetic algorithm (GA) used to find optimized QoS value of services and cuckoo optimal (CO) algorithm used to find optimized WS composition. Genetic algorithm is an inquiry technique related to principles of natural selection, used to produce satisfactory clarifications for optimization problems [5]. It performs on a population created with individuals of arbitrary values.

QoS optimization is done over the credibility of WS, to identify the optimal set of the population comprising of best candidate WS, to be considered in the service composition process. For optimal service composition, a cuckoo optimization algorithm is implemented [6]. It is an efficient metaheuristic approach used to accomplish optimal solutions at the point of global optimization.

Following Sect. 2 explains background study of related works, Sect. 3 elaborates system design of optimised service composition using evolutionary algorithms, Sect. 4 illustrates implementation results with the input dataset at different stages, and Sect. 5 concludes with future enhancements.

## 2 Background

Web service composition as a methodology is used for merging and reprocessing available WS to generate new value-added complex service that fulfils the needs of users [7]. Effectiveness of composition depends on the selection of high-quality candidate services. Process of selecting top-notch WS from a massive set of WS becomes more challenging for users due to the leveraging of services. So, prediction of QoS values is considered as a crucial part in selecting quality WS [8].

Nowadays, numerous methodologies have been analyzed to appraise quality expectation. Personalized prediction of QoS is foremost in serving users to progress high superior SOA systems [9]. Collaborative filtering (CF) is a major expertise for personalized QoS prediction; it is used for predicting QoS which maintain a strategic distance from a tedious and costly assessment.

In [9], Zainab provides a bootstrapping framework which is used to examine QoS of freshly registered WS. Solicitations with high performance are estimated, and their QoS values are assigned to new services. Zhang [10] recommends a structure for reliable WS evaluation, created based on QoS closeness of accurate and broadcast values. To find reliable services, similarity measures are scrutinized which appraises the reputation of the service supplier. Marzieh [11] devised a structure to choose WS for each task by utilizing Simple Additive Weighting strategy. Selection is done based on weightage score allotted to quality aspects based on necessities of the user. WS with a high score is selected which realizes a result close to user preference. It considers user preferences and ignores other factors, so it is not an optimal solution.

In [12], Mustafa planned for scheming static and dynamic composition of services. Components of distributed computing are compared with WS composition to define execution ability. It adopts data circulation policies among services and proposes a structure model to handle composition and execution. Suresh Kumar [13] elaborated an investigation of different methods of QoS-based service selection (QSS) methodology [14]. It is examined and presented to estimate the effectiveness of those service selection procedures.

AlSedrani [15] represented composition as a comprehensive classification technique based on various segments by observation and categorize the solutions based

on their compliance with the user’s request. It constitutes four phases—composition planning, service discovery, service selection and execution. This work failed to guarantee the overall quality of service composition. Above methods strive to find the best quality services and perform service composition which satisfies requirements of the defined problem, but they all ignored the aspect of optimal service selection and composition. To overcome this issue, an optimized service composition methodology is implemented with evolutionary algorithms.

### 3 System Implementation

To obtain an optimized web service composition, it is essential to identify the quality of services. Finding QoS values of services is an incredible task to be addressed in this context. To have optimal solutions, two evolutionary algorithms are used: genetic algorithm and cuckoo search algorithm. Genetic algorithm and cuckoo algorithm are used to optimize QoS prediction and service composition, respectively. An implementation methodology of the work is elaborated in the following sections.

#### 3.1 QoS Prediction

QoS is the explanation or calculation of the global performance of service. First, credibility evaluation is adopted to find QoS values and weight calculation to determine the significance of different quality attributes of WS in the context of business solution. Improved QoS value is assessed by applying weights on QoS value which is calculated based on the credibility of service. Figure 1 depicts a design methodology of QoS prediction.

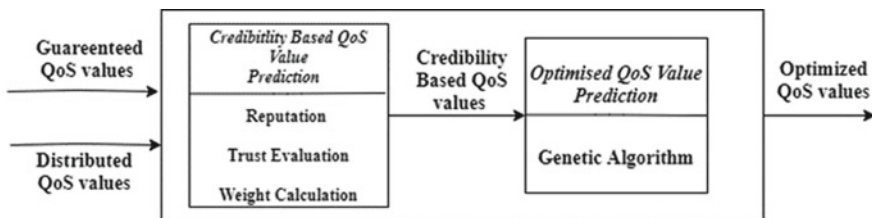


Fig. 1 Optimized QoS prediction

### 3.1.1 Credibility Evaluation

QoS of WS depends on various quality factors pertaining to service provided by the service providers. Service credibility is one such factor which greatly influences QoS values. Credibility can be realized through the reputation of a service provider and trustworthiness of the user about the service. Reputation is based on comparative of QoS data experienced by users who solicited these services in the past, with promised QoS assured by providers. It refers to an obligation range of completeness with regard to service rendered by providers [16]. By using UDDI QoS extension service monitors, information of available WS are obtained. Immanent issues faced while adopting above approach is, evaluation is done by users may agonize from malevolent evaluations done by other users and then reputation should not be treated equally as other attributes as it depends on various other factors. Service credibility is also realized through subjective evaluations done by users called user trust. Trust indicates aggregation of quality experiences by different users about WS. Here, past experiences of users are given more importance. The credibility of service should specify the global capability of the provider to render its guaranteed QoS.

(a) **Reputation Evaluation**

The target of reputation evaluation is to identify the range of integrity of each QoS characteristics of WS. It is done by associating actual distributed values and guaranteed attribute values. Distributed value is denoted as  $q_{di}$  for  $i$ th QoS attribute after each invocation, and guaranteed QoS attribute value as  $q_{gi}$  and is well-known as it is given by providers. Integrity level is represented as  $R_i \in [1, -1]$ , then greater  $R_i$  is the improved level of integrity that represents better reputation. QoS attributes can be subdivided into two different groups: benefit-type attribute value (throughput is high) and cost-type attribute value (response time is less). In benefit-type attribute, the integrity level is calculated as,

$$R_i = q_{di} - q_{gi}/q_{gi}. \tag{1}$$

In cost type attribute, integrity level is calculated as

$$R_i = q_{gi} - q_{di}/q_{gi}. \tag{2}$$

According to the value of  $R_i$ , reputation level for each QoS attribute is described as:

1. Excellent—if  $(0.5 \leq R_i < 1)$
2. Good—if  $(0.5 \leq R_i < 0)$ .
3. Acceptable—if  $(R_i = 0)$
4. Poor—if  $(-0.5 \leq R_i < 0)$
5. Very Poor—if  $(-1 \leq R_i < -0.5)$

(b) **Trust Evaluation**

In user trust evaluation, the trust of the user about the service reflects the user's certitude range for the service. Each QoS characteristics have a trust value based on antiquity of data about previously invoked WS. Trust of different QoS attributes for service  $ws$  is defined as  $T = \{T_{q_1}, T_{q_2}, \dots, T_{q_n}\}$  and it is calculated as,

$$T_{q_i} = \sum_{m=1}^n t_{im}(ws)/n. \quad (3)$$

where  $n$  indicates the total number of the evaluation carried out for service  $ws$  and  $t_{im}(ws)$  indicate  $m$ th evaluations for QoS attribute  $q_i$  of service  $ws$ . After finding trust value, overall credibility evaluation is carried out by accumulating both reputation evaluation and trust of the user about that service as shown below:

$$C_{q_i} = \beta R_{q_i} + (1 - \beta)T_{q_i}, \quad 0 \leq \beta \leq 1. \quad (4)$$

where parameter  $\beta$  is used to stabilize the reputation and trust of the user.

1.  $\beta = 1$ : Trustworthiness depends on historic data of previously invoked WS.
2.  $0 < \beta < 1$ : Credibility relies on the reputation of WS and trust of the user.
3.  $\beta = 0$ : Credibility depends on the experience of the user.

After obtaining overall credibility, calculate credibility-based QoS value as shown:

$$Q_{Ci} = q_{gi} \times C_{qi}. \quad (5)$$

where  $Q_{Ci}$  represents the  $i$ th credible QoS attribute value,  $q_{gi}$  is the guaranteed QoS value, and  $C_{qi}$  is the estimated credibility of  $i$ th QoS attribute.

### 3.1.2 Weight Calculation

Weightage for different QoS attribute plays an essential role in choosing the best WS. Improved weight calculation method depends on rough set theory for weight estimation [17]. Higher weights are allotted to more discriminate attributes. An essential degree of importance of quality attributes for each WS is reflected by weights of attributes. Weight is based on the count on different classes of QoS values present for QoS attribute related to the set of WS. The main goal is to discriminate different services and discover contrasts and similar among all applicant WS based on their QoS values. Find the number of different classes  $x$  for each attribute  $i$ , as  $P(i) = x$ . Then, attribute discernibility is calculated using:

$$\text{Dis}(P_i) = 1 - \text{class}/\text{class}^2. \quad (6)$$

where  $Dis(P_i)$  is discernibility of attribute  $P_i$ . Weight of each attribute  $W_{gi}$  is calculated using:

$$W_{gi} = Dis(P_i) / \sum_{i=0}^n Dis(P_i). \tag{7}$$

Finally, QoS of each quality attribute with weight factor over credible QoS value is calculated by,

$$QoS_i = \sum_{i=0}^n Q_{ci} * W_{gi}. \tag{8}$$

where  $Q_{ci}$  is credible QoS and  $W_{gi}$  is the global weight for attribute  $i$  of web service  $w_i$ . Thus, weighted credibility based QoS of each attribute for WS is calculated.

### 3.2 Genetic Algorithm for Optimized QoS

QoS prediction using weighted credibility calculation is a novel methodology, and it is further enhanced with optimal prediction. The advent of new technologies and development has relatively increased the expectation of end-users by providing plentiful services in their living rooms in a more efficient manner. This ultimately has elevated their comfort level and has made them lazy. As an after effect of this, users refuse or does not like to spend time and money in realising services of less optimal solutions. At this juncture, our focus gradually moves to the optimization process in providing the best services to end-users. The system thrives to perform optimization in two levels: QoS prediction and service composition, using evolutionary algorithms.

Genetic algorithm is used for optimization of QoS values which are based on maximization or minimization process depending on the type of quality attributes [18]. It is more effective in dealing with intricate objective and constraints. To solve an issue, they inspire subsistence of the fittest among the individual of successive generations [19]. Each generation consists of the populace of individuals. Operations like selection, fitness calculation, crossover, mutation, objective function attainment are performed iteratively, to get a set of web services that have better QoS values, with the superior fitness values after each generation. The populace with the weighted credibility-based QoS values of WS is fed into the genetic algorithm module to realize the populace of WS with optimal values which are carried over to the service composition module. The implementation details of the genetic algorithm which uses the dataset of QoS values of response time and throughput of WS are given below:

**Algorithm 1: Genetic Algorithm**

**Input:** Dataset  $D$  with WS predicted QoS values  $Q_{pi}$  for quality factors, where  $i$  is 1 to  $N$ ,  $N$  is a total of available WS.



**Output:** Population with optimized QoS factor value, fitness score of all WS.

1. Generate initial population from  $D$
2. Set standard value for each QoS attribute, i.e., minimum RT and maximum TP.
3. Evaluate the fitness score  $F_i$  for each WS in the population with relevance to corresponding predicted  $Q_{pi}$  of ws and standard value.
  - (a) Compare the place value of standard value and attribute value.
  - (b) If the same number in both place values, add it to get the fitness value of WS.
4. Sort all WS in the population with regard to fitness value.
5. Generate a new population by following the given steps:
  - (a) Choose two parent services  $p1, p2$  with highest fitness score.
  - (b) Crossover  $p1, p2$  to obtain new offspring  $c1, c2$ .
  - (c) Mutate offspring  $c1, c2$  at an arbitrary position in chromosome to get offspring  $m1, m2$ .
  - (d) Assess fitness score of new offspring.
  - (e) If fitness score of offspring  $>$  replace predicted QoS value with  $m1, m2$ .
6. Add that two WS to a new population, and remove from the old population.
7. Repeat step 5 until the population is created with optimized value for all WS.
8. Return optimized QoS of WS with a fitness score  $F_j$  where  $j$  refers to WS index.

WS with the highest fitness score is selected as the best services, and it passed on to the next module whose main goal is to find out an optimized web service composition. With those optimized QoS values, the composition of WS is invoked which adopts cuckoo optimization algorithm described in the following section.

### 3.3 Service Composition

#### 3.3.1 Cuckoo Optimization Algorithm

Cuckoo optimization algorithm (CO Algorithm) is a heuristic optimization algorithm used to render a composition that provides maximum benefit to end-users. It works based on reproduction approach of cuckoo birds which do not own nest, and they don't feed their kids. It uses a scrounging performance to incubate young birds. They lay their eggs in the nest of different winged creatures [20] or sometimes in the nest of host bird which may have various species. Unconsciously hatched cuckoos will thrust other eggs out of the nest and made a lurid cry to get host mothers feeding which develops the rate of existence. If host bird identifies that the egg is not its own, then it pushes other eggs out of the nest. Sometimes they either destroy eggs or abandon them from the nest. It results in the evolution of eggs of cuckoos which imitate eggs of native host birds. Levy flights an essential part is used in CO algorithm [21] for both local search and global search. It is defined as an arbitrary walk described in

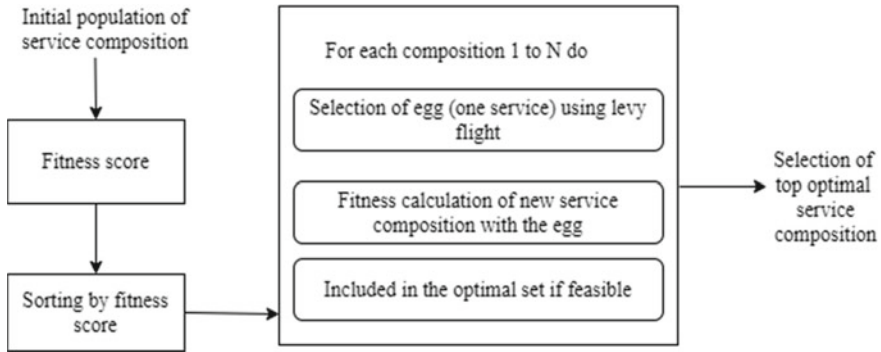


Fig. 2 CO algorithm for optimized service composition

terms of the length of the steps. The process by which the optimized composition obtained is depicted in Fig. 2.

The algorithm is applied to the optimized QoS values got from genetic algorithm. Initially, services are categorized based on their functionality domain, and required participant services are identified. Based on the requirements and availability of services in respective domains, the composition is performed in a normal manner to get the list of all possible combinations of services. Fitness score of WS is considered as local optima and fitness score of the web service composition as global optima. Objective or fitness function is set up with standard fitness value which is a vector created by appending minimum RT and maximum TP of WS identified for every domain which is the part of service composition. The fitness value of every web service composition is calculated. For each composition, the process selecting a WS using levy flight, replacing selected services with existing service to create a new composition, calculating fitness score, comparing with standard fitness score and appending composition with high fitness score to new population is performed. Thus, a new population with promising composition is got. Input to CO Algorithm have optimized QoS values of response time (RT) and throughput (TP) for all WS in that composition.

**Algorithm 2: Cuckoo Optimization Algorithm**

**Input:** Set of WS from the required domains, with its optimized QoS factor value  $Q_{opi}$ , where  $opi$  is 1 to  $N$  and  $N$  is the quantity of WS.

**Output:** Best optimal web service combination.

1. Arrange WS based on domain and find a combination of WS.
2. Find standard fitness value [Ambulance [ $\min(RT)$ ,  $\max(TP)$ ], Blood Bank ( $\min(RT)$ ,  $\max(TP)$ ), Pharmacy ( $\min(RT)$ ,  $\max(TP)$ ).
3. Finalize standard fitness vector and fitness function.
4. Each combination of WS
  - (a) Compute fitness value as vector of QoS value of RT, TP of all services in that composition.

5. Sort all web service combinations with fitness value.
6. Initial population contains list of WS got from step 5.
7. For each combination of WS in the initial population.
  - (a) Using levy flight, select a WS from list WS maintained for each domain.
  - (b) Replace WS selected with existing WS in the composition chosen.
  - (c) Find fitness value, i.e., create the vector of QoS value.
  - (d) Execute fitness function with fitness value of selected composition and standard fitness vector to get the fitness score.
  - (e) Add WS composition with highest fitness score to the new population if it does not already exist.
8. New population contains the list of optimized composition with fitness score.
9. Sort the new population with respect to fitness score.
10. Return the best WS composition with highest fitness score.

### 3.3.2 Levy Flight

Levy flight algorithm used in CO algorithm for selecting WS from the list of all WS in the selected domain is explained below:

#### Algorithm 3: Levy Flight Algorithm

**Input:** List of participant WS for each domain and its respective pointers.

**Output:** Selected web service.

1. Set  $r1 =$  global walk distance with a random value higher than 15 and  $S = 1$
2. While number of steps  $S < r1$ 
  - (a) Set  $r$  with random number between 5 to 8 and do the following  $rtimes$ :
    - (i) Take one step - LOCAL WALK - select any one participant services to move the respective pointer of that service list to next position performed circularly.
    - (ii) Store  $w =$  pointer value of WS list of service used in previous step.
    - (iii) Increase number of steps  $S = S + 1$
    - (iv) Store  $ws =$  Domain name of service whose pointer is moved.
  - (b) Take one step - GLOBAL WALK—go to list with domain  $ws$  and move pointer to 10 locations ahead, pointer movements is in circular fashion.
  - (c) Store  $w =$  pointer value of WS list of services used in previous step
  - (d) Increase number of steps  $S = S + 1$
3. End While
4. Return selected WS  $w$  with its domain  $ws$ .

CO algorithm can find the best composition of WS which gives maximum benefit to end-users, as it considers different combinations with its fitness score and selects combinations with high scores ignoring less promising combinations.

### 4 Experiment Result and Analysis

This research work experimented for the composition of services on the medical domain, which considers ambulance, pharmacy, and blood bank services. It is done with a dataset of WS in the above three domains which includes guaranteed and distributed quality factor value of response time and throughput. Those values of WS are considered for calculating reputation evaluation, trust evaluation to all WS, and it is done separately for throughput and response time. Reputation is calculated by correlating real distributed attribute values and guaranteed attribute values. Trust evaluation is done based on feedback given by users for WS with their experience. Trust value reflects the user’s confidence level related to the WS. Table 1 shows a sample set of WS with their guaranteed TP, distributed TP, reputation TP, trust TP and credibility values.

In similar manner, credibility for RT is also calculated an optimized QoS value result of the genetic algorithm is shown in Table 2. The values are normalized based on the maximum TP and minimum RT.

Service composition realized with different candidate services and fitness score of the initial population is calculated. Then, those compositions are subject to the cuckoo optimization process, in which individual candidate services are replaced with WS selected with levy flight. Implementation details of levy flight algorithm with three services blood bank, pharmacy and ambulance are shown as follows:

**Table 1** Predicted QoS values using credibility

S. No.	WS	Guaranteed TP	Distributed TP	Reputation TP	Trust TP	Credibility TP
1	WS7	120.90	437.72	0.7237	116.88	468.44
2	WS5	9.1741	180.1	0.9490	9.6368	118.80
3	WS4	7.2777	25.352	0.7129	6.9425	16.711
4	WS134	32.179	117.39	0.7258	32.179	75.987
5	WS202	4.7557	7.9641	0.4028	4.7557	47.479
6	WS130	20.104	48.213	0.5830	20.104	65.271
7	WS228	56.742	900.1	0.9369	56.742	660.59
8	WS231	60.341	900.1	0.9329	60.341	93.73
9	WS232	60.197	900.1	0.9331	60.197	625.42
10	WS257	63.229	787.5	0.9197	63.229	464.36

**Table 2** Predicted QoS using genetic algorithm

WS7	WS5	WS4	WS134	WS202	WS130	WS228	WS231	WS232	WS257
420	113	19	71	43	69	610	90	624	425

1. Assign 4 variables count, a, b, p with values 0, -1, -1, -1.
2. Randomly select a number between 15 - 30, store the number in a variable r1.
3. While (*count* <= *r1*)
  - For loop up to 5 times
    - Set *r2* = random number in 1, 2, 3
    - If *r2* = 1 then check if *a* <= no. of *ambulance ws* then *a* = *a* + 1  
Otherwise *a* = -1, assign *w* = *a*, *ws* = *ambulance*.
    - If *r2* = 2 then check if *p* <= no. of *pharmacy ws* then *p* = *p* + 1  
Otherwise *p* = -1, assign *w* = *p*, *ws* = *pharmacy*.
    - If *r2* = 3 then if *b* <= no. of *bloodbank ws* then *b* = *b* + 1  
Otherwise *b* = -1 and assign *w* = *b*, *ws* = *bloodbank*.
  - Set *count* = *count* + 1
  - End for
  - Check the value of *ws* variable
    - If *ws* = *ambulance* then *a* = (*a* + 5)% 10, *count* = *count* + 1, *w* = *a*
    - If *ws* = *pharmacy* then *p* = (*p* + 5)% 10, *count* = *count* + 1, *w* = *p*
    - If *ws* = *bloodbank* then *b* = (*b* + 5)% 10, *count* = *count* + 1, *w* = *b*
  - Set *count* = *count* + 1
4. Return WS stored in *ws* to be used for replacing existing WS in the composition subject to the optimization evaluation process.

New composition created with replacement is subject to fitness score evaluation and if promising it is included in the new population, otherwise, it is discarded and only the old composition is included. This process continues to get a population of optimal values. Table 3 shows a comparative study of composition before and after optimization. It shows candidate services from each domain selected for the composition, with their fitness value specified as a vector of RT and TP.

## 5 Conclusion

Dynamic service composition is widely used to solve varied real-time problems and to find the best suited business solutions which seem to be more complex. It involves identification of best candidate services and integration of services as per the requirements of the problem. A novel method of QoS-based service selection is adopted. It predicts QoS of WS with regard to credibility, reputation and trustworthiness. Weight calculation and optimization of the QoS prediction are also performed. The services with high QoS values are taken for service composition. This leads to a reasonable set of services competing to be part of service composition. To get an optimal service composition, cuckoo-based algorithm with levy flight is adopted. A cuckoo algorithm is implemented to get the best composition from an existing population which satisfies fitness function, and Levy flight is used to select individual services to be included in service composition. It is proved that a population with an optimal composition of services is created, and it helps to select required composition with the highest fitness score. In future, service composition can be focused and improved with contextual-based semantic service selection.

**Table 3** Composition of WS before and after cuckoo composition

No.	Top scoring composition and fitness score before cuckoo optimization	Top scoring composition and fitness score after cuckoo optimization
1	<b>WS7, WS134, WS228</b> [54.37, 57.46, 87.65, 11.78, 76.03,81.56]	<b>WS10, WS125, WS229</b> [400.37, 20.46, 352.65, 11.78, 502.03, 14.56]
2	<b>WS7,WS134, WS231</b> [54.37, 57.46, 87.65, 11.78, 100.90,73.53]	<b>WS10, WS120, WS233</b> [400.37, 20.46, 310.68, 13.73, 443.90, 15.53]
3	<b>WS7, WS134,, WS232</b> [540.03, 57.46, 87.65, 11.78, 71.46,75.21]	<b>WS10, WS132, WS232</b> [400.37, 20.46, 309.65, 11.78, 42.46, 25.21]
4	<b>WS4 WS130 WS228</b> [19.27, 12.19, 17.29, 50.98, 76.03, 81.56]	<b>WS19, WS145, WS229</b> [38.27, 22.19, 310.29, 15.98, 439.03, 20.56]
5	<b>WS5, WS134, WS232</b> 137.05,, 11.41, 87.65, 11.78, 71.46, 75.21]	<b>WS10, WS145, WS232</b> [137.05, 21.41, 310.29, 15.98, 421.46, 25.21]
6	<b>WS5, WS202, WS257</b> [137.05, 11.41, 154.77, 60.95,135.67,38.69]	<b>WS19, WS206, WS250</b> [389.27, 22.47, 354.77, 19.95, 403.67, 26.69]
7	<b>WS4, WS202, WS228</b> [19.27, 12.19, 54.77, 660.95, 76.03, 81.56]	<b>WS19, WS201, WS234</b> [389.27, 22.47,304.75, 23.95, 332.03, 23.56]
8	<b>WS7, WS134, WS257</b> [540.37, 57.46, 87.65, 11.78, 135.67, 81.56]	<b>WS25, WS132, WS252</b> [348.37, 17.46, 287.65, 11.78, 335.67, 28.69]
9	<b>WS5, WS130, WS257</b> [137.05, 11.41, 17.29, 50.98, 135.67, 38.69]	<b>WS25, WS132, WS244</b> [348.37, 17.46, 287.65, 11.78, 215.67, 21.69]
10	<b>WS5, WS130, WS231</b> [137.05, 11.41, 17.29, 50.98, 100.90, 73.53]	<b>WS25, WS129, WS244</b> [348.37, 17.46, 275.29, 15.98, 332.03, 20.56]

## References

1. Mumbaikar S, Padiya P (2003) Web services based on soap and rest principles. *Int J Scient Res Publ* 11:17–32
2. Rostami NH, Kheirkhah E, Jalali M (2013) Web service composition methods and techniques: a review. *Int J Comput Sci Eng Inf Technol* 3(6):10–5121
3. Kaewbanjong K, Intakosum S (2015) QoS attributes of web services: a systematic review and classification. *J Adv Manage Sci* 3(3):194–202
4. Hanna S, Alawneh (2010) An approach of web service quality attributes specification. *Commun IBIMA J*. ISSN:1943-7765
5. Subbulakshmi S, Ramar K, Krishna VCK, Sanjeev S (2018) Optimized QoS prediction of web service using genetic algorithm and multiple QoS aspects. In: 2018 international conference on advances in computing, communications and informatics (ICACCI). IEEE

6. Mareli M, Twala B (2018) An adaptive Cuckoo search algorithm for optimization. *Appl Comput Inf* 14(2):107–115
7. Subbulakshmi S, Elsa Saji A, Chandran G (2020) Methodologies for selection of quality web services to develop efficient web service composition. In: 2020 fourth international conference on computing methodologies and communication (ICCMC). IEEE
8. Subbulakshmi S, Ramar K, Renjitha R, Sreedevi TU (2016) Implementation of adaptive framework and WS ontology for improving QoS in recommendation of WS. In: The international symposium on intelligent systems technologies and applications, pp 383–396. Springer, Cham
9. Aljazzaf Zainab (2015) Bootstrapping quality of web services. *J King Saud University-Comput Inf Sci* 27(3):323–333
10. Zhang H, Shao Z, Zheng H, Zhai J (2014) Web service reputation evaluation based on QoS measurement. *Sci World J* 2014, Article ID 373902, 7 pages, <https://doi.org/10.1155/2014/373902>
11. Karimi M, Esfahani FS, Noorafza N (2015) Improving response time of web service composition based on QoS properties. *Indian J Sci Technol Indian J Sci Technol* 8(16):1–8
12. Mustafa F, McCluskey TL (2008) Dynamic web services composition: current issues. University of Huddersfield, pp 48–54
13. Sathya M et al (2010) Evaluation of QoS based web-service selection techniques for service composition. *Int J Softw Eng* 1(5):73–90
14. Veena G et al (2016) A concept-based model for query management in service desks. *Innovations in computer science and engineering*. Springer, Singapore, pp 255–265
15. AlSedrani A, Touri A (2016) Web service composition processes: a comparative study. *Int J Web Service Comput (IJWSC)* 7(1): 1–21
16. Aggarwal S et al (2014) Providing web credibility assessment support. In: Proceedings of the 2014 european conference on cognitive ergonomics
17. Wang R, Zhang L, Lu X (2014) Crop evaluation system optimization: attribute weights determination based on rough sets theory. In: Proceedings of European conference on cognitive ergonomics
18. Geetha Lekshmy V, Anusree PK, Varunika VS (2018) An implementation of genetic algorithm for clustering help desk data for service automation. In: International conference on advances in computing, communications and informatics (ICACCI)
19. Ai L (2011) QoS-aware web service composition using genetic algorithms. Diss. Queensland University of Technology
20. Kamoona AM, Patra JC, Stojcevski A (2018) An enhanced cuckoo search algorithm for solving optimization problems. In: 2018 IEEE congress on evolutionary computation (CEC). IEEE
21. Guerrero M, Castillo O, Garcia M (2015) Cuckoo search via levy flights and a comparison with genetic algorithms fuzzy logic augmentation of nature inspired optimization met heuristics. Springer, Cham, pp 91–103

# Emotion Recognition from Speech Signal Using Deep Learning



Mayank Chourasia, Shriya Haral, Srushti Bhatkar, and Smita Kulkarni

**Abstract** Emotions play a vital role in a human's mental life. Speech is a medium through which expression of perspective and identification of one's mental state is possible. Recognizing the feelings that others are trying to convey through speech is essential. There are various parameters of the speech signal that define the feelings of a person. Thus, speech emotion recognition (SER) from the speech signal is a challenging task. This paper proposed an SER system based on the features extracted and obtained by Mel-frequency cepstral coefficient (MFCC) spectrograms. As for a complex model, the audio features obtained by MFCC play an important role in emotion recognition. The 1D-CNN model architecture is implemented in this paper. The work is performed on "The Ryerson Audio-Visual Database of Emotional Speech and Song" (RAVDESS) dataset. Six emotions based on their gender are classified (i.e.,  $6 \times 2$  emotions) with 82.3% accuracy. The emotions classified are happy, sad, angry, calm, fear, and nervous.

**Keywords** Speech emotion recognition (SER) · Mel-frequency cepstral coefficient (MFCC) · Deep learning · Convolution neural network (CNN)

---

M. Chourasia · S. Haral (✉) · S. Bhatkar · S. Kulkarni  
School of Electrical Engineering, MIT Academy of Engineering, Pune, India  
e-mail: [shharal@mitaoe.ac.in](mailto:shharal@mitaoe.ac.in)

M. Chourasia  
e-mail: [mkchourasia@mitaoe.ac.in](mailto:mkchourasia@mitaoe.ac.in)

S. Bhatkar  
e-mail: [shbhatkar@mitaoe.ac.in](mailto:shbhatkar@mitaoe.ac.in)

S. Kulkarni  
e-mail: [sskulkarni@entc.maepune.ac.in](mailto:sskulkarni@entc.maepune.ac.in)



# 1 Introduction

Feeling assumes a major job in day by day relational human communications. It helps us comprehend and coordinate our feelings by passing on our sentiments and offering input to other people. Research has uncovered the ground-breaking job that feelings play in forming human social interaction. Emotions display the pass on data about the emotional condition of an individual and have extended a new analysis of the field called computed feeling recognition, which has a fundamental objective to comprehend and reclaim the desired feelings. SER expects to recognize the basic soul of a speaker from the voice. Expanding research enthusiasm during the recent years right now has been gotten by the world. There are numerous utilizations of distinguishing the feeling of the people like, inside the interface with robots, banking, call focuses, PC games, and many more. For study hall organization or E-learning, data about the soul of researchers can be spent significant time in the upgrade of educating quality. For example, to settle on a choice on what subjects are regularly instructed and should be prepared to create techniques, an instructor can utilize SER for overseeing feelings inside the preparation condition.

There are two utilized delineations of feeling persistent and discrete. Inside the persistent portrayal, the feeling of articulation is frequently communicated as constant qualities along with different mental measurements. Considering the investigation of [1], “feeling is frequently portrayed in two measurements: initiation and valence.” Activation is that the “measure of vitality required to exact a specific feeling” and research has demonstrated that regularly connected to high vitality and delve in discourse are an outrage, satisfaction, and dread, while pity is regularly connected to low vitality and moderate discourse. Valence gives more subtleties and recognizes feelings like being furious and glad since expanded enactment can show both inside the discrete portrayal, feelings are regularly discretely communicated as explicit classes, as irate, tragic, upbeat, and so forth. The exhibition of a feeling acknowledgment framework depends on highlights separated from the sound signal. This examination has some expertise in recognizing the more exact sound feeling classifier. Figure 1 shows the rate chosen for the different emotions for the actor intended for eight different emotions which were neutral, calm, happy, sad, angry, fearful, disgust, and surprise.

		Actor intended emotion							
		Neutral	Calm	Happy	Sad	Angry	Fearful	Disgust	Surprise
Rater chosen emotion	Neutral/Calm	86.6	69.9	14.25	17.12	4.03	4.5	4.36	7.03
	Happy	0.63	17.27	68.44	1.48	0.23	0.59	0.59	6.56
	Sad	4.65	6.06	2.29	60.85	1.02	6.58	8.65	0.76
	Angry	3.82	1.02	1.79	2.9	81.32	4.79	6.48	2.78
	Fearful	0.63	0.66	1.67	9.64	1.39	70.71	2.31	2.22
	Disgust	1.15	1.46	0.78	3.09	8.37	1.81	69.77	3.28
	Surprise	0.28	0.33	7.88	0.69	1.2	7.76	4.13	72.29
	None	2.26	3.3	2.9	4.24	2.45	3.26	3.72	5.07

Fig. 1 Rate chosen for emotions [15]

This paper proposed to recognize emotion, based on the speech signal. For recognition of emotion, the features are extracted from speech and are classified using a deep learning model. The paper is divided into 3 sections. In Sect. 2, a literature survey, the related work by different researchers is being studied. In Sect. 3, the selection of the dataset which is RAVDESS, studying the dataset and then preparing the 7 segments numeric value for the data augmentation. In Sect. 4, features are extracted using MFCC from the audio file then training the CNN model for testing to find the accuracy for different input to the CNN.

## 2 Literature Survey

Different studies have been performed in this SER domain, on various datasets and based on the diverse aspects and features of the speech signal. Different examinations have been done in the past to recognize feelings from discourse for various dialects and accents. The contemplated exhibition of LPCC in distinguishing the feelings utilizing HMM and SVM. The created model yielded a 61% precision on SAVEE database [2]. The proposed DNNs, RNNs, and 1D-CNN models based on MFCC highlights on the Chinese language dataset accomplished 56% exactness [3]. The HMM on MFCC and LPCC to extract the features from Burmese language speakers and the Mandarin language speaker dataset which obtained the accuracy of 90% and an average accuracy of 78% [4]. Using the gaussian classifier, the classification of the emotions was made for the English audiovisual emotional database, which consisted of 4 male actors in 7 emotions. Features are extracted with the help of principal component analysis and linear discriminant analysis. The accuracy obtained was 98% [5]. The model was implemented by extracting robust and effective features using MFCC, then normalized for isolating digits recognition. For the English language database, with an accuracy of 95% on augmented data [6]. Quranic recitation recognition was done on the Quranic database with features extraction using MFCC and DFT [7].

The experiment carried on the IITKGP-SEC and IITKGP-SEHSC speech corpus databases for speech emotion recognition. Different models were proposed for feature extraction and also for implementation using GMM-HMM, AANN-GMM, and HMM-AANN. Results found that the AANN when combined with GMM and HMM, shows better performance. The accuracy obtained for GMM-AANN was 44.52% and HMM-AANN was 39% [8]. The system proposed is for emotion recognition from the speech on the English database with five emotional states for a person. Feature extraction made using the traditional approach of MFCC. Using KNN, the classification made depending on the frequency range. The overall success rate was 63.63%. For males, the success rate of 72.72% and for females, it was 54.54% [9]. Speech recognition was made using acted and real emotion database. Feature engineering made using GMM and MFCC for the classification of emotions and gender. A total of 21 features were obtained using MFCC, and using GMM 16 components were constructed with an accuracy of 95% [10].

This examination centers around distinguishing the best sound component and model design for feeling acknowledgment in discourse. The work was performed on the RAVDESS dataset. In this paper, a one-layer CNN model proposed with Log-Mel spectrogram sound highlights.

### 3 Dataset

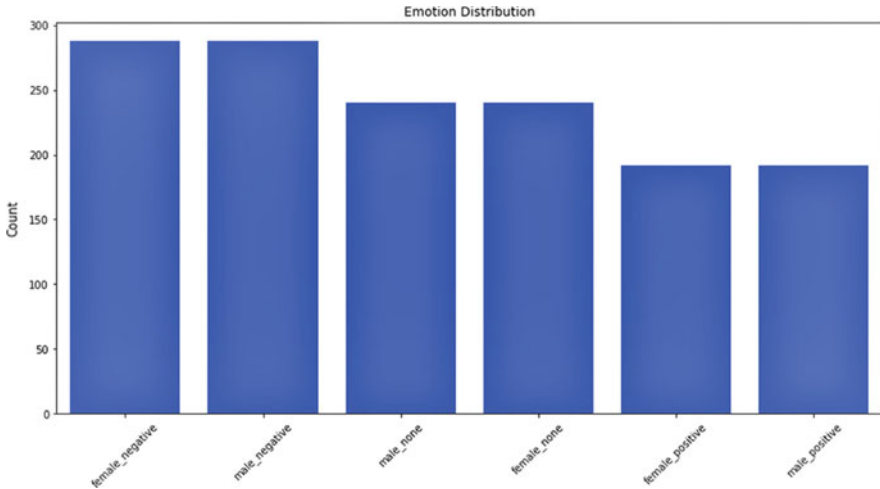
#### 3.1 Dataset Selection

The fundamental parts for structuring discourse feeling acknowledgment are the database, highlights removed with the correct classifier for sound, and video documents. Utilizing these segments, one can have limitless models with different exactnesses. The dataset utilized is RAVDESS [11]. The dataset suits 7356 documents. The database contains 24 skilled on-screen characters with a North American inflection. This sexual orientation adjusted skillful on-screen character's database comprises 12 males and 12 females having two sorts of records which are discourse and tune. The discourse comprises feelings happy, sad, neutral, calm, angry, fearful, disgust, and surprised [11]. For a tune, feelings are calm, happy, sad, angry, and fearful [11]. Every on-screen character captivated two lexical articulations with various "Kids are talking by the door" and "Dogs are sitting by the door" [11]. Each vocalization by the entertainers was rehashed twice. There is a whole of 1440 discourse expressions and 1012 tune articulations. Every articulation is worked at two degrees of enthusiastic force (ordinary, amazing), with a further impartial articulation. All conditions are accessible in three methodology positions, Audio-Video (720p H.264, AAC 48 kHz, .mp4) [11], Audio-just (16bit, 48 kHz .wav) [11] and Video-just (no stable) [11].

Distinctive datasets, for instance, SAVEE [5] and TESS [12], Berlin database [13], and Chinese dataset [14] composed of sounds from just male and female entertainers individually. The RAVDESS dataset is similarly appropriated overall feeling classes (approximately 15%), so, it does not experience the ill effects of any class-awkwardness issues.

Figure 2 shows the distribution graph of gender and emotion based on the number of recordings available in the RAVDESS dataset. The emotions plotted are in terms of positive, negative, and none type. In a positive type, emotions include happiness and surprise. The negative emotions consist of sad, disgust, angry, and fearful. For none type, the emotions are neutral and calm. The plot gives the number of positive males, positive females, negative male, and so on.

The dataset has hardly any deficiencies. At first, the way where an inclination is indicated is unequivocally settled on the language, supplement, vernacular, and social establishment. A model arranged to perceive feelings on the English dataset presumably won't have the choice to recognize sentiments in the Chinese language. In this paper [15], RAVDESS [11] shows strong North American qualities. Likewise, the dataset is prepared by arranging performance as a basic event of emotion.



**Fig. 2** Emotion Distribution [11]

### 3.2 *Speech Data Augmentation*

Every one of the 7356 documents has an interesting filename which is a 7 section numerical identifier. These identifiers characterize the boost in attributes based on the various emotions like happy, sad, disgust, neutral, clam, fearful, and surprised. These identifiers also consist of various parameters like modulation, vocal channel, intensity, redundancy, gender, and sentences.

The initial input speech signal is added with some noise and pitch signals. This concatenation helps to obtain the features more accurately. This procedure is performed to compensate for any unvoiced signal frame in the input. This also helps in obtaining diversity in the same input voice signal. The graph of this native speech signal, the noise, and pitch are shown in Figs. 3, 4 and 5 as follows.

## 4 Feature Engineering and Modeling

### 4.1 *Features*

**MFCC:** Speech is a convolved signal. The vocal cords produce the excitation function, due to vibration produced by the air, and the vocal tract system functions as the filter. The transfer function of the vocal tract the human speech mechanism considers the vocal tract system and source of excitation as an independent. The speech signal presented in various features; few of them are zero-crossing rate, minimum energy, maximum amplitude, the energy of the signal [7].

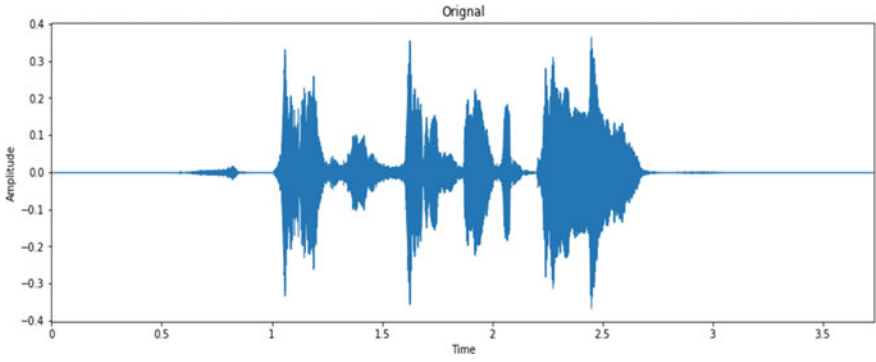


Fig. 3 Original sound signal

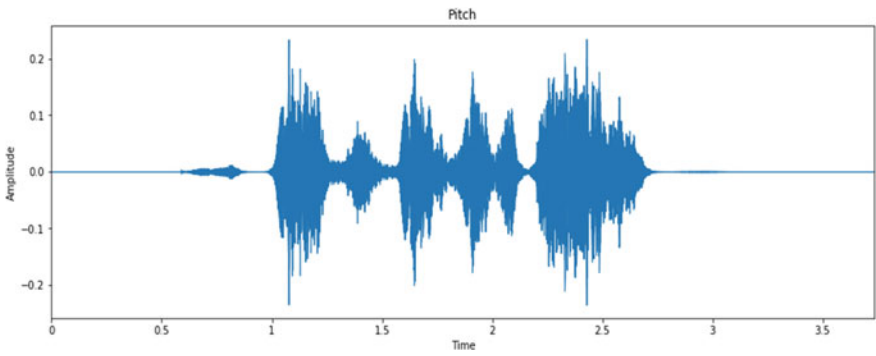


Fig. 4 Pitch + Original

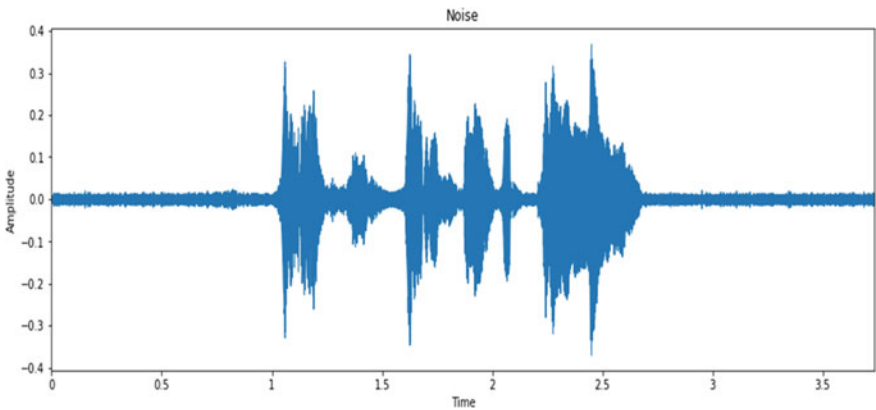
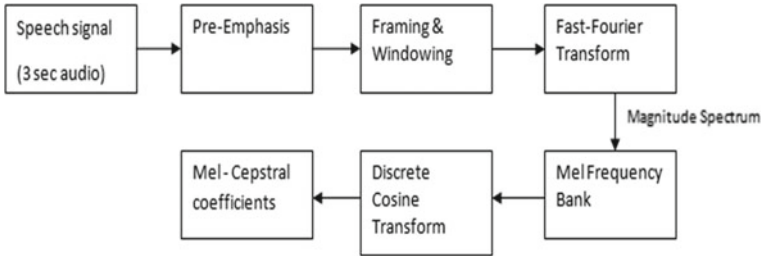


Fig. 5 Noise + Original



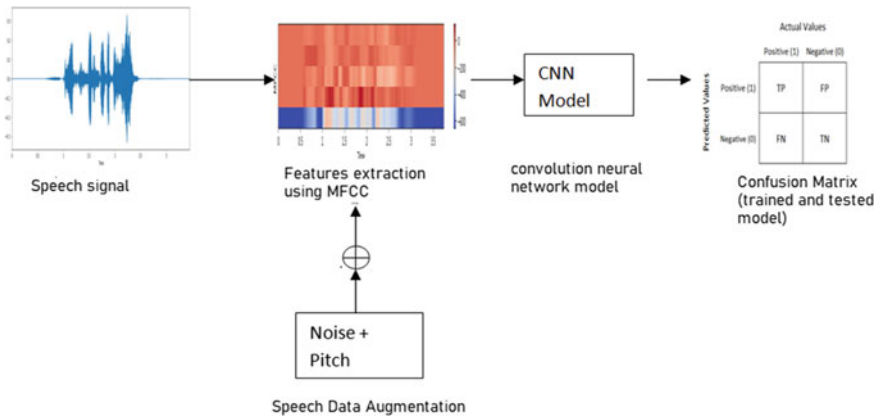
**Fig. 6** Steps of MFCC [6]

Mel-frequency cepstrum (MFC) presents a short-term power spectrum of a sound. MFCC is commonly used as a feature extraction technique in the speech recognition system. The human sound produced depends upon the cross-section area as well as the length of the vocal tract. The cepstrum is the representation of the sound after taking the fourier transform of the log spectrum. MFC coefficients (MFCC) represents the coefficient of a short-time power spectrum as a seize envelope. MFCC helps to identify the emotions through the features extracted.

MFCC is obtained when the audio signal passes through various operations as pre-emphasis, framing, and windowing, FFT, Mel-frequency bank, DCT, and Mel-cepstral coefficient. As shown in Fig. 6, the audio signal considered is of 3 s duration. Emphasizing the higher frequencies in the speech signal is the objective of pre-emphasizing. This framing is based on two factors: the number of frames and the distance between the two consecutive frames. Multiplying the framed signal with the window helps to maintain the continuity in a signal. There are various types of windows such as Hanning, Kaiser, Hamming, and so on which according to the application requirement can be used. The main purpose of the windowing technique is for tapering of the signal and not segmentation. The window used in this work is the Kaiser Window. A Mel-filter bank is used to get an overview. Mel-scale cepstral coefficients the DCT is applied to the output obtained at the triangular bandpass filter.

## 4.2 Deep Learning Model

Deep learning CNNs model which is current progressive models. CNN model is a combination of normally N layers depending upon the usage of applications like 1D-CNN, 2D-CNN, AlexNet [16], LSTM [17] are the various types of architecture proposed in CNN. In our proposed model of CNN, 1D-CNN architecture is used because it gives better results for speech due to the 1D-vector. In our model, the convolution layer is the first layer that contains the number of filters used to extract important features. The second layer is the pooling layer which is used for downsampling especially of the options mapped. The third layer is the fully connected layer of CNN that is mainly proposed for extracting the output option fail to Soft-Max



**Fig. 7** Speech emotion recognition model

classifier [16] to speak out the likelihood for every category. Accuracy also depends on which optimization technique is being preferred and feature selection of speech signal.

As shown in Fig. 7, the model presented the input speech signal; i.e., the .wav file from which features are extracted through the MFCC technique. In the speech, lots of features can be abstracted but focused on MFCC only. As to provide versatility, with the noise and pitch re-added to these extracted MFC coefficients are then given as input to the CNN 1D model. After the input is given to the model, it is then trained and tested.

## 5 Results

The proposed research work carried out on the MFCC features for modeling SER is an overall survey of various features and modeling for SER. Thus, the obtained results are based on the MFCC features that improved performance due to deep learning.

Comparing the results among the different models based on the state-of-art method shown below in Table 1.

The 1D-CNN model gives a better result as presents in Table 1. It is observed that emotion classification with gender specification performance is higher. Additionally, it is observed the energy and pitch having differences in the voice of the female and male. This leads to differences in the pattern of females from male emotions. Also appending features like energy and pitch to the MFCC improves the overall performances of the model. It is observed that the proposed 1D CNN prediction performance improved with an accuracy of 82.3% while testing the data samples of the given dataset as shown in Fig. 7. Whereas, the training accuracy obtained is

**Table 1** Comparing the result among the different approaches

Approach	Classifier	Feature	Accuracy (%)
Approach 1 [8]	GMMs	MFCCs	73.68
Approach 2 [9]	Nearest neighbor algorithm	MFCC	63.63
Approach 3 [10]	Simple SVM classifier	Pitch and prosody features	81.132
Proposed approach	1D-CNN	MFCC	82.3

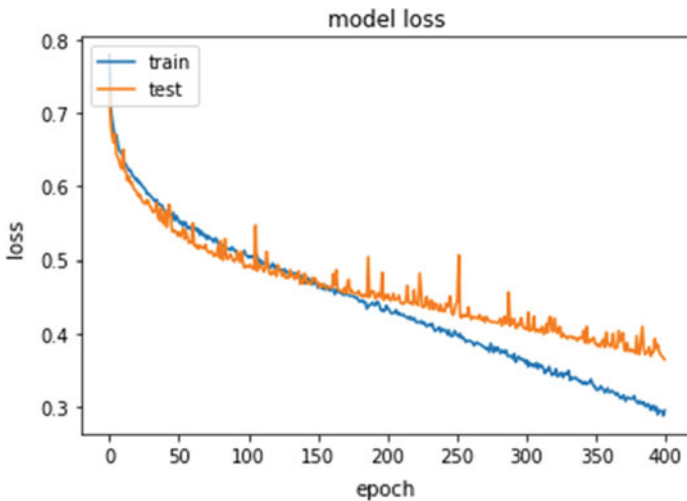
83.3%. This accuracy obtained is for the classification of 6 emotions: anger, happy, sad, fearful, calm, and nervous.

As stated previously, the low precision and accuracy can be explained by the number of features that define a speech signal. Also, the variety of emotions expressed by different people that increase the complexity of the problem statement.

Comparing the results with the DCNN (2 convolution and 2 pooling layers) [18] that extracted features using PCA, the overall accuracy obtained was 40%.

In this work, the obtained output graph of the training and testing model is shown in Fig. 8. This graph gives us a comparison between the training and testing model based on epochs. As epoch plays an important role, it helps us improve the accuracy. However, after a specific epoch, accuracy tends to remain constant.

A confusion matrix is a performance measurement for the deep learning classification problem where the output can be two or more classes. It is a table with 4



**Fig. 8** Training and testing comparison





**Fig. 9** Confusion matrix

different combinations of predicted and actual values represented as true positive, true negative, false positive, and false negative, respectively, after training and testing. It helps us to visualize the insight of our CNN model as shown in Fig. 9. It also helps in understanding the feature engineering using MFCC for the proposed model.

## 6 Conclusion

In this paper, speech emotion recognition is presented by using MFCC based on 1D-convolutional neural networks. Speech signal represented in terms of array values acts as input to the CNN. The CNN model consists of three convolutional as well as fully connected layers extract features from the spectrograms and predictions at the output is for six emotion classes which are anger, happy, sad, fearful, calm, and nervous. In this regard, the experiment was executed with a satisfactory result with a testing accuracy of 82.3% was achieved based on the RAVDESS dataset. Further experimentation is required to improve the accuracy of emotion recognition efficiently. In the proposed work plan to make use of the different database with more data having varying statements and a relatively multifarious model to improve the SER performance further.

## References

1. Ayadi ME, Kamel MS, Karray F (2011) Survey on speech emotion recognition. In: Features, classification schemes, and databases. *Pattern Recog* 44(3):572–587. <https://doi.org/10.1016/j.patcog.2010.09.020> (2011)

2. Chenchah F, Lachiri Z (2015) Acoustic emotion recognition using linear and nonlinear cepstral coefficients. *Int J Adv Comput Sci Appl* 6(11):135–138
3. Parthasarathy S, Tashev I (2018) Convolutional neural network techniques for speech emotion recognition. In: 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), Tokyo, pp 121–125. <https://doi.org/10.1109/IWAENC.2018.8521333>
4. New TL, Foo SW, De Silva LC (2003, November) Speech emotion recognition using hidden Markov models. *Elsevier Speech Commun J* 41(4):603–623
5. Haq S, Jackson PJB (2009) Speaker-dependent audio-visual emotion recognition. In: Proceedings of international conference on auditory-visual speech processing, pp 53–58
6. Lokhande NN, Nehe NS, Vikhe PS (2012) MFCC based robust features for English word recognition. In: 2012 Annual IEEE India conference (INDICON), Kochi, pp 798–801. <https://doi.org/10.1109/INDCON.2012.6420726>
7. Razak Z, Ibrahim NJ, Tamil EM, Idris MYI, Yusoff MY (2007) Quranic verse recitation feature extraction using Mel frequency cepstral coefficient (MFCC). University Malaya
8. Bhaykar M, Yadav J, Sreenivasa Rao K (2013) Speaker dependent speaker independent and cross language emotion recognition from speech using GMM and HMM. In: Communications, NCC, National Conference on IEEE, pp 1–5
9. Sapra A, Panwar N, Panwar S (2013, February) Emotion recognition from speech. Jaypee Institute of Information Technology, Noida. *Int J Emerg Technol Adv Eng* 3(2). [www.ijetae.com](http://www.ijetae.com). ISSN 2250-2459, ISO 9001:2008 Certified Journal
10. Lalitha S, Madhavan A, Bhushan B, Saketh S (2014) Speech emotion recognition. Department of ECE, Amrita School of Engineering, Amrita Vishwa Vidyapeetam, Bangalore, India. In: International conference on advances in electronics, computers and communications (ICAEECC)
11. Fux T, Jouvét D (2015) Evaluation of PNCC and extended spectral subtraction methods for robust speech recognition. In 23rd European signal processing conference (EUSIPCO) and nice, pp 1416–1420. <https://doi.org/10.1109/eusipco.2015.7362617>
12. Dupuis K, Pichora-Fuller MK (2010) Toronto emotional speech set (TESS). In: University of Toronto, Psychology Department
13. Berlin Database of emotional speech. <http://pascal.kgw.tu-berlin.de/emodb/index-1280.html>
14. Jing S, Mao X, Chen L et al (2015) Annotations and consistency detection for Chinese dual-mode emotional speech database. *J Beijing Univ Aeronaut Astronaut* 41(10):1925–1934 (in Chinese)
15. Venkataramanan K, Rajamohan HR (2019) Emotion recognition from speech. [arXiv:1912.10458v1](https://arxiv.org/abs/1912.10458v1) [cs.SD] 22 Dec 2019
16. Zheng WQ, Yu JS, Zou YX ADSPLAB/ELIP. In: School of Electronic Computer Engineering Peking University Shenzhen, China
17. Sopra Storia. <https://blog.soprasteria.se/2019/10/07/using-cnn-for-speech-emotion-recognition/>
18. Zheng WQ, Yu JS, Zou YX An experimental study of speech emotion recognition based on deep convolutional neural networks

# COVID-19 Database Management: A Non-relational Approach (NoSQL and XML)



Priya Shah, Rajasi Adurkar, Shreya Desai, Swapnil Kadakia,  
and Kiran Bhowmick

**Abstract** The advancing COVID-19 pandemic caused by the novel coronavirus has taken the world by a storm due to its unprecedented nature. In order to increase the understanding of the disease and create countermeasures for the same, collecting and storing data in a proper and efficient format is of utmost importance. However, this tremendous amount of data is obtained from various heterogeneous sources and is usually dynamic in nature. Traditional RDBMS might not be the most efficient choice for the sporadic and ever-changing clinical data associated with COVID patients due to its highly rigid nature. This paper utilized a primary dataset acquired from COVID-19 patients as a premise to portray the inefficiencies of RDBMS and further proposes two new schemaless, unstructured databases, NoSQL and XML databases, as an offset to this drawback. The intention is to propose the two most efficient technologies and delineate the findings through a sample implementation.

**Keywords** COVID-19 · Clinical data · Non-relational databases · NoSQL · XML

---

All authors have contributed equally to the paper.

---

P. Shah · R. Adurkar · S. Desai · S. Kadakia (✉) · K. Bhowmick  
Department of Computer Engineering, D J Sanghvi College of Engineering, Mumbai, India  
e-mail: [swapnilkadakia@gmail.com](mailto:swapnilkadakia@gmail.com)

P. Shah  
e-mail: [pshah3103@gmail.com](mailto:pshah3103@gmail.com)

R. Adurkar  
e-mail: [adurkar.rajasi562@gmail.com](mailto:adurkar.rajasi562@gmail.com)

S. Desai  
e-mail: [shreyadesai1202@gmail.com](mailto:shreyadesai1202@gmail.com)

K. Bhowmick  
e-mail: [kiran.bhowmick@djsce.ac.in](mailto:kiran.bhowmick@djsce.ac.in)

## 1 Introduction

The COVID-19 pandemic is an advancing pandemic of the novel coronavirus which has been said to be caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). First detected in Wuhan, China, in December 2019, the outbreak has become a matter of earnest agitation with cases rising at an alarming rate in different parts of the world. The World Health Organization reported the outbreak as a Public Health Emergency of International Concern on January 30, 2019, and a pandemic on March 11, 2019. Globally, as of June 14, 2020, there have been 7,690,708 confirmed cases of COVID-19, including 427,630 deaths, reported in more than 188 countries and territories across the world [1]. The majority of the population tested positive with COVID-19 endure mild to moderate respiratory sickness and recover without requiring hospital treatment. However, older people and those with previous medical records like heart diseases, diabetes, chronic respiratory disease, and cancer are more likely to be severely ill and need to be hospitalized.

In view of the foregoing pandemic, it has become extremely important to document all the medical records of the COVID patients, thus providing a deeper insight into the disease and its cure. The clinical data is procured from several heterogeneous sources. The data may be highly structured like a patient's demographic information or his blood platelet count, as well as unstructured like the descriptive text of the patient's symptoms and the patient's radiology images. Moreover, the data obtained is sporadic and dynamic [2]. The health of some COVID patients deteriorates within one week of illness onset [3]. The constantly changing values of test results and the heterogeneity of data pose a challenge to the database management system currently being used.

One of the most commonly adopted systems worldwide for clinical data storage is the Relational Database Management System (RDBMS). The data in RDBMS is highly structured in the form of tables with relations incorporated among them, where SQL is used to communicate with the stored data. While there is a possibility to store some of the clinical data in the structured format, due to the sporadic nature, the relational model is not practical when the requirement of fields is high. This is because it will lead to empty fields resulting in insufficient storage [2]. However, it faces a drawback of rigidity for which the data should always be in the form of tables. Since it is evident from the recent searches and experiments that the n-COV data is heterogeneous in nature, a shift from the traditional storage in a relational database to an advanced non-relational database format becomes a necessity.

The effective management of clinical data and the transformation of the data into a structured format for data analysis are extremely challenging issues in electronic health records development [4], thus a solution with non-relational databases, NoSQL and XML database models, is proposed. Non-relational databases have been put forward as a flexible alternative to the traditional relational data models since they allow related data to be nested within a single data structure. In NoSQL database models, there is no need to store related data elements like tables. NoSQL allows a schemaless database design, and XML involves no transition of natural text data and is

also a subclass of document-oriented databases. NoSQL systems are the most suitable for query speed because their performance is efficient, and they are more scalable. Furthermore, XML technologies are beneficial in terms of flexibility, extensibility, and robust mark-up language to deal with the characteristics of clinical data and encode documents electronically such that the data can be delineated in a structured manner.

This paper proceeds as follows—Sect. 2 briefly explains the various technological research done in the COVID pandemic as well as the work done in the database management systems. Section 3 mentions the advantages of non-RDBMS over RDBMS in the COVID scenario. Section 4 deals with the clinical document standard set by the HL7 CDA. Sections 5 and 6 discuss the NoSQL and XML databases and their benefits over RDBMS. And finally, Sect. 7 provides implementation and Sect. 8 states the conclusion.

## 2 Literature Review

### 2.1 COVID-19 Pandemic

The novel coronavirus has been proved by the epidemiologists as a tricky illness due to its emergence in various forms, ranging from mild symptoms to high risk of organ failure or death [5]. Since n-COV is caused due to SARS, lung infection is one of the commonly observed symptoms detected using imaging techniques such as computed tomography (CT), positron emission tomography—CT (PET/CT), lung ultrasound, and magnetic resonance imaging (MRI) [6]. Furthermore, to accelerate the diagnosis and treatment process, advanced artificial intelligence (AI) methods using deep learning and bioinformatics approaches have been introduced [1]. Prompt actions like identifying high-risk individuals employing the Social Internet of Things (SIOT) with the relationship among mobile devices can be of great advantage to suppress the contagion [7]. Also, to forecast the number of infected cases, deaths, and recoveries, prediction models such as exponential smoothing (EN), least absolute shrinkage and selection operator (LASSO), and linear regression (LR) have been implemented [8]. In addition, with the help of social media, the Natural Language Process (NLP) method is utilized to unveil various issues regarding public opinion surrounding COVID-19 [9]. Surveys adopting ML approach prove that this pandemic has had effects on mental health, learning styles, and activities in countries like India due to a complete shutdown from March 24, 2020 [10]. According to our research to date (June 21, 2020), not much exploration has been performed on the storage of this enormous data engendered from n-COV.

## 2.2 Non-relational Databases

One primary problem, industries all across the world are facing due to the advancement and augmentation of technology is the storage and management of millions of data that is being produced at intervals of less than a nanosecond. Data is being engendered and processed more actively than ever. Data generation will continue to rise in volume in the mere future at an exponential rate [11]. NoSQL databases have been developed to deal with the extensive data along with proficiencies like adaptivity, join free queries, and to deliver flexibility with the help of no fixed schema [12]. NoSQL database systems have emerged parallel to the major Internet companies, such as Google, Amazon, and Facebook. These companies faced a major challenge in managing the copious amount of data that the traditional RDBMS could not deal with [13]. A variety of activities, consisting of experimental and prophetic analysis, ETL-style data transformation, and non-mission-critical OLTP (for instance, handling protracted or inter-organization transactions) are assisted by NoSQL databases [14]. In the current scenario, e-commerce has developed its application with forums like Magento, Zen Cart, Prestashop, Spree, etc. to run a booming online store. Customer e-commerce apps also use them and thus an extendable database is required for storing the data. Thus, e-commerce applications widely use NoSQL databases to handle extensive business applications by providing competent storage access and processing background with horizontal hierarchy and transferable strategy over RDBMS [15]. Oracle and MongoDB are the two prominent NoSQL databases. Oracle offers a better distribution model and is most preferred for implementing storage nodes intended to provide adaptability, accessibility then performing on warehouse and distribution models. Conversely, MongoDB depends on sharding for resizing and assigns a definite server to hold bits of data when data burgeons unremittingly [16].

## 3 Advantages of Non-RDBMS Over RDBMS

See Table 1.

**Table 1** Difference between RDBMS and non-RDBMS

RDBMS	Non-RDBMS
Table-oriented with fixed, predetermined, and restrictive schema	Document-oriented databases that are schemaless
Can be only scaled vertically which is limited by budget	Can be scaled horizontally to provide more resilience and lower costs
A very rigid schema and making regular changes is not feasible	It has no constraints and provides adaptability.
Can handle data coming in low velocity	Can handle data coming in high velocity

### 3.1 *Disadvantages of RDMS*

One of the main drawbacks of RDBMS is the cost of maintenance of complex software. In addition to the management of the high volume of data, it also possesses a property of rigidity for the formation of the schema. Lastly, the data processing hampers the speed, and it is difficult to recover the data.

## 4 Clinical Document Architecture

The collection of massive amounts of clinical data during COVID-19 is not enough; it has to be stored in an efficient format for future reference. The HL7 Version 3 Clinical Document Architecture (CDA) is a document mark-up standard that specifies the structure and semantics of “clinical documents” for exchange between healthcare providers and patients [17]. It is an XML document, consisting of a header and a body.

The CDA is beneficial because of its striking features of re-usability, flexibility, and conciseness. The clinical document contains data like pathology report, imaging report, symptom description, and alternative parts of a multimedia system—all integral components of electronic health records (EHRs) and have the following six characteristics, set forth by HL7:

- Persistence—A clinical document remains unaltered for a long period of time defined by local and regulatory requirements [17].
- Stewardship—A clinical document is maintained by someone or organization vouchsafed with its care [17].
- Potential for authentication—A clinical document is a collection of information that is intended to be legally attested [17].
- Context—A clinical document is a default context of the recorded data including the creator of the document, and patient’s identity, etc. [17].
- Wholeness—A clinical document can be authenticated as a whole and is not just restricted to certain parts of the document [17].
- Human Readability—A clinical document is easily read by humans or can be browsed on devices [17].

## 5 NoSQL

NoSQL, which stands for ‘Not Only SQL’, databases came into existence due to the limitations of the traditional relational database systems. Though they’ve been in existence for many years, they’ve recently gained popularity in the era of cloud and big data [12].

**Table 2** A table in NoSQL

Patient ID	Key	Value1	Subkey1	Subkey Value1	....
1	Cough	Yes	Influenza	A	...
2	Breathlessness	Yes	Pneumonia	25%	...
3	Fever	101°F	Duration	5 Days	...

It enables agile storage and pre-processing along with swiftness in utilization. Since these functionalities are very essential in the management of the COVID-19 data, it is one of the best systems to tackle unstructured, semi-structured, or structured data.

According to data models, NoSQL databases are classified as “key-value store,” “column-oriented store,” “document-oriented store” and “graph databases” [18].

- **Key-value store:** It is the most fundamental data model where data is stored as a key-value.
- **Column-oriented store:** In this data model, columnar manner is preferred over the traditional row manner.
- **Document store database:** It provides an efficacious way to administer document-oriented information in a semi-structured data format. It has a layer that manages the association between these documents.
- **Graph store database:** This type of data model records data in a graph structure to depict the relationship between data by warehousing data in the form of nodes, edges, and properties.

The following table depicts the general representation of data of COVID patients along with their commonly observed symptoms in NoSQL (Table 2).

Here, the dynamic nature of the sporadic symptoms is organized without being confined to a predefined structure. Thus, it helps in decreasing redundancy and turn improving efficiency.

### 5.1 Advantages of Using NoSQL for Storing COVID-19 Data

- (1) **Scalability:** The burgeoning number of COVID-19 cases and the concurrent increase in healthcare data demands an expandable EHR system as most of the current systems are based on relational databases that restrict scalability. The number of COVID-19 cases has increased exponentially and continues to rise. Thus, NoSQL database systems are critical as they allow scaling up to large datasets without any amendments in the comprehensive structure of data or architecture. Hardware requirements and expenses can develop rectilinearly as storage demands grow. Thus, in the time of economic crisis, cost-efficient scaling can be made possible, and preliminary investment in hardware requirements can be avoided by the already encumbered medical systems with the help of NoSQL.



Conventional relational database systems expand their capacity by acquiring more expensive and potent servers, whereas NoSQL database systems are based on a shared-nothing approach. In a shared-nothing architecture, servers have their resources and thus do not divide a common RAM, processor or warehouse. Thus, a large number of read/write operations can be made feasible with the help of horizontal scaling, dissemination of data and handling operations over many servers [19]. Hence, the capacity to store accurate data about a large number of patients can be elevated by the addition of more commodity servers dynamically without any reconfiguration or mitigation in performance.

- (2) Flexibility: On studying the humongous amount of data on n-Cov, it can be inferred that due to sporadicity and homogeneity, a rigid data storage system like RDBMS should be supplanted with a much more flexible model. To avoid redundancy, it compels the user to prioritize flexibility for the management of such disparate data. NoSQL databases provide pliability in the development of the schemas by avoiding the traditional table-like format for the data storage. Given that the COVID-19 data is voluminous and dynamic in nature, NoSQL proves to be efficient for quick iterations and frequent code pushes [12]. Due to the absence of a predefined structure, NoSQL helps to easily add and make changes with no need for any regard to the structure/schema of the database. Thus, it provides ad hoc schema changes that are often difficult and complex to carry out by using RDBMS [12].
- (3) High Functionality: COVID-19 data needs to be analyzed for the following reasons:
  - To comprehend accurate responses: With the right analytics capabilities, healthcare professionals can answer queries such as where the next cluster is most likely to arise, which demographic is most vulnerable, and how the virus may mutate over time.
  - To see the inconspicuous: Heterogeneous data from various sources has led to novel sharing of visualizations and messages to enlighten the public and to track the situation to understand its gravity.
  - To comprehend accurate responses: With the right analytics capabilities, healthcare professionals can answer queries such as where the next cluster is most likely to arise, which demographic is most vulnerable, and how the virus may mutate over time.
  - To see the inconspicuous: Heterogeneous data from various sources has led to novel sharing of visualizations and messages to enlighten the public and to track the situation to understand its gravity.

NoSQL databases offer many highly functional APIs and data types that are specially designed for each of their corresponding data models [20]. New application archetypes can be more easily supported with the help of NoSQL. The extensibility of NoSQL databases enables a single database to serve both transactional and analytical workloads from the same database as opposed to SQL databases that require a separate data warehouse to substantiate analytics. Since the NoSQL databases have

been developed during the era of cloud computing, they have accustomed themselves quickly to the automation that is part of the cloud. NoSQL also makes it easier to deploy databases extensively in a way that supports microservices.

NoSQL databases support polyglot persistence, which means combining various types of NoSQL databases depending on the requirements of a specific health-care system. For example, some hospitals storing most of their data in a document database like MongoDB, but supplement that with a graph database to seize innate relationships between patients and symptoms.

(4) **Security:** Data breaches are a major concern that needs to be taken into consideration while selecting a database system. A database system needs to be extremely secure and provide the four features of security—authorization, authentication, encryption, and auditing. MongoDB, a very popular non-relational database, provides these security features through the MongoDB Enterprise Advanced service. Advanced security controls like LDAP integration and AWS private link can be integrated with MongoDB Enterprise Advanced [21].

- **Authorization:** Access to the database by an entity is governed by MongoDB using the Role-Based Access Control (RBAC).
- **Authentication:** Authentication mechanisms like Kerberos and LDAP are supported by MongoDB for validation of entity.
- **Encryption:** Data can be encrypted while it is in transit over the network or at rest in storage and backups by the administrators.

**Auditing:** MongoDB Enterprise Advanced provides an auditing framework which can log all the actions (DDL and DML) and accesses made to the database.

(5) **Retrieval of archived data:** The emergence of big data technologies to handle gigantic volumes of structured and unstructured data all at low cost has right suited it to take the position of data archival solution. MongoDB is designed to establish long-term storage needs, an effective and prompt search of content with the help of keywords or full texts and cost-efficient services. For instance, COVID data produces a huge amount of content each day, X-ray images, symptom details, doctor’s comments and chat transcripts. Not only would such an institution produce such varied content, but it would also need to archive the content for long-term retention and serve as precedence. By leveraging MongoDB, the agility of the organization’s business can be benefited. The challenges associated with the velocity, volume and variety of data can be tracked down in a swift, elegant, and agile manner, thereby making MongoDB a scalable back-end data archival solution to such clinical data.

## 6 XML

XML databases are generally used to store information that is in varied forms. These databases are document-centric and are a subset of the NoSQL database. XML documents are marked by the heterogeneity of data records, extensibility by allowing different data types in a single document, larger grained data, and flexibility in size [22]. Similar characteristics are observed in the data obtained from COVID patients (Fig. 1).

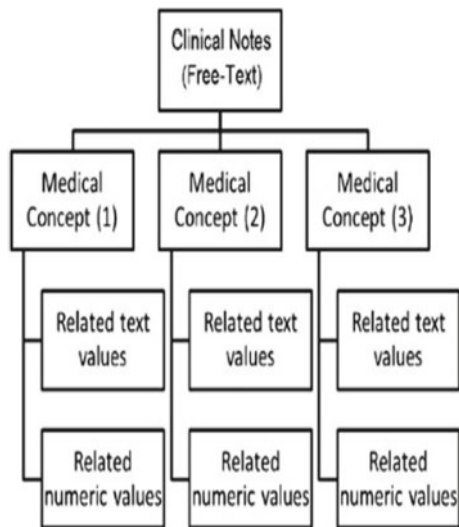
Raw data obtained is highly irregular and contains lots of mixed content. It consists of measures and values that are structured and narrative descriptions that are unstructured. It is crucial to gather all this information and convert it into a computer-processable structure for storage and processing. Data can be transformed into knowledge only if it is understandable by the machine and humans. Hence, the use of a tree data structure to represent this information is recommended.

As seen from the above figure, clinical notes in its raw form are situated at the absolute apex level, followed by the medical concepts. These consist of narrative text descriptions and their related numeric values. XML databases are broadly classified into two categories:

- XML-enabled database: Data is stored in tables consisting of rows and columns.
- Native XML database: Document-centric databases. Data is stored as a list of files.

XML databases are a relatively newer data storage technology and overcome certain drawbacks faced by the relational databases. The processing time required for the execution of queries is found to be a lot lesser in XML databases as compared

**Fig. 1** A general structure for clinical notes [23]



to RDBMS [23]. They support a hierarchical structure of data that allows a high level of granularity and is also flexible.

## 6.1 XML Document Architecture

The HL7 clinical document architecture is a standard approach for storing and exchanging various healthcare information. The header of the document is consistent across all clinical documents. The COVID-19 narrative data is the primary thing that is stored in the header. The CDA body indicates the human-readable content. The data accumulated from the corona tested patients with a patient's ID and their ages is incorporated within the header. From a patient's blood count of the red blood cells, plasma and platelets to the deficiency in each of the nutrients in the body are the human-readable narrative data blocks. These attributes encompass the whole body. This facilitates in satisfying CDA standards as it excludes encoding.

Flexibility is the most conspicuous as well as an astonishing feature of this document, where

- Each COVID patient record can have zero to many medical records concomitants to them.
- Each medical record can have zero to many symptoms correlated.
- Each symptom can comprise zero or more properties which can be textual as well as numeric.

Dealing with XML as large strings is ineffectual, native XML is customized for its storage and querying. Native XML databases are specially adapted to the XML data. They are highly capable of storing, maintaining, and querying the XML document. Relational databases fail to espouse such adaptability and even cannot provide an efficacious way of handling, storing, recovering, and analyzing such fitful data. Consequently, XML databases unlock a remarkable path for treating medical data that is voluminous and sporadic (Fig. 2).

## 7 Implementation

Scenario: Consider the following contrived dataset consisting of patients and their various COVID-19 attributes. Due to the sporadic nature of this data, not every single patient will have all the attributes populated for him/her. It is represented as follows in the RDBMS:

As it is visible, for example, patient "04235e8a80d92ed," the attribute influenza B is null. Due to the rigid structure of RDBMS, that particular column cannot be ignored for the patient which makes the database bulky, inefficient, and full of unwanted null values. Moreover, considering irregular addition and reduction of different attributes,

```

HEADER
<Clinical Document>
  <id><code><title><recordTarget>
  <patient>

BODY
<structuredBody>
  <section>
    <code>
      <title>Covid Symptoms</title>
      <text>Dry Cough</text>
      <entry>
        <observation><code:value,system_name>
          <classcode><moodcode>
            <statusCode>
              <effectiveTime>
                <value>
      <text>High Fever</text>
      <entry>
        <observation><code:value,system_name>
          <classcode><moodcode>
            <statusCode>
              <effectiveTime>
                <value>
        .
        .
        .
    
```

Fig. 2 Example of CDA document [24]

a constant usage of Data Definition Language (DDL) is needed to alter the structure of database schema. This makes the maintenance of the database complex.

Therefore, NoSQL is being proposed, a non-structured/schemaless, fluid and a flexible database management system. Using NoSQL, you can simply add attributes, delete attributes and scope in the attributes on an as-need basis, thus providing a superior and efficient method to handle sporadic databases.

Following are the points that show how NoSQL has helped in storing the patient data displayed in Fig. 3 (Fig. 4).

- **Scalability:** The nature of symptoms of the COVID patients being admitted is dynamic. Since NoSQL is schemaless, it is easy to add new attributes (symptoms) to the database in contrast to RDBMS where the entire schema would have to be revised as shown in Figs. 5 and 6.
- **Flexibility:** Flexibility is a boon when NoSQL is considered. With the key-value pair and the document-based NoSQL, it is an approach that provides flexible storing and retrieving of the data. Null/Unrelated values take up a lot of space in

Patient ID	Patient age quantile	SARS-Cov-2 exam result	Patient admitted to regular ward (1=yes, 0=no)	Platelets	Red blood Cells	Influenza B	CoronavirusNL63	CoronavirusOC43	Influenza B, rapid test	Influenza A, rapid test
d7834ed75f2da44	16	positive	1			not_detected	not_detected	not_detected		
d3adcb9067a5f4	3	negative	0	0.248857364	-1.608140349	not_detected	not_detected	not_detected		
04235e8a80d92ed	12	negative	0	0.198610127	0.472241729	not_detected	detected	not_detected		
33729cd2658ca64	15	positive	0							
3fc734fd99bafbd6	18	positive	1							
e586524f71166bc	18	negative	0	-0.45460397	-0.356384665	not_detected	detected	not_detected		
ff342c79fc6f0	7	positive	1						negative	negative

Fig. 3 COVID-19 patient data in RDBMS

```

Command Prompt - mongo
{
  "_id": ObjectId("5f1269e10e34333fd38f0ed1"),
  "Patient ID": "d3adcb9067a5f4",
  "Patient age quantile": "3",
  "SARS-Cov-2 exam result": "negative",
  "Patient admitted to regular ward (1=yes, 0=no)": "0",
  "Platelets": "0.248857364",
  "Red blood Cells": "-1.608140349",
  "Influenza B": "not_detected",
  "CoronavirusNL63": "not_detected",
  "CoronavirusOC43": "not_detected"
}

{
  "_id": ObjectId("5f1269e10e34333fd38f0ed2"),
  "Patient ID": "04235e8a80d92ed",
  "Patient age quantile": "12",
  "SARS-Cov-2 exam result": "negative",
  "Patient admitted to regular ward (1=yes, 0=no)": "0",
  "Platelets": "0.198610127",
  "Red blood Cells": "0.472241729",
  "Influenza B": "not_detected",
  "CoronavirusNL63": "detected",
  "CoronavirusOC43": "not_detected"
}

{
  "_id": ObjectId("5f1269e10e34333fd38f0ed3"),
  "Patient ID": "d3729cd2658ca64",
  "Patient age quantile": "15",
  "SARS-Cov-2 exam result": "positive",
  "Patient admitted to regular ward (1=yes, 0=no)": "0"
}
    
```

**Blank and Null values can be disregarded due to schema less nature of the databases**

Fig. 4 COVID-19 patient data in MongoDB

Patient ID	Patient age quantile	SARS-Cov-2 exam result	Patient admitted to regular ward (1=yes, 0=no)	Platelets	Red blood Cells	Influenza B	CoronavirusNL63	CoronavirusOC43	Influenza B, rapid test	Influenza A, rapid test	Dry Cough	Diarrhoea
d7834ed75f2da44	16	positive	1			not_detected	not_detected	not_detected			Yes	Yes
d3adcb9067a5f4	3	negative	0	0.248857364	-1.608140349	not_detected	not_detected	not_detected				
04235e8a80d92ed	12	negative	0	0.198610127	0.472241729	not_detected	detected	not_detected				
d3729cd2658ca64	15	positive	0									
3fc734fd99bafbd6	18	positive	1								Yes	
e586524f71166bc	18	negative	0	-0.45460397	-0.356384665	not_detected	detected	not_detected				
ff342c79fc6f0	7	positive	1						negative	negative		

Fig. 5 Updating relational table with new attributes (symptoms)

```

> db.patients.find().pretty()
{
  "_id" : ObjectId("5f1269e10e34333fd38f0ed0"),
  "Patient ID" : "d7834ed75f2da44",
  "Patient age quantile" : "16",
  "SARS-Cov-2 exam result" : "positive",
  "Patient admitted to regular ward (1=yes, 0=no)" : "1",
  "Influenza B" : "not_detected",
  "CoronavirusNL63" : "not_detected",
  "CoronavirusOC43" : "not_detected",
  "Diarrhoea" : "Yes",
  "Dry Cough" : "Yes",
  "lastModified" : ISODate("2020-07-19T12:53:13.603Z")
}
{
  "_id" : ObjectId("5f1269e10e34333fd38f0ed1"),
  "Patient ID" : "d3adcbd9067a5f4",
  "Patient age quantile" : "3",
  "SARS-Cov-2 exam result" : "negative",
  "Patient admitted to regular ward (1=yes, 0=no)" : "0",
  "Platelets" : "0.248857364",
  "Red blood Cells" : "-1.608140349",
  "Influenza B" : "not_detected",
  "CoronavirusNL63" : "not_detected",
  "CoronavirusOC43" : "not_detected"
}
    
```

Fig. 6 Updating NoSQL with new attributes (symptoms)

RDBMS, however, with NoSQL having the option of storing only the values of attributes related to the record.

- Unstructured data storage: Every single patient has different attribute assigned with it which classifies data. Rigid traditional RDBMS is not an efficient way to store the data as shown above. On the contrary, schemaless database allows the highly customized association of each patient and its idiosyncratic attribute, thus making storage of unstructured data highly efficient.
- Ease of access: As each primary key, i.e., patient ID has its customized attribute uniquely connected and NoSQL facilitates such intertwining of the individual entity and attribute. Every single patient record can be considered as a unique entity, and its personalized attribute can be retrieved on an as-need basis thus amplifying ease of access.

## 8 Conclusion

Over many years, relational databases have been providing noteworthy results in large enterprise scenarios due to its simplicity, compatibility, and robustness in handling generic data. The COVID data being dynamic and ever-changing enhances the complexities of a traditional database system. The speed, cost, and potential



of the non-relational databases, like NoSQL and XML, finely capture data that are unstructured and semi-structured too. In this paper, the two augmenting databases are highlighted that are worth considering as an alternative to the traditional one.

The inclusion of HL7 represents a structured format of the COVID record of patients for targeting the exchange of healthcare data. Having an XML base, CDA is extremely flexible and clearly distinguishes the COVID-related data from patient's data.

The key-value pair storage of NoSQL is a convenient way of storing and managing complex COVID data. NoSQL being a schemaless database has many advantages. Moreover, the persistence of XML databases provides security for XML-formatted documents, and the extensibility characteristic of XML is beneficial for storing the COVID data.

NoSQL and XML are the emerging technologies that surmount certain limitations of RDBMS. Though it may require considerable efforts to replace the current database systems used in health care with the newer ones, a gradual shift or hybrid of the two can be employed to make the best of both the systems. The COVID pandemic has been a wake-up call to all the healthcare organizations to reconsider their data storage infrastructure because of the large influx of data that the world is being bombarded with. It may be time to venture into newer technologies to increase the efficiency of the database systems and provide the best possible medical care to the patients.

## References

1. Coronavirus disease 2019—World Health ..... <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed 16 June 2020
2. Bageshwari M, Adurkar P, Chandrakar A (2014) Clinical database: Rdbms v/s newer technologies (NoSQL and XML database); Why look beyond Rdbms and consider the newer. *Int J Comput Eng Technol (IJCET)* 5(3):73–83
3. Management of Patients with Confirmed 2019-nCoV CDC. 2 June 2020. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html>. Accessed 15 June 2020
4. Lee KK, Tang W, Choi K (2012) Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage. *Comput Methods Programs Biomed* 110. <https://doi.org/10.1016/j.cmpb.2012.10.018>
5. Jamshidi MB et al (2020) Artificial Intelligence and COVID-19: deep learning approaches for diagnosis and treatment. *IEEE Access*. <https://doi.org/10.1109/access.2020.3001973>
6. Dong D et al (2020) The role of imaging in the detection and management of COVID-19: a review. *IEEE Rev Biomed Eng*. <https://doi.org/10.1109/rbme.2020.2990959>
7. Wang B, Sun B, Duong TQ, Nguyen LD, Hanzo L (2020) Risk-aware identification of highly suspected COVID-19 cases in social IoT: a joint graph theory and reinforcement learning approach. *IEEE Access*. <https://doi.org/10.1109/access.2020.3003750>
8. Rustam F et al (2020) COVID-19 future forecasting using supervised machine learning models. *IEEE Access* 8:101489–101499. <https://doi.org/10.1109/ACCESS.2020.2997311>
9. Jelodar H, Wang Y, Orji R, Huang H (2020) Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J Biomed Health Inf*. <https://doi.org/10.1109/jbhi.2020.3001216>



10. Khattar A, Jain PR, Quadri SMK (2020) Effects of the disastrous pandemic COVID 19 on learning styles, activities and mental health of young indian students—a machine learning approach. In: 2020 4th international conference on intelligent computing and control systems (ICICCS), Madurai, India, pp 1190–1195. <https://doi.org/10.1109/iciccs48265.2020.9120955>
11. Kalid S, Syed A, Mohammad A, Halgamuge MN (2017) Big-data NoSQL databases: a comparison and analysis of “Big-Table”, “DynamoDB”, and “Cassandra”. In: 2017 IEEE 2nd international conference on big data analysis (ICBDA), Beijing, pp 89–93. <https://doi.org/10.1109/icbda.2017.8078782>
12. Sharma M, Sharma VD, Bunde MM (2018) Performance analysis of RDBMS and No SQL Databases: PostgreSQL, MongoDB and Neo4j. In: 2018 3rd international conference and workshops on recent advances and innovations in engineering (ICRAIE), Jaipur, India, pp 1–5. 10.1109/ICRAIE.2018.8710439
13. [online] Available <https://en.wikipedia.org/wiki/NoSQL>
14. Moniruzzaman ABM, Hossain S (2013) NoSQL Database: new era of databases for big data analytics—classification, characteristics, and comparison. *Int J Database Theor Appl* 6
15. Ramesh D, Khosla E, Bhukya SN (2016) Inclusion of e-commerce workflow with NoSQL DBMS: MongoDB document store. In: 2016 IEEE international conference on computational intelligence and computing research (ICCIC), Chennai, pp 1–5. <https://doi.org/10.1109/iccic.2016.7919652>
16. Padhy S, Kumaran GMM (2019) A quantitative performance analysis between MongoDB and Oracle NoSQL. In: 2019 6th international conference on computing for sustainable global development (INDIACom), New Delhi, India, pp 387–391
17. HL7 Standards Product Brief-CDA® Release 2 ... - HL7.org. [https://www.hl7.org/implementation/standards/product\\_brief.cfm?product\\_id=7](https://www.hl7.org/implementation/standards/product_brief.cfm?product_id=7). Accessed 22 June 2020
18. Ercan M, Lane M (2014) An evaluation of NoSQL databases for electronic health record systems
19. [online] Available <https://www.ibm.com/cloud/learn/nosql-databases>
20. What is NoSQL? Amazon Web Services. <https://aws.amazon.com/nosql/>. Accessed 17 July 2020
21. MongoDB Atlas Security. <https://www.mongodb.com/cloud/atlas/security>. Accessed 26 July 2020
22. Saba A, Shahab E, Abdolrahimpour H, Hakimi M, Moazzam A (2017) A comparative analysis of XML documents, XML Enabled Databases and Native XML Databases
23. Leea K, Tangb W, Choia K (2012) Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage” Centre for Integrative Digital Health, School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, Information Technology Committee, Hong Kong Doctors Union, Hong Kong, 31 October 2012
24. Clinical Document Architecture (CDA)-Office of the National .... 28 July 2010. [https://www.healthit.gov/sites/default/files/resources/cda\\_c-cda\\_theirrole\\_in\\_mu.pdf](https://www.healthit.gov/sites/default/files/resources/cda_c-cda_theirrole_in_mu.pdf). Accessed 17 July 2020

# Study of Effective Mining Algorithms for Frequent Itemsets



P. P. Jashma Suresh, U. Dinesh Acharya, and N. V. Subba Reddy

**Abstract** “Frequent Itemset Mining” is a domain where several techniques have been proposed in recent years. The most common techniques are tree-based, list-based, or hybrid approaches. Although each of these approaches was proposed with the intent of mining frequent itemsets efficiently, as the number of transactions increases, the performance of most of these algorithms gradually declines either in terms of time or memory. In addition, the presence of redundant itemsets is another crucial problem where a limited investigation has been carried out in recent years. There is thus a pressing need to develop more efficient algorithms that will address each of these concerns. This paper aims to survey the different approaches highlighting the advantages and disadvantages of each of them so that in future effective algorithms may be designed for extracting frequent items while addressing each of these concerns effectively.

**Keywords** Frequent itemset mining · Tree · List · Hybrid · Redundancy

## 1 Introduction

Finding items that frequently occur together is the essence of “Frequent Itemset Mining”. This process is carried out by scanning the entire database and calculating the support of individual itemsets. Several algorithms have been designed for this

---

P. P. Jashma Suresh (✉) · U. Dinesh Acharya · N. V. Subba Reddy  
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal  
Academy of Higher Education, Manipal, Karnataka 576104, India  
e-mail: [jashma.suresh@manipal.edu](mailto:jashma.suresh@manipal.edu)

U. Dinesh Acharya  
e-mail: [dinesh.acharya@manipal.edu](mailto:dinesh.acharya@manipal.edu)

N. V. Subba Reddy  
e-mail: [nvs.reddy@manipal.edu](mailto:nvs.reddy@manipal.edu)

purpose. These can be applied to various disciplines. These include market basket analysis, medicine, fraud detection, intrusion detection, web content mining, and biomedical applications.

In the field of market basket analysis, mining frequent itemsets can be used to find items that are purchased frequently together. This will help retailers in deciding how the items have to be shelved in the grocery store. Apart from this, “Frequent Itemset Mining” (FIM) can also be used to detect “outliers” which in turn can be applied to detect cancerous cells in medicine and detect fraudulent behavior of people while making financial transactions. Web content mining and web usage mining are other areas where FIM can be applied. Analyzing the patterns in web logs aids in understanding the browsing habits of a particular individual. This can help in the extrication of semantic information about the user and a given community which can be used by companies for personalized marketing.

Although considerable work was done in the past concerning designing algorithms for mining different types of frequent items (FI), their performance tends to decline as the mining progresses. Identifying the appropriate data structure and designing the appropriate framework for carrying out parallel mining on a multi-core processor remains as an area where there is scope for further study to be carried out.

While mining frequent itemsets, there is usually a trade-off between memory and performance. These are the two main metrics for mining frequent itemsets. Most of the existing techniques either result in good run time but may not provide satisfactory improvement in terms of memory consumption. Designing an algorithm that finds a balance between the two is a critical problem that needs to be addressed. This also necessitates the exploration of ensembling and hybrid approaches for the enhancement of frequent itemsets. Apart from this, another metric that determines the efficiency of FIM is the number of itemsets generated. This number may also include some redundant itemsets and removing such items is another area where there is scope for further research to be carried out.

There is also limited investigation in terms of the nature of the datasets used for the extrication of frequent itemsets. A data structure that can be used for handling static data need not be appropriate for the mining of data from a streaming environment. Developing a data structure that incrementally mines exact frequent patterns from a streaming environment with negligible loss of information remains a challenging task.

This paper aims to provide an insight into the recent techniques that have been developed for mining frequent items with an emphasis on the advantages and disadvantages of each of the proposed approaches.

## 2 Frequent Itemset Mining

There are several approaches for mining frequent itemsets. This study classifies it into mainly three categories based on the data structure used for frequent itemset mining. They are—Tree, list, and hybrid approaches (Fig. 1).

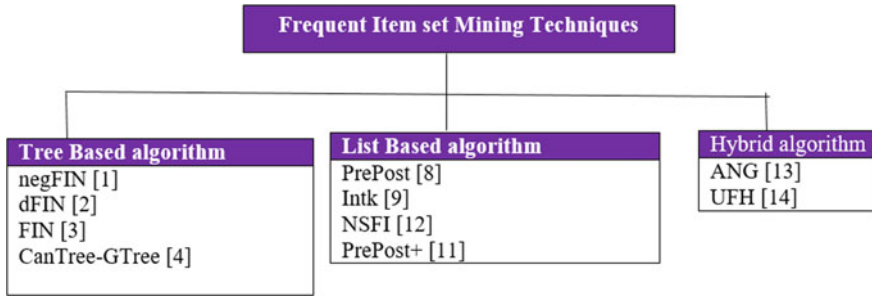


Fig. 1 Frequent itemset mining techniques

## 2.1 Tree-Based Frequent Itemset Mining

Tree-based algorithms scan the database to build a tree data structure. A tree-based algorithm “negfin” was put forward by Aryabarzan et al. [1] for mining “Frequent Itemsets.” A data structure “NegNodeset” was developed in this paper. This relies on the construction and traversal of a prefix tree. All the nodes are encoded in bitmap form resulting in a concise representation of itemsets.

An algorithm dFIN [2] was also developed that used a data-structure DiffNodeset for mining “Frequent Itemsets.” This algorithm involves building and scanning of a “set enumeration tree” and also adopted a “hybrid search paradigm” for finding “frequent itemsets.” In some scenarios, this approach also had the pros of not having to find out “candidate itemsets.” This algorithm is an advancement of the algorithm FIN [3] that employed “pre-order” and “post-order” traversal methods. This circumvented overconsumption of memory.

A tree-based algorithm called “CanTree-GTree” was presented by Kim and Hwang [4] The algorithm discovers the entire collection of frequent itemsets from real-time transactions using the “sliding window” mechanism. Two data structures are used in this algorithm—“CanTree” and “GTree.” “CanTree” depicts the transactions in a condensed manner with the help of a sliding-window technique using one scan. Another data-structure GTree has been proposed to use as a projection tree for finding frequent itemsets. Meng and Sha [5] presented a novel tree-based approach that will evaluate the influence factors that affect the “life satisfaction” and “loneliness of retired athletes.” An extended prefix tree is constructed for mining frequent itemsets. An efficient data-structure called SO-Sets based on SO-Tree was designed by Tan and Qin [6]. An algorithm called FISO has been put forward to mine “frequent itemsets” by making use of the “set enumeration tree structure.” It applies the conceptualization of superset equivalence to minimize the search space. A broader analysis has shown that FISO outshines “Prepost” and “FIN” with respect to “time” and “memory.”

Khawaja et al. [7] presented a “Tree Data Structure (TDS)” for cumulating “candidate itemsets” for pruning. A multi-core framework has been proposed for the running “TDS-based Apriori Algorithm.” The principle of divide and conquer has

been used on this framework. Each core is assigned a separate task and all these cores work in parallel. Each of these cores produces a subset result, and these results are collaborated to obtain the final global result.

A framework SUSHI [8] was proposed by Francia et al. that mines multi-level and multi-dimensional frequent items. This framework involves using an integration of tree and graph-based approaches that emphasize the relationship between two clusters. Shah and Halim [9] proposed a methodology for “mining Frequent Itemsets” from large uncertain databases. The proposed method uses a 3D linked array that traverses the database once and also employs a “tree-based method” that stores the data pertaining to the support and association of itemsets. In addition, it also calculates the “average probability factor” and uses it to map the “uncertain database” to a tree.

## ***2.2 List-Based Frequent Itemset Mining***

List-based data structures involve the usage of vertical mining technique for the extrication of frequent items. The entire database is scanned, and items are inserted into the list structure. Deng et al. [10] proposed a “vertical format” data representation called “N-list.” This structure relies on the building and scanning of a “PPC Tree.” An algorithm “PrePost” was developed that made use of this tree structure. Huynh-Thi-Le et al. [11] presented the iNTK algorithm for extracting “top-rank-k frequent patterns.” This approach is an advanced variant of the “NTK algorithm” and uses an “N-list structure” to characterize patterns. The mining time of “top rank-k patterns” has been improvised by using the subsume concept.

Bui et al. [12] presented a method that extracts “Frequent Weighted Itemsets” using a “Weighted N-list structure.” The method uses a few theorems to compute the weighted support of the “itemsets” and to find out these values immediately without having to use “WN-list intersection” in some specific cases. Deng et al. enhanced the “PrePost” algorithm by incorporating a pruning strategy called “Child Parent Equivalence testing” and designed the “PrePost+” [13] algorithm. The same “N-list” data structure was used here along with a “set enumeration tree” for mining “frequent itemsets.”

Vo et al. [14] designed an algorithm that is based on the concept of “N-list” and “Subsume Concept” for extracting “Frequent Itemsets.” The algorithm put forward makes use of a hash table. This provides a boost to the procedure of building N-list associated with “1-itemsets” and also produces an enhanced “N-list intersection algorithm.”

## ***2.3 Hybrid Frequent Itemset Mining***

While mining “frequent itemsets” relying on a single “data structure” may be detrimental to the performance of the mining process. Some data structures perform well

on low values of threshold while some others may perform well on higher values of threshold. So, using a combination of data structures can help in enhancing the overall efficacy of “mining frequent itemsets.” This necessitates the need for hybrid approaches. Using a combination of two or more approaches will exploit the benefit of all of them thereby enhancing the mining activity.

Zhang et al. [15] presented a hybrid technique that combines Apriori and Graph Computing for extracting “frequent itemsets.” The “Apriori algorithm” is employed for the computation of the support of  $k$ -item set candidates when  $k$  is small. As the value of  $k$  increases and becomes very large, the graph computation method is employed to calculate the support of  $k$ -item sets.

Another hybrid framework was put forward by Dewar et al. [16]. This structure uses a combination of “tree-based” and “inverted list algorithm” for mining “high-utility itemsets” from database transactions. The proposed methodology combines UPGrowth+ and FHM algorithm and experimental analysis prove that this technique outperforms contemporary algorithms.

Kalpna and Natarajan [17] proposed an approach that employs two algorithms Hybrid Miner I and Hybrid Miner II that operates on a vertical database for effective mining of frequent itemsets. Hybrid Miner I is a bottom-up algorithm for identifying “maximal frequent itemsets”. An outline of the advantages and disadvantages of some of these algorithms have been highlighted (Table 1).

Although various methods have been proposed of late for “mining frequent itemsets,” there has been limited investigation done in the type of itemsets that are mined. Some of the most commonly mined itemsets include “closed,” “non-derivable,” “closed and non-derivable,” and “maximal itemsets.” These generate a concise set of patterns that can help in avoiding redundancy among the itemsets. This is elaborated in the next section.

### 3 Concise Mining of Frequent Itemsets

While extracting “Frequent Itemsets” setting a very “low threshold” would produce a voluminous collection of “itemsets” that are too massive for user interpretation. To address this challenge, many approaches were presented to generate a “compressed” and “lossless representation of itemsets.” This section gives an overview of the techniques that have been adopted for the extraction of “concise frequent itemsets” (Fig. 2).

#### 3.1 Non-Derivable Frequent Itemset Mining

Mining frequent itemsets may sometimes result in “itemsets” whose “support” can be extrapolated from that of their “subsets” by using deduction rules. Such item sets are called derivable itemsets and result in redundancy of the dataset. The following

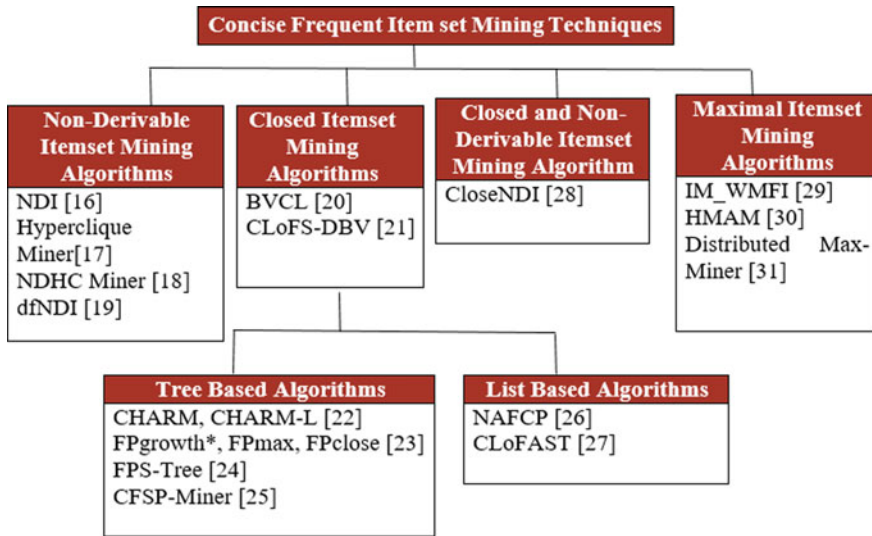
**Table 1** Comparison of “Frequent Itemset Mining Algorithms”

Algorithm	Advantages	Disadvantages
3D Array PFI, Tree-based PFI, average probability PFI [9]	<ul style="list-style-type: none"> <li>• Mines FI from uncertain databases</li> <li>• Consumes lesser time than MB, FIN, Prepost+, and UApriori</li> </ul>	<ul style="list-style-type: none"> <li>• Memory consumption was more on some datasets</li> <li>• A limited investigation was done with respect to incremental databases. Parallelization and constrained dependent FIM</li> </ul>
negfin [1]	Has got a time complexity of $O(ln)$ , where “l is the number of generated nodes,” “n. is the cardinality of the base set of nodes”	<ul style="list-style-type: none"> <li>• No improvement in memory consumption</li> <li>• Not tried on closed, maximal, erasable itemsets</li> </ul>
dFIN [2]	<ul style="list-style-type: none"> <li>• Size of DiffNodeset is much smaller</li> <li>• It has a faster runtime</li> </ul>	Not explored on the concise generation of itemsets
FIN [3]	<ul style="list-style-type: none"> <li>• Outperforms FP-growth, PrePost with respect to time</li> <li>• Consumes less memory</li> </ul>	• Does not generate non redundant itemsets
CanTree-GTree [4]	• Requires only one database scan	Needs more memory in some datasets
PrePost [10]	• Uses <i>N</i> - list intersection and subsequently has better runtime than FP-growth, Eclat, dEclat	<ul style="list-style-type: none"> <li>• Consumes more memory</li> <li>• Has not been tried on compressed mining of frequent itemsets</li> </ul>
iNTK [11]	• Uses less memory than NTK	• Not effective for sparse datasets
NSFI [14]	• Has better runtime than PrePost	<ul style="list-style-type: none"> <li>• Less improvement in memory consumption</li> <li>• Not effective on sparse datasets</li> <li>• Not tried on closed/maximal itemsets mining</li> </ul>
PrePost+ [13]	<ul style="list-style-type: none"> <li>• Runs faster than PrePost, FIN, FP-growth* since it uses “N-list intersection” and “child parent equivalence pruning technique”</li> <li>• Has lesser memory consumption than PrePost, FIN</li> </ul>	<ul style="list-style-type: none"> <li>• Consumes more memory than FP-growth*</li> <li>• Not explored with respect to mining HUIs</li> </ul>
ANG [15]	• Runs faster than Apriori, Graph Computing algorithm used alone	• Performs well as a hybrid technique. However, Apriori performs well only when “k” is small, and graph computing performs well only when “k” is large

(continued)

**Table 1** (continued)

Algorithm	Advantages	Disadvantages
UFH [16]	<ul style="list-style-type: none"> <li>Performs better than FHM and EFIM on sparse datasets</li> </ul>	<ul style="list-style-type: none"> <li>Does not show good performance on dense datasets</li> </ul>



**Fig. 2** Concise frequent itemset mining techniques

section gives an outline of the different techniques adopted for the extraction of “non-derivable frequent itemsets.”

A technique that relies on “deduction rules” was presented by Calder’s and Goethals [18] to discard “itemsets” whose support can be deduced from their “subsets.” This generates a set of “itemsets” called “Non-Derivable Itemsets.” Itemsets whose “lower bound” equals the “upper bound” are “derivable.” The proposed algorithm aims to generate a set of “lossless” and “concise” representation of itemsets called “Non-Derivable Itemsets.”

A framework for extracting highly co-related patterns has been presented by Xiong et al. [19]. A pruning method that combines Hypercliques and non-derivable itemsets has been put forward by Koufakou [20] to generate a collection of non-derivable hypercliques. An algorithm called NDHCMiner that makes use of a single scan has been employed for the generation of NDHC sets.

Calders [21] proposed an algorithm called dfNDI. The algorithm is based on the concept of Eclat and mines “non-derivable itemsets.” The presented algorithm has the advantage that it has lesser costly candidate generation and refrains from scanning the database repeatedly. It also incorporates the concept of item re-ordering to abstain from the production of too many candidates. With the help of experiments, it has



been shown that the presented approach outperforms the NDI algorithm by allowing fast discovery of non-derivable frequent itemsets.

### 3.2 *Closed Frequent Itemset Mining*

An item set is “closed” if there exists no superset that has the same support count as that of the actual item set. Like mining non-derivable item sets, mining closed itemsets can also result in a reduced collection of item sets. An outline of the different approaches that have been adopted for the mining of closed itemsets is described as follows.

Hashem et al. [22] proposed a new data structure that makes use of “dynamic bit vector” and “dynamic superset bit vector” that infers relationships between “frequently closed itemsets” in a “lattice structure.” This had the advantage of minimizing the “search space” and to evict the “non-closed itemset generators.”

Recent studies indicate that mining “closed sequences” is more advantageous than mining “all frequent itemsets.” Tran et al. [23] proposed an algorithm called “CLoFS-DBV” that employs “dynamic bit vectors.” There are two steps to this algorithm. In the first stage, the “actual database sequence” is remodeled into a vertical-based format called “CLoFS-DBVPattern.” In the second stage, “frequent closed sequences” are extracted.

Zaki and Hsiao [24] employed a dual-itemset tid-set search tree and designed a charm algorithm for mining “frequent closed itemsets.” A hybrid strategy that eludes various layers is developed here to extract “closed itemsets.” This had the advantage of being able to produce “closed itemsets” without having to produce “non-closed itemsets.” As the computation progresses a vertical format that relies on “diffsets” was used to hold the “differences in tids” for visualizing the rule generation process an algorithm “Charm-L” was also put forward.

A collaboration of “FP-Tree” and “FP Array” was presented by Gosta and Zhu [25] to improve the performance of “FP-tree-based algorithms.” “Algorithms” have been presented for the mining of “all,” “closed,” and “maximal frequent itemsets.” The proposed approach shows adequate performance on “sparse datasets.” In addition, various techniques for “optimization” have been put forward to enhance the performance of the proposed approach.

ur Rehman et al. [26] proposed an algorithm that uses FPS-Tree to mine “top-k closed frequent itemsets.” This technique incorporates the sliding window approach. The “frequent closed itemsets” are generated with user-defined “minimum” and “maximum lengths” using a bitmap-based data-structure. This eludes the requirement of scanning the database multiple times.

A tree-based “CFSP-miner” algorithm was put forward by Rodríguez et al. [27] for extracting “Closed Frequent Similar Patterns” for producing a “condensed” and “lossless” set of “frequent similar patterns.”

Studies from the literature indicate that although several techniques have been adopted for mining “closed frequent itemsets,” the mined collection of itemsets may

still have some redundant information. Eliminating such redundancy will help in generating a concise representation of itemsets which will result in lesser memory requirements. Moreover, there is also room to improvise further the efficiency of these algorithms to speed up the mining operation.

Le and Vo [28] presented an “N-list-based algorithm” for mining “Frequent Closed Patterns” (FCP) called NAFCP. Two theorems were introduced for finding out FCPs based on “N-list structure.” This structure produces a more condensed representation in comparison to vertical structures, thereby minimizing the memory used and the time needed for mining to complete.

An algorithm called CLOFAST was presented by Fumarola et al. [29] for mining “closed frequent sequences” of itemsets. It combines “sparse id list” and “vertical id list” to obtain the support of “sequential patterns.” A “one-step technique” is employed to examine “sequence closure” and to “prune” the “search space.” The initial step involves the mining of all “closed frequent itemsets” to generate an initial collection of sequences of size 1. Then, new sequences are extracted without having to extricate additional frequent itemsets.

A parallel row enumerated algorithm was proposed by Vanahalli and Patil [30] for mining colossal closed itemsets. This algorithm uses an “efficient pruning” technique for efficiently reducing the “search space” and was designed to mine the colossal items from high dimensional datasets. This work was further extended in [31] to work on high dimensional datasets by incorporating an efficient pruning technique to reduce the “search space” along with “closure checking” for identifying closed itemsets among rowsets. This work can further be extended to mine colossal closed itemsets in a parallel and distributed environment as well.

Li et al. [32] constructed a topology-transaction tree for mining frequent closed itemsets. The proposed approach involves the analysis of the “binary relation” on the set of itemsets and demonstrates the co-occurrence relation between itemsets. This helped to deduce the associative relationship among itemsets. Although this algorithm performs well, there has been limited investigation carried out in parallelizing this algorithm.

### ***3.3 Closed and Non-Derivable Frequent Itemset Mining***

Mining of “Frequent itemsets” typically produces a large number of itemsets. Several techniques have been adopted for generating a condensed and lossless rendition of itemsets such as mining “closed,” “non-derivable itemsets.” Muhonen and Toivonen [33] proposed a “Closed Non-Derivable Itemset mining algorithm” that merges both the concepts of “closed” and “non-derivable itemset” mining to produce a “concise” and “lossless” representation of frequent itemsets. With the help of experimental study, it has been proven that the proposed algorithm generates a “condensed representation of itemsets” that will never exceed the smaller of the two. This collection is considered to be “truly lossless” since the complete set of “frequent patterns” can be retrieved from this set.

### 3.4 Maximal Frequent Itemset Mining

An itemset is “maximal” if none of its supersets is frequent. Mining maximal frequent itemsets are another commonly used approach to minimize the size of the mined itemsets.

Incremental mining often involves a large amount of data and interpreting the entire collection is a cumbersome task. This resulted in a dire need to mine “maximal frequent itemsets” from “data streams.” Yun and Lee [34] proposed an algorithm for mining “weighted maximal frequent itemsets” from “incremental databases.” A tree structure called “IM\_WMFI” tree along with the SC tree was used for carrying out the mining of frequent itemsets.

Liu et al. [35] proposed an approach for handling the problems associated with big data processing by mining maximal itemsets. The contemporary work in big data processing had issues such as redundancy, increased time complexity and require voluminous space for storage. To handle these issues, a novel approach called “Heuristic MapReduce-based Association rule” approach through “Maximal Frequent Itemset mining” was put forward.

To deal with the challenges involved in Bitmap Join Indexing Necir and Drias [36] put forward a method that comprises of two phases. The first phase involves pruning the “search space” to minimize the number of candidates produced. To achieve this, a “distributed maximal itemset mining approach” based on “multi-agent system” was employed. The second phase involves the selection of an index that reduces the total response time of queries to the storage constraint.

Chung and Luo [37] proposed two parallel algorithms for mining “maximal frequent itemsets” from the database. One called “Distributed Max-Miner” and another called “Parallel Max-Miner.” Nguyen et al. [38] proposed a “Maximal High-Utility Itemset mining algorithm” that extracts maximal itemsets in a single parse by employing efficient pruning techniques, thereby minimizing the search space effectively. This algorithm was found to generate a lossless and compact representation of frequent patterns that can be used for mining association rules. An outline of the advantages and disadvantages of some of these algorithms have been highlighted (Table 2).

## 4 Conclusion

This study gives an overview of some of the techniques that are adopted for the mining of different types of frequent itemsets. The objective of this study was to review the pros and cons of the existing techniques and algorithms that are employed for mining frequent itemsets.

As a part of the future scope, the cons of these techniques may be considered as a motivation for the design of a data structure that effectively mines frequent itemsets. Although most of the existing algorithms perform well in terms of run time, they

**Table 2** Comparison of “Concise Frequent Itemset Mining Algorithms”

Algorithm	Advantages	Disadvantages
CFSP-Miner [27]	<ul style="list-style-type: none"> <li>Generates closed frequent patterns without information loss</li> </ul>	<ul style="list-style-type: none"> <li>Number of features has a negative effect on the scalability of the algorithm</li> </ul>
FPS-Tree algorithm, [26]	FPS-Tree has better runtime than LDS, FCI-Max, TOPSIL-Miner	Memory usage is high
NAFCP [28]	N-list algorithm compressed the input data	Not much improvement in memory while working on large datasets
CLOFAST [29]	CLOFAST performs better than CLASP, BIDE in terms of “run time” and “memory”	Does not handle noise in the data
NDHCMiner, NDHCderiveAll [20]	<ul style="list-style-type: none"> <li>Generates a lossless representation without having to perform additional scans on the database</li> <li>Performs well on dense datasets</li> <li>Generates a much concise set of patterns than HCMiner</li> </ul>	<ul style="list-style-type: none"> <li>No measures adopted to further reduce the size of the patterns</li> </ul>
IM_WMFI [34]	<ul style="list-style-type: none"> <li>Improved runtime</li> </ul>	<ul style="list-style-type: none"> <li>Performance degrades on large databases</li> </ul>
PFCCIM [30]	<ul style="list-style-type: none"> <li>Mines frequent colossal closed itemsets using a parallel approach</li> <li>Prunes irrelevant rows</li> <li>Reduces the search space</li> </ul>	<ul style="list-style-type: none"> <li>A limited investigation was done on designing a dynamic switching algorithm when the number of rows and features are very large</li> </ul>
POWER [38]	<ul style="list-style-type: none"> <li>Generates a concise and lossless set of patterns</li> <li>Has a better run time than contemporary algorithms</li> </ul>	<ul style="list-style-type: none"> <li>Generates unnecessary candidate itemsets</li> </ul>

either consume too much memory or their performance decreases with the difference in density of the datasets. This could be taken up for further investigation. Exploring the difference in performance to the nature of the dataset is another area where there is room for further study. The items that are mined are usually large and redundant. Designing effective algorithms that handle this seems to be the hour of the need especially with the massive volumes of data generated today. This also implies that the frameworks being developed should be scalable so that the large volumes of data may be handled effectively.

## References

1. Aryabarzan N, Minaei-Bidgoli B, Teshnehlab M (2018) negFIN: an efficient algorithm for fast mining frequent item sets. *Expert Syst Appl* 105:129–143
2. Deng ZH (2018) DiffNodesets: an efficient structure for fast mining frequent item sets. *Appl Soft Comput* 48(9):214–223
3. Deng ZH, Lv SL (2014) Fast mining frequent item sets using Nodesets. *Expert Syst Appl* 41(10):4505–4512
4. Kim J, Hwang B (2016) Real-time stream data mining based on CanTree and Gtree. *Inf Sci (Ny)* 367–368:512–528
5. Meng Q, Sha J (2017) Tree-based frequent item sets mining for analysis of life-satisfaction and loneliness of retired athletes. *Cluster Comput* 20(4):3327–3335
6. Tan L, Qin Q (2016) A new algorithm for fast mining frequent item sets based on SO-Sets. In: *ICEICT*, pp 342–346
7. Khawaja SG, Tehreem A, Akram MU (2017) Multicore framework for finding frequent item-sets using TDS. In: *Proceedings of the 16th international conference on Hybrid Intelligent Systems (HIS 2016)*, vol 552 (no. His 2016)
8. Francia M, Golfarelli M, Rizzi S (2020) Summarization and visualization of multi-level and multi-dimensional itemsets. *Inf Sci* 520:63–85
9. Shah A, Halim Z (2019) On efficient mining of frequent itemsets from big uncertain databases. *J Grid Comput* 17(4):831–850
10. Deng Z, Wang Z, Jiang J (2012) A new algorithm for fast mining frequent item sets using N-lists. *Sci China Inform Sci* 55(9):2008–2030
11. Huynh-Thi-Le Q, Le T, VO B, Le B (2015) An efficient and effective algorithm for mining top-rank-k frequent patterns. *Expert Syst Appl* 42(1):156–164
12. Bui H, Vo B, Nguyen H, Nguyen-Hoang TA, Hong TP (2018) A weighted N-list-based method for mining frequent weighted item sets. *Expert Syst Appl* 96:388–405
13. Deng ZH, Lv SL (2015) PrePost+: an efficient N-lists-based algorithm for mining frequent item sets via children-parent equivalence pruning. *Expert Syst Appl* 42(13):5424–5432
14. Vo B, Le T, Coenen F, Hong TP (2016) Mining frequent item sets using the N-list and subsume concepts. *Int J Mach Learn Cybern* 7(2):253–265
15. Zhang R, Chen W, Hsu TC, Yang H, Chung YC (2017) ANG: a combination of Apriori and graph computing techniques for frequent item sets mining. *J Supercomputer* 1–16
16. Dewar S, Goal V, Bear D (2017) A hybrid framework for mining high-utility item sets in a sparse transaction database. *Appl Intel* 47(3):809–827
17. Kalpana B, Natarajan R (2008) Incorporating heuristics for efficient search space pruning in frequent item set mining strategies. *Curr Sci* 94(1):97–101
18. Calder's T, Goethals B (2007) Non-derivable item set mining. *Data Min Knowl Disc* 14(1):171–206
19. Xiong H, Tan PN, Kumar V (2006) Hyperclique pattern discovery. *Data Min Knowl Disc* 13(2):219–242
20. Koufakou A (2014) Mining non-derivable hypercliques. *Knowl Inf Syst* 41(1):77–99
21. Calders T, Goethals B (2005) Depth-first non-derivable itemset mining. In *Proceedings of the 2005 SIAM international conference on data mining*. *SocInd Appl Math* 250–261
22. Hashem T, Rezaul Karim M, Samiullah M, Farhan Ahmed C (2017) An efficient dynamic superset bit-vector approach for mining frequent closed item sets and their lattice structure. *Expert Syst Appl* 67:252–271
23. Tran MT, Le B, Vo B (2015) Combination of dynamic bit vectors and transaction information for mining frequent closed sequences efficiently. *Eng Appl Artif Intell* 38:183–189
24. Zaki MJ, Hsiao CJ (2005) Efficient algorithms for mining closed item sets and their lattice structure. *IEEE Trans Knowl Data Eng* 17(4):462–478
25. Grahne G, Zhu J (2005) Fast algorithms for frequent item set mining using FP-trees. *IEEE Trans Knowl Data Eng* 17(10):1347–1362

26. ur Rehman Z, Shahbaz M, Shaheen M, Guergachi A (2015) FPS-tree algorithm to find top-k closed item sets in data streams. *Arab J Sci Eng* 40(12):3507–3521
27. Rodríguez-González Y, Lezama F, Iglesias-Alvarez CA, Martínez-Trinidad JF, Carrasco-Ochoa JA, de Cote EM (2018) Closed frequent similar pattern mining: reducing the number of frequent similar patterns without information loss. *Expert Syst Appl* 96:271–283
28. Le T, Vo B (2015) An N-list-based algorithm for mining frequent closed patterns. *Expert Syst Appl* 42(19):6648–6657
29. Fumarola F, Lanotte PF, Ceci M, Malerba D (2016) CloFAST: closed sequential pattern mining using sparse and vertical id-lists. *Knowl Inf Syst* 48(2):429–463
30. Vanahalli MK, Patil N (2019) An efficient parallel row enumerated algorithm for mining frequent colossal closed itemsets from high dimensional datasets. *Inf Sci* 496:343–362
31. Vanahalli MK, Patil N (2020) An efficient colossal closed itemset mining algorithm for a dataset with high dimensionality. *J King Saud Univ Comput Inf Sci*
32. Li B, Pei Z, Qin K, Kong M (2018) TT-miner: topology-transaction miner for mining closed itemset. *IEEE Access* 7:10798–10810
33. Muhonen J, Toivonen H (2006) Closed non-derivable itemsets. In: *PKDD*, Springer-Verlag, Berlin Heidelberg, 601–608 (LNAI 4213)
34. Yun U, Lee G (2016) Incremental mining of weighted maximal frequent itemsets from dynamic databases. *Expert Syst Appl* 54:304–327
35. Liu Z, Hu L, Wu C, Ding Y, Wen Q, Zhao J (2018) A novel process-based association rule approach through maximal frequent itemsets for big data processing. *Futur Gener Comput Syst* 81:414–424
36. Necir H, Drias H (2015) A distributed maximal frequent itemset mining with multiagents system on bitmap join indexes selection. *Int J Inf Technol Manage* 14(2–3):201–214
37. Chung SM, Luo C (2008) Efficient mining of maximal frequent itemsets from databases on a cluster of workstations. *Knowl Inf Syst* 16:359–391
38. Nguyen LT, Vu DB, Nguyen TD, Vo B (2020) Mining maximal high utility itemsets on dynamic profit databases. *Cybern Syst* 1–21

# A Novel Method for Pathogen Detection by Using Evanescent-Wave-Based Microscopy



Vijay A. Kanade

**Abstract** The research proposal discloses a novel microscopic device and method for making a colorless pathogen specimen (i.e., virus, bacteria) visible to the human eye. The novel device fits like a headgear on a user's head. Further, the novel method involves the usage of evanescent waves, that are generated by surfaces of the objects that undergo excitation on the incidence of a certain light wavelength. The microscopic device uses optical sensors to detect the evanescent waves generated by the object surface. Further, the microscopic device also uses nano projecting units for projecting the captured evanescent waves on the projection screen that in turn magnifies the captured object's evanescent waves. Hence, the disclosed research proposal serves as an effective solution for viewing tiny viruses that are not visible to the human eye, such as COVID-19 virus.

**Keywords** Microscopic device · Evanescent wave · Total internal reflection microscopy (TIRF) · Bacterial specimen · COVID-19 virus · Graphene aerogel (Aerographene)

## 1 Introduction

Viruses, bacteria are fundamentally colorless by nature. Unlike bacteria (which are single-celled beings), a virus is usually not defined under the umbrella of creatures, as they cannot survive independently. They need a host to thrive on.

Dimension wise, the virus structure ranges from 100 to 400 nm in diameter. While bacteria are about 1000 nm to 1  $\mu\text{m}$ . As per the research study on the COVID-19 virus, its dimension measures about 120–160 nm in size.

Such nanoscale sized objects are dealt with the lower band of the visible light spectrum, that the objects can reflect. Visible light falls under a specific wavelength. Lower threshold and upper threshold of the visible spectrum have a wavelength in

---

V. A. Kanade (✉)  
Intellectual Property Research, Pune, India  
e-mail: [kanade.science@gmail.com](mailto:kanade.science@gmail.com)

the range of 400–700 nm. Lower threshold defines the violet-blue, while the upper threshold defines red. Hence, the colors in the visible light spectrum lie within the range of 400–700 nm ( $\pm 50$  nm).

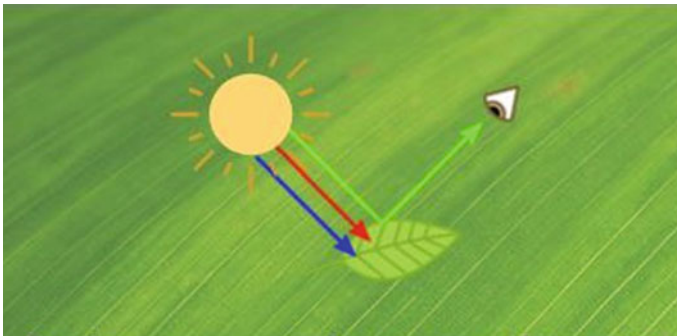
Now, viruses or a certain class of bacteria fall under this category wherein these creatures are too small to reflect any form of light. The lightwave incident on these biological creatures just passes through without being intercepted by their presence. This makes such viruses, bacteria invisible, even under the strongest of microscopes, having good magnifying power. As per the current research on COVID-19, the coronavirus also falls under colorless category [1].

Further, living beings have evolved a visible sense since their inception on this planet, wherein beings are sensitive to a specific wavelength which today is termed as the visible spectrum. This spectrum is in turn aligned to the light received from the sun [2].

Now, how do humans see such microscopic viruses, bacteria with such limited visual sense? Human perception of any colored object is dependent on white light emitted by sun, which is a combination of all wavelengths of the visible spectrum. When light is intercepted by an object, a certain portion of the light is reflected by the object. The reflected light is received and absorbed by human eyes, and thus, the color of the object is perceived (Fig. 1).

Hence, to tackle the scenarios such as an epidemic, COVID-19 pandemic, etc., it is important to develop a technique that makes the viruses visible enough so that a solution can be worked out for disinfecting the surfaces that are highly susceptible or exposed to viruses. Currently, disinfection of surfaces or areas is performed randomly, without much knowledge of whether the surface has viruses clung to it or not.

The current research discloses a novel method for detecting the colorless pathogen specimens such as viruses, bacteria that not usually visible to humans even after magnification.



**Fig. 1** Leaf reflects green wavelength from the received solar light [2]



## 2 Principle

The research proposal operates on the principle of ‘Total Internal Reflection Microscopy (TIRF)’ discussed below.

### 2.1 Total Internal Reflection Fluorescence Microscopy (TIRF)

TIRF microscopy is a fluorescence phenomenon for imaging objects within the range of 50–250 nm that are placed adjacent to a surface. Such a microscopy technique detects evanescent waves generated by total internal reflection of the incident light. TIRF microscopy is used for studying the biological specimens, molecular activities occurring near the cell surface, etc. (Fig. 2).

When light is incident on a surface of another medium at an angle, light undergoes refraction. However, when light hits the surface with the angle of incidence greater than the critical angle (i.e., refraction angle is  $90^\circ$ ), the light undergoes total internal reflection and the entire spectrum of the incident light is reflected from the surface.

Here, as total internal reflection occurs, a certain portion of the reflected light is absorbed by the surface, thereby allowing the light to penetrate through the surface. Such penetration of light develops an electromagnetic field ( $<250$  nm) across the surface. The thin electromagnetic field consequentially generates evanescent waves adjacent to the surface. The developed evanescent waves travel parallel to the surface with a frequency that is equivalent to the frequency of the incident light. Further, these evanescent waves die down in their intensity as the wave travels away from the surface [3].

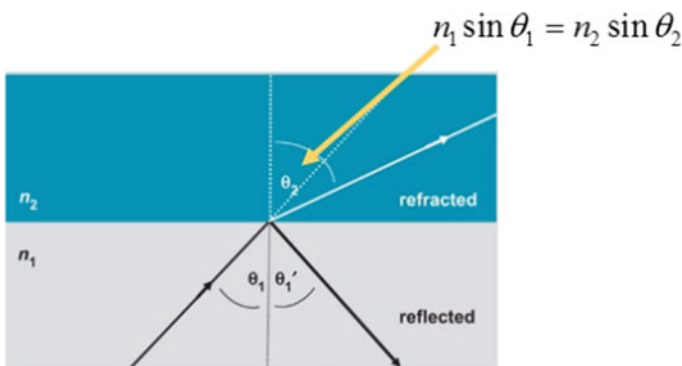


Fig. 2 Total internal reflection

Further, it is important to note that the evanescent waves have sufficient energy for excitation of specimen lying adjacent to the surface. The specimen (i.e., micro-organisms, viruses, bacteria, etc.) discussed here may have natural fluorescence or induced fluorescence. Here, certain microscopic organisms have enough internal fluorescence that may trigger excitation by the incidence of the external light at a certain angle. In another scenario, the micro-organisms may be impregnated with the fluorochromes or chemical that may undergo fluorescence [4].

The researchers at the University of Manchester have developed a white light nanoscope that makes the live virus visible by using evanescent wave. However, the designed microscopic unit is bulky in nature and expensive as conventional microscopes [5].

## 2.2 Total Internal Reflection Parameters

The conditions required for total internal reflection (TIR) to occur are:

- The light must be traveling from a denser medium into a less dense medium (i.e., glass to air) since as light travels from one medium to another, the speed of light is altered. Now, as light travel from denser to less dense media, the speed of light increases and the light ray bends away from the normal. This ‘away bending’ bending phenomenon is critical for total internal reflection, as in case of light traveling from rarer medium to denser media, the light ray bends towards the normal, hence total internal reflection cannot occur under such circumstances
- The angle of incidence must be greater than the critical angle.

## 2.3 Evanescent-Wave Mathematics

As light passes through the two different media, the phenomenon of refraction, refraction at a critical angle ( $\theta_c$ ), and total internal reflection is depicted in Fig. 3.

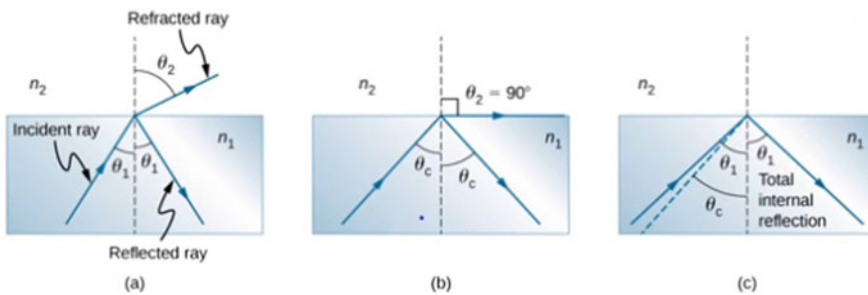


Fig. 3 Light propagation

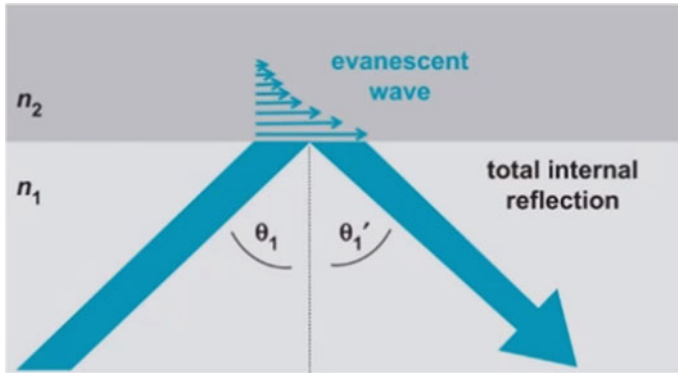


Fig. 4 Evanescent waves

Mathematically, according to Snell’s Law,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1$$

Note: Here, the maximum value must be  $\sin \theta_2 \leq 1$ . If  $\sin \theta_2 = 1$ , then the refracted wave travels at  $90^\circ$  and eventually glides along the surface of the medium.

Now, if  $n_1 > n_2$ , then the condition  $\sin \theta_2 \leq 1$  of Snell’s Law is not satisfied.

Further, when  $\theta > \theta_c$  (critical angle), there is a proper total internal reflection.

Now, for wave continuity,

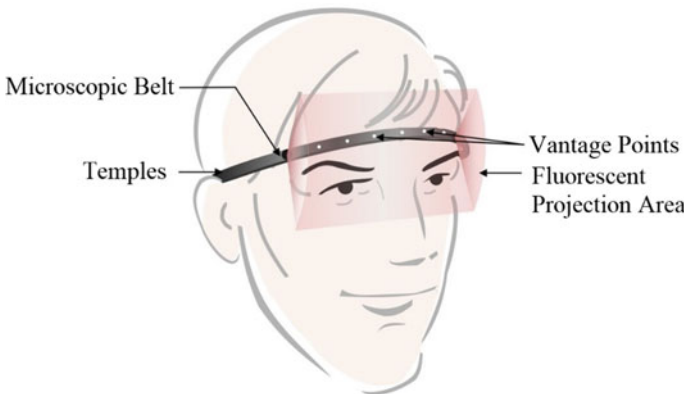
$$\varepsilon \geq 1 \tag{1}$$

Hence, in case of total internal reflection, there is no light component passing through the surface, and therefore to obey the principle in (1), there must be a component along the surface. It turns out that in case of total internal reflection, it is evanescent waves that are decaying waves along the surface of the medium. Here, the length of decay along the surface is approx. equivalent to  $\lambda/2$ . The evanescent waves thus disappear quickly along the surface [6] (Fig. 4).

### 3 Inventive Microscopic Technology

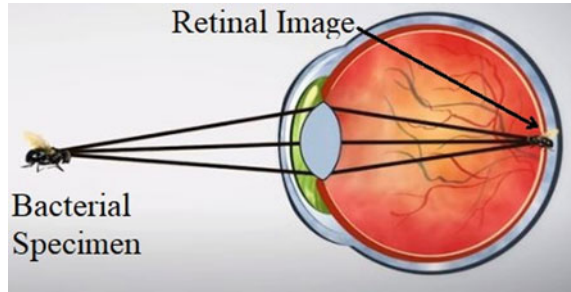
The advancement in imaging technology in the current IoT era knows no bounds. There is an upsurge in the usage of the IoT-based solutions for healthcare applications [7, 8]. The proposed research is one such IoT-based healthcare solution that is deployed to facilitate rapid pathogen detection.

The research proposal discloses a novel microscopic belt that is used for making the colorless biological specimen (i.e., virus, bacteria) visible to the normal human eye. The microscopic belt serves as a headgear that a user can wear over the head. The microscopic device includes a fluorescent tube along its length with various vantage points on it. The vantage points emit the fluorescent light generated by the underlying fluorescent tube. Furthermore, the microscopic belt has built-in graphene aerogel storage unit along with a spraying apparatus for spraying a thin film of the aerogel on the target surface [9]. The graphene aerogel is colorless and less dense medium than air, hence as the aerogel film is sprayed on the surface, the surface film becomes less dense than outside air [10]. Therefore, as the fluorescent light is emitted from the vantage points of the microscopic device, the light rays travel through the air (i.e., denser medium as compared to aerogel sprayed surface) to the target surface layered by the aerogel. The thin-film coating of aerogel allows the total internal reflection of light rays that are incident on its surface, thereby creating evanescent waves along the surface. Furthermore, the temples of the microscopic belt have in-built optical sensors that detect and record the evanescent waves generated adjacent to the surface on which the fluorescent light is incident. Here, the optical sensors detect the intensity of the evanescent waves by recording the pixel intensity of the entire evanescent waves that pass parallel to the surface. Notably, the evanescent waves make the virus, bacteria or the pathogen lying on the surface visible as evanescent waves traverse through the biological specimen or pathogen. Furthermore, the microscopic belt also has in-built nano-projecting units that project the optical sensor captured pixel intensity on the fluorescent projection area. This makes the bacterial specimen along the target surface viewable on the portable microscopic device (Fig. 5).



**Fig. 5** Microscopic belt

**Fig. 6** Visible bacterial specimen



### 3.1 Operative Mechanism

Humans can perceive any object due to the light reflected by the object. The reflected light passes through the lenses of the eye and forms an inverted image on the retina. This inverted retinal image sends signals to the brain of the presence of the object, and thereby the object is perceived.

Consider a bacterial specimen which is at a comfortable distance from the eye. The specimen is visible since the light from the bacterial specimen forms a retinal image over the retina (Fig. 6).

Now, consider a scenario, wherein the microscopic device emits fluorescent light and the light is incident on the surface that is infected with a colorless bacterial specimen. The bacterial specimen is hit by the fluorescent light at an angle greater than the critical angle. This causes total internal reflection and further, a certain portion of the fluorescent light penetrates the surface. Since the surface has the bacterial specimen on it, the fluorescent light passes through the specimen's body. This, in turn, generates evanescent waves nearby to the specimen. The generated evanescent waves are captured by the optical sensors of the microscopic device. Once captured, the nano-projecting unit in the microscopic device project the virtual projection of the captured pixel intensities of the evanescent waves generated over the bacterial specimen. The virtual image of the specimen is projected on the adjacent screen attached to the microscopic device. The projecting screen, therefore, serves as a magnifying apparatus, wherein the evanescent waves from the real specimen at a distance are captured and, are projected right in front of the microscopic device in such a way that the retinal image of the virtual specimen is created by the screen. The microscopic device has a half convex lens for focusing light on the retina of the user to create the retinal image. The retinal image of the real specimen makes the image visible to the human eye (Fig. 7).

The diagrammatic representation of the microscopic device with the magnifying and projecting screen is disclosed in Fig. 8.

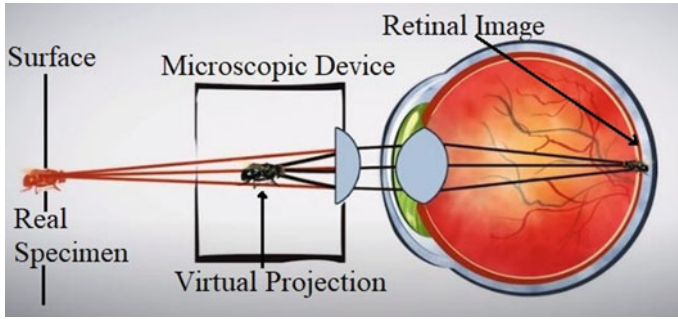


Fig. 7 Microscopic device magnifying the bacterial specimen at a distance

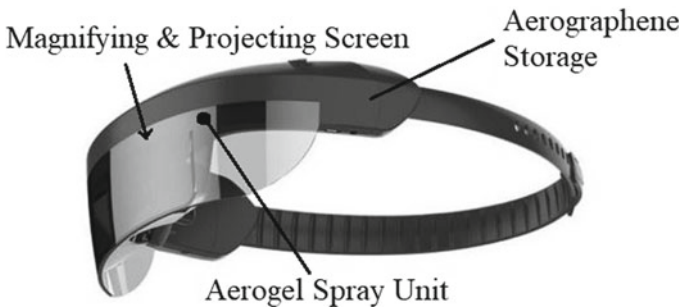


Fig. 8 Microscopic device

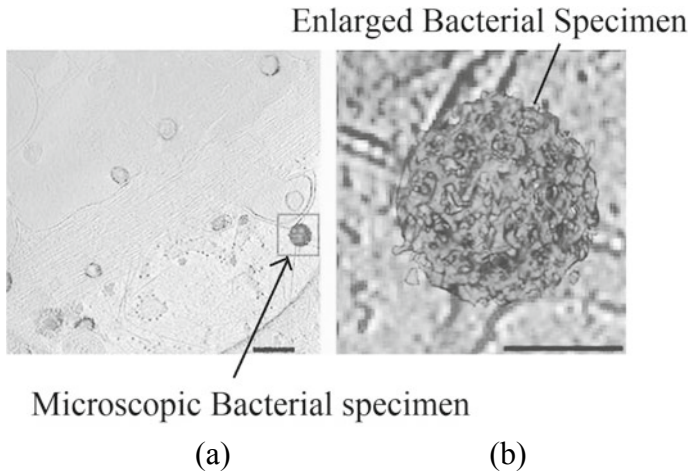
### 4 Preliminary Results

Some of the preliminary results obtained by using the TIRF microscopy are presented in the figure below. The results presented here specifically pertain to a pathogenic (bacteria) specimen, however, the procedure proposed in the research proposal is applicable for pathogens of any form, i.e., bacteria, virus, or micro-organism.

Initially, the biological specimen (i.e., pathogen/bacteria in this case) at a distance was hit by the fluorescent light. This procedure generated the evanescent waves on the biological specimen (Fig. 9a). Further, the captured fluorescence of the specimen was observed to be magnified by the projecting unit as seen through the microscopic device in Fig. 9b below.

### 5 Advantages

The disclosed solution is inexpensive and effective as compared to the conventionally used standard microscopes. The traditional microscopes are bulky and expensive.



**Fig. 9** **a** Fluorescent light falling on the specimen, **b** magnified image of the specimen

However, the designed microscopic device uses an inexpensive fluorescent tube, optical sensors, and nano-projectors for viewing the tiny viruses. The inexpensive nature of the microscopic device allows the device to be easily adopted by anyone (i.e., layman).

## 6 Conclusion

The research proposal discloses a novel device and method for making the colorless pathogen specimen visible. The disclosed solution can be used in scenarios wherein the novel viruses such as COVID-19 virus can be seen irrespective of its color and size. The research proposal provides an effective alternative to standard microscope technology that is traditionally used in research labs.

**Acknowledgements** I would like to extend my sincere gratitude to Dr. A. S. Kanade for his relentless support during my research work.

## References

1. Weaving S (2020) Scary red or icky green? We can't say what color coronavirus is and dressing it up might feed fears 31 March 2020
2. Otap L (2020) Virus have no color—not even gray 28 May 2020
3. Roberts Research Institute Types of fluorescence molecular imaging
4. Darken MA (1960) Natural and induced fluorescence in microscopic organisms

5. Wang Z, Guo W, Li L et al (2011) Optical virtual imaging at 50 nm lateral resolution with a white-light nanoscope. *Nat Commun* 2:218
6. Kay Law Y (2017) Total internal reflection and evanescent waves 5 March 2017
7. Chandy A (2019) A review on IoT based medical imaging technology for healthcare applications. *J Innovative Image Proc* 1(01):51–60
8. Smys S (2019) Big data business analytics as a strategic asset for health care industry. *J ISMAC* 1(02):92–100
9. Carey T et al (2018) Spray-coating thin films on three-dimensional surfaces for a semi-transparent capacitive-touch device 10 May 2018
10. Anthony S (2013) Graphene aerogel is seven times lighter than air, can balance on a blade of grass 10 April 2013



# Analyzing Social Media Data for Better Understanding Students' Learning Experiences



T. Ganesan, Sathigari Anuradha, Attada Harika, Neelisetty Nikitha, and Sunanda Nalajala

**Abstract** Social media keeps on increasing in size and demands automation in data analysis. Student shares their opinions, concerns, and emotions in the social Web site, because it has a variety of opinions that are central to most of the human activities and a key influence of behavior in their day-to-day life. Many of the tweets made by students have some sort of meaning, but some category does not have a clear meaning such as a long tail. In this paper, a different classification model is developed to analyze student's comments which are available in social media. This paper mainly focused on emotions. Data is taken from 15,000 tweets of student's college life and categories—study load of all majors, antisocial, depression, negative emotion, external factors, sleep problems, diversity problems. These multi-label emotional comments are to be classified, analyzed, and compared with the support vector machine and Naïve Bayes algorithm to show student learning problems. The experimental results show that major students' learning problems make better decisions for future education and service to them.

**Keywords** Sentiment · Social media · Classification · Data collection · Education · Learning experience · Feedback · Emotional data · Text analysis

---

T. Ganesan (✉) · S. Anuradha · A. Harika · N. Nikitha · S. Nalajala  
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,  
Vaddeswaram, Guntur, Andhra Pradesh 522502, India  
e-mail: [tganesanit@gmail.com](mailto:tganesanit@gmail.com)

S. Anuradha  
e-mail: [sathigarianuradha@gmail.com](mailto:sathigarianuradha@gmail.com)

A. Harika  
e-mail: [harika.attada@gmail.com](mailto:harika.attada@gmail.com)

N. Nikitha  
e-mail: [nikithaneelisetty@gmail.com](mailto:nikithaneelisetty@gmail.com)

# 1 Introduction

The social media Web sites like Twitter (TW), Facebook (FB), and YouTube (YT) offer good platforms for various learners like students to share their happiness, struggle, stress, and emotions. Sentimental analysis analyzes people’s feedback and reviews which is usually an opinion poll or opinion mining. Here, student learning experience can assess outside the classroom by social gadgets and inform institutions to take better decisions of education quality [1, 14]. The main challenge in social media data is the large volume of data, Internet local slang, different locations, short comments, posting time, and repeated comments [2]. Normally, educational institution collects student learning experience by surveys, interview, feedback, group discussion, classroom activities, and one-to-one interaction [3, 4]. These methods are the very oldest and time taken and sometimes the same kind of usual feedbacks because there is no privacy of feedback [5, 14].

The modern learning analytics of education data mining is focused on course management and time-limited online classrooms to take decisions for better education [6, 7]. The sentimental analysis characterization is done by natural language processing or machine learning methodology. Sentimental analysis is generally a supervised learning algorithm, in which classifications are done on negative, positive words, booster words and idioms, etc. Sentimental analysis is known as a text mining technique that describes the mining of the text and analyzing the sentiment of the text. There are several social media Web sites like TW, FB, Instagram, WhatsApp, and YouTube [2, 6, 7].

Sentimental analysis is termed as people’s opinions and views, typically as opinion mining. This methodology is the time taken, more redundant data, unformed data, and different local language-related data [7, 8, 10]. The sentimental analysis characterization is usually done by natural language processing or machine learning methodology. Sentimental analysis is generally a supervised learning algorithm in which classifications are done on negative, positive words, booster words and idioms, etc. [12, 13]. In students, sentimental analysis plays a major role because every student has their feelings, emotions which they would share through their social media Web sites. Many of them share through Twitter in the way of tweets. The tweets may be of positive, negative, or neutral tweets. Table 1 shows the different categories of tweets.

The real-time collecting feedback on a particular field is mainly done by students. Let us consider an example; it was a tweet by a student like “Laying on the bed ... feeling happy ... chilled out today” which comes under positive emotion as the person is in a happy mood. The sentiment mining can include identification of simple

**Table 1** General examples for classification of tweets

Event set	Example	Emotion
Fun	You are going be the first twitter;) cause your amazing lol.	Positive
Sadness	Ok... the passengers... no one is alive... they’re all dead	Negative
Neutral	Cannot fall asleep	Neutral

statements, negation, and the modifiers [10, 11]. Based on the text, the text is assessed as positive or negative or neutral. They are many applications of sentimental analysis in different fields such as business application, product prediction, student emotions, and customer services.

The traditional approach of collecting any feedback is done through public opinions, surveys, and some sort of interviews, which is generally a huge task and time-consuming. But nowadays, the collection of data is done through many social media Web sites. Here, let us consider dataset based on the tweets by students. These tweets which contain all types of emotions are tweeted by the students.

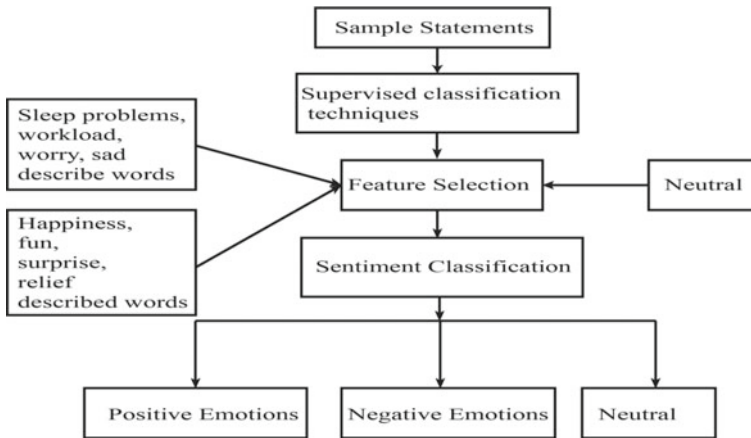
In this paper, the student's reviews, feedback, and comments are classified into three different categories as mentioned above in Table 1. This classification is further classified into a multi-label of Naïve Bayes and SVM classification methods to predict student emotional factors. The classification data represents a different category of study load of all majors, antisocial, depression, negative emotion, external factors, sleep problems, diversity problems presentation to find the accuracy of methods which analysis emotional factor for better learning. The remaining of this paper have a data collection model, related works done by researchers, algorithm evaluation then conclusion of results.

## 2 Related Background Work

The students learning experience for sentimental analysis and traditional methods are replaced by machine learning-based algorithms to increase more accuracy by using different classification techniques. Sometimes social media data are not real data compared to original identity [2, 9] and to get more comments to compare to students in classroom data. Researchers are from various fields to analyze Twitter content having a specific domain of their profession. The same kind of study has done in health care, marketing, sports, etc. These studies are to be done based on content, profession, language-specific, friends circle, short forms, symbolic represent [3, 12].

The classification consists of training data and testing data, where the relevant tweets are extracted and trained on the training data and then tested on unseen data. Figure 1 shows that data for supervised classification model and feature selection is based on tweets categorized into positive, negative and neutrals. This collected dataset from different source applied for further classification into multi-labeled study load of all majors, antisocial, depression, negative emotion, external factors, sleep problems, diversity problems sets [1, 4].

The sentimental analysis starts with collection labeled tweets that are handled by natural language processing (NLP) techniques [11, 12]. The features for analysis are extracted, and different classifiers are trained with known data and tested on unknown data. The tweets may be noisy, so they need to be preprocessed by using preprocessing techniques such as removal of stop words, removal of punctuation, stemming, lemmatization, and spelling corrections. After the tweets get preprocessed, the extraction of feature is done for relevant tweet analysis.



**Fig. 1** Process flow for sentimental analysis for learning student's experience

Student conversations on social media like TW, Instagram share their educational learning experiences, feelings, opinions, and issues concerning the educational method. Analyzing such information might be difficult. Increasing information demands automatic information analysis techniques. Focusing on tweets is posted by various engineering students to grasp issues in their instructional practices. Qualitative analysis is finished considering 25,000 tweets linked with student's life [1, 11, 13]. Issues like serious lack of social commitment, more study load, and sleep deprivation. A multi-label classifier algorithm is employed to classify the various tweets of reflective students' issues. An algorithm is to verify student issues from concerning 35,000 tweets at Purdue University.

A social media data classification system is an inclusive approach to classification. The description of free-text metadata and generation of tags and resource discovery refer to a loose superset [10, 12]. Decomposition of compound terms to be done by retrieving information from within compound terms. More limitation in stemming algorithms suffers from exceptions in grammar.

With the efficient development of the internet, most of the individuals prefer to categories their thoughts, like about social and personal, views and attitude over the Web, that increase the more users generated content and opinionated information. The progressive approach permits associate degree previous results to be newly updated victimization solely new information instances, which cannot be re-process the past instances [5, 9].

Social media networks such as FB, TW, and Instagram have changed the way our society works. Social media is generally used for commercial perspective and different from customer perspectives, i.e., to collect the business review and shares and services provided to the public and analysing general people's feedback and reviews [2, 5]. The opinion mining for sentiment analysis follows the data extraction, right data identification, or characterization of the sentiment from text or file and

proposes a well-defined feature extraction and reduction algorithm by an evolutionary algorithm-based approach with a customized fitness function.

The quantitative analysis is to do in most popular websites that are based on different UGC-referred types. Internet users having the flexibility to access and to contribute content revolution such a way information or data is created and distributed among social. The famous resource person in social media or content in UGC-referred sites like YouTube has been already quantitatively analyzed [12, 13].

Feedback in education may be categorized into the lecturer to the scholars; this can be for the self-improvement of the students and feedback from the scholars to the lecturer; this enables them to guide the lecturer into teaching the course in ways in which they perceive best [9, 10]. Student response system (SRS) is employed for feedback within the room, given by students to the lecturer, via devices like clickers and mobiles. Sentiment analysis has not been enforced on the academic sector, however.

First, users' activity is glanced at with relevance to their tag use and next tend to examine tags themselves in larger detail. To tend to flip our attention to URLs, the bookmarks that reference them, and therefore the tags that describe them [7].

### 3 Data Collection and Classifier Summary

The data collection is a very challenging task in social media because different languages are used, many short forms of comments, and a variety of emotional, feeling of student feedbacks [1]. The data can be collected from different sources such as Twitter, Facebook, WhatsApp, LinkedIn, etc., the data here have been used for understanding students learning experience is from Twitter data [8]. Table 2 shows that different classifier labels, descriptions, and comments. Generally, the data from Twitter is collected from Twitter API, which is used to allow access to Twitter features without going through the website and they should meet Twitter applications. Data are collected from different source ratios. These data can be categories of classroom data, lab, assignments related, homework, and hashtags, etc. But this dataset is having more noise data [9, 10].

## 4 Content Analysis Based on Classifier Algorithms

### 4.1 Text Preprocessing

Students can express their learning views on Twitter to convey certain symbols of ad short forms. For example, # symbols used for a hashtag, @ used to indicate person and some emoji. Twitter users used some repeated letters for expressing feeling like “happyyy”, “hellooooo”, “nooooo” and “classssss”. And also used some stop

**Table 2** Students learning experience on sentimental analysis

S. No.	Classifier labels	Label description	Students comments	Data source ratio (%)
1	Study load of all majors	Students express the study load of classes, homework, assignment, projects, exams, labs, quiz of all the major subjects in the class	<i>“Not enough one night a week”, “At first it was hard to pick out subjects”, “I felt rushed and stressed over that material”. “It was difficult at first to keep all the postings threaded”</i>	Tweets: 60 Facebook: 10 #hashtag: 10 Others: 20
2	Antisocial	Students need to sacrifice the summer or weekend holidays, special family occasions, function, party, family, and their friends for their exams	<i>“I feel like I’m hidden from the world”, “math’s for every day”, “nothing is fun in my life”, “call me as machine”</i>	
3	Depression	Students state of mind, enjoyment of life, inability, and about the future	<i>“Now I’m mostly content with my life”, “Worst course in engg”, “I think my future is dark”</i>	
4	Negative emotion	Students express anger, hatred, stress, sickness, depression, disappointment	<i>“What a hell subject, when I can reach heaven”, “I need end of my study”, “I think u have to find better faculty”</i>	
5	External factors	The learning process can affect the Internet, friends, rain, and some controllable and uncontrollable external factors. Students can express these factors	<i>“tnks for my breakfast in M3 class”, “I pray for today power cut afternoon”, “Hey I didn’t study for today exams”, “wow beautiful nature climate”</i>	
6	Sleep problems	Due to more homework and assignments, students frequently feel suffer from lack of sleep and nightmares	<i>“I think my mind is out”, “I feel my class is rhythm”, “afternoon math needs heavy lunch”, “I’m flying in the sky”</i>	

(continued)

**Table 2** (continued)

S. No.	Classifier labels	Label description	Students comments	Data source ratio (%)
7	Diversity problems	Students have feels difficulties embracing diversity, because of so many cultural conflicts	<i>“I’m sorry to say that. were not used to having girls around in our university”, “I pity the 1 girl in my lab with 25 guys. It smells like a man in here.. .. And that’s not in a good way”, “Finally talked to a girl today!!! It was Siri”, and “Let’s start with an example, tell me something you know nothing about’– Professor ‘girls.’ – Students. lol”</i>	

words as “a, an, and, for, of, hey, they, he, she and it”. For text preprocessing is removed all the symbols, emoji, punctuation, and repeated letters by using Lemur toolkit and kept only text for analysis.

### 4.2 Naive Bayes Classification

The multiple single label classification is transformed into a multi-label classification problem. Suppose the data collection having a total  $N$  number of words  $W = \{W_1, W_2, W_3, \dots, W_N\}$  and the total number of label categories  $C = \{C_1, C_2, C_3, \dots, C_L\}$ . A word  $W_N$  can appear in some label category  $C_L$  in  $M$  times, the probability of frequency estimation is given Eq. (1) and the same word is categories other than  $C_L$  is represent Eq. (2).

$$P(W_N|C_L) = \frac{M_{W_N C_L}}{\sum_{N=1}^N M_{W_N C_L}} \tag{1}$$

$$P(W_N|C_L'') = \frac{M_{W_N C_L''}}{\sum_{N=1}^N M_{W_N C_L''}} \tag{2}$$

For the testing set in document  $D$  having  $K$  number of word, so that  $W_K$  and  $W_D$  is the  $K$  number of the word in document and subset of  $W$  respectively. Therefore in Naive Bayes' theorem, the probability of  $D_i$  belongs to category  $C$  and other than  $c$  is

$$P(C|D_i) = \frac{P(D_i|C) \cdot P(C)}{P(D_i)} \propto \prod_{K=1}^K P(W_{iK}|C) \cdot P(C) \quad (3)$$

$$P(C''|D_i) = \frac{P(D_i|C'') \cdot P(C'')}{P(D_i)} \propto \prod_{K=1}^K P(W_{iK}|C'') \cdot P(C'') \quad (4)$$

### 4.3 Support Vector Machine Classification

The support vector can be used for mostly multi-class and multi-label classification and regression-based problems. The main part is that finding hyperplane is the best separate feature from others. Technically, hyperplane called as margin maximizing hyperplane. The classification task labeled  $X$  labels from  $N$  classes, where  $X$  can be 0 to  $N$ . Formally, binary values are assigned to each class, where positive classes are 1 and negative classes are 0 or  $-1$ . This approach threatens each label independently, whereas a multi-label classifier threatens the multi-class simultaneously.

### 4.4 Label-Based Evaluation

The label-based measure has to be calculated for the average of each category in  $C_L$ . Here one with rest binary classification to predict and not predicted in  $C_L$  based on the confusion matrix True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Based on that the accuracy, precision, and recall value to be calculated.

$$\text{Accuracy(Acc)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (5)$$

$$\text{Precision(Pr)} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (6)$$

$$\text{Recall(Rc)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$



### 5 Comparison of Classifier Algorithms

The classifier analysis for knowing students learning experiences, the algorithms comparisons are done based on Twitter dataset which consists of 9684 records and has the attribute values as Twitter id, tweets, and sentiment. The comparisons are made on the class label (sentiment is the class label in the dataset considered). Among all the algorithms used the more accurate algorithm found is logistic regression.

Figure 2 represents the Twitter and Facebook data classification based on Positive, Negative, and Neutral datasets.

Figure 3 represents data collection to detect potential student problems from TW, FB, different hashtags, and other sources. The student learning experience collected from all categories of students in different sources. The trained dataset detector can

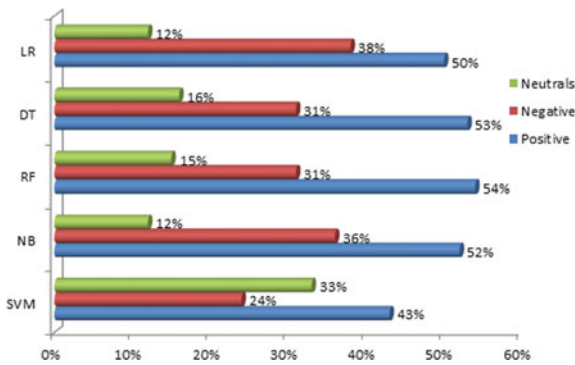


Fig. 2 Positive, negative and neutrals Twitter and FB data sets

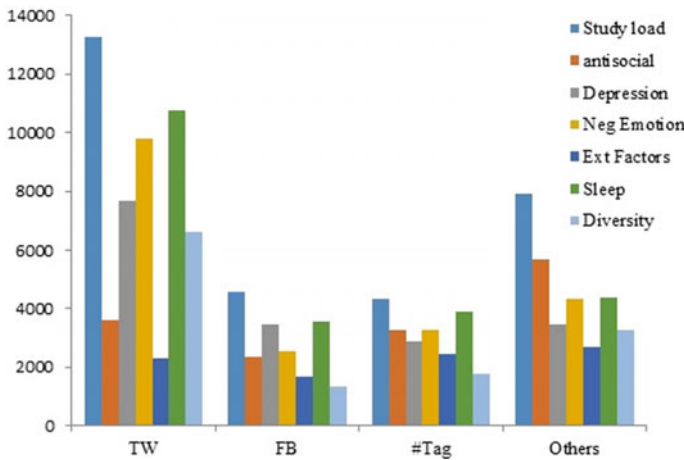


Fig. 3 Data collected from different source

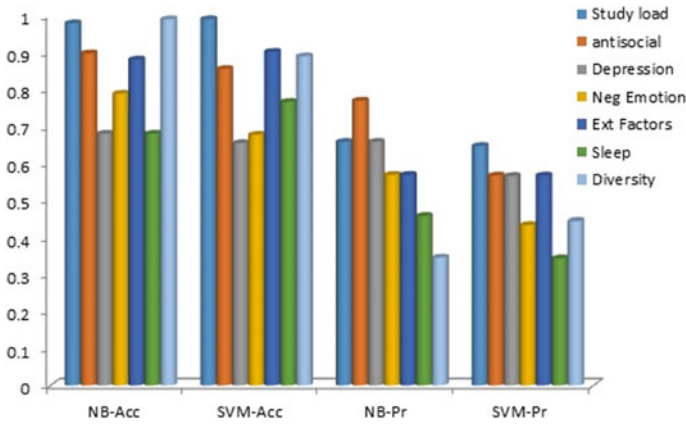


Fig. 4 Label-based Naïve Bayes and SVM accuracy and precision

be applied as monitoring data for the long run, and it can able to identify the severe cases of at-risk students. For example, in future mostly students may post a very large number of tweets, comments, and reviews about students’ study problems or negative emotions.

However, Fig. 4 represents multi-label classifier accuracy and precision about training data and testing datasets as in the above section, this NB multi-label and one-versus-all SVM multi-label classifier classified all the student’s tweets, FB, #tag, and some other source accuracy and precision to be measured. Based on a multi-label classifier algorithm like Naïve Bayes has not only attained significant improvement but also somewhat exceeded the performance of one-versus-all multi-label classifiers in support vector machine.

## 6 Conclusion

The experimental study works to help many researchers’ analytics of learning experience, educational data mining, and a new way of teaching-learning techniques. According to previous researches, naive Bayes has given more accuracy and it is best used for text mining in sentimental analysis. To our observation, most of the students are in a state that they could not properly express the emotions most of them are in a neutral state. To know more about the students learning experiences their imbalanced state mostly study load, negative emotion about subjects, and sleeping problems are needed to understand. This would help most colleges and institutions to understand student’s level of thinking and their performance.

## References

1. Chen X, Mihaela V, Krishna M (2013) Mining social media data for understanding students' learning experiences. *IEEE Trans Learn Technol* 7(3):1939–1382
2. Rost M, Barkhuus L, Cramer H, Brown B (2013) Representation and communication: challenges in interpreting large social media datasets. In: *Proceedings of the conference on computer supported cooperative work*, pp 357–362
3. Ferguson R (2012) The state of learning analytics in 2012: a review and future challenges. Technical Report KMI-2012-01, Knowledge Media Institute
4. Vorvoreanu M, Clark Q (2010) Managing identity across social networks. In: *Proceedings of the poster session at the ACM conference computer supported cooperative work*
5. Siemens G, d Baker RS (2012) Learning analytics and educational data mining: towards communication and collaboration. In: *Proceedings of the second international conference on learning analytics and knowledge*, pp 252–254
6. Arnold KE, Pistilli MD (2012) Course signals at purdue: using learning analytics to increase student success. In: *Proceedings of the second international conference on learning analytics and knowledge*, pp 267–270
7. Tsoumakas G, Katakis I, Vlahavas I (2010) Mining multi-label data. In: *Data mining and knowledge discovery handbook*. Springer, Berlin, pp 667–685
8. Van Asch V (2012) Macro- and micro-averaged evaluation measures. [http://www.cnts.ua.ac.be/\\_vincent/pdf/microaverage.pdf](http://www.cnts.ua.ac.be/_vincent/pdf/microaverage.pdf)
9. Cristianini N, Shawe-Taylor J (2000) *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, Cambridge
10. Chang C-C, Lin C-J (2012) LIBSVM—a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
11. Hariharan B, Zelnik-Manor L, Vishwanathan SVN, Varma M (2010) Large scale max-margin multi-label classification with priors. In: *Proceedings of the international conference on machine learning*
12. Hariharan B, Zelnik-Manor L, Vishwanathan SVN, Varma M (2013) Code for large scale multi-label classification with priors. <http://www.cs.berkeley.edu/~bharath2/codes/M3L/download.html>
13. Chen X, Vorvoreanu M, Madhavan K (2013) A web-based tool for collaborative social media data analysis. In: *Proceedings of the third international conference on social computing and its applications*
14. Sridevi K, Ganesan T, Samrat BVS, Srihari S (2019) Traffic analysis by using random forest algorithm considering social media platforms. *Int J Recent Technol Eng* 7(6S):620–625

# Load Balancing in Cloud Computing Environment: A Broad Perspective



Minakshi Sharma, Rajneesh Kumar, and Anurag Jain

**Abstract** Cloud computing is a recent buzzword in the field of information technology (IT) that has changed the way of computing from personal computing to virtual computing done with the help of cloud service providers via the Internet. It is evolved from distributed computing included with some other technologies such as virtualization, utility computing, service-oriented architecture, and data center automation. It is based upon the concept of computing delivered as a utility encapsulated with various other characteristics such as elasticity, scalability, on-demand access, and some other prominent features. To keep these properties intact during high demand for services, the load must be balanced among the available resources in the cloud environment. This load can be of various types such as CPU load, network load, memory load, etc., and balanced by executing a load balancing mechanism after detecting the overloaded and underloaded nodes. To achieve this researchers design different types of load-balancing algorithms for optimizing different performance parameters. The paper deals with a broad perspective of various load-balancing approaches done in the field by assuming the different performance metrics. The authors discuss that these approaches are multi-objective and there should be a good trade-off among these metrics to improve the performance.

**Keywords** Cloud computing · Load balancing · Elasticity · Performance · Optimization criterion

---

M. Sharma (✉) · R. Kumar

Department of Computer Science and Engineering, MMEC, Maharishi Markandeshwar Deemed to be University, Mullana, Ambala, Haryana 134002, India

e-mail: [Minakshi.sharma.12@gmail.com](mailto:Minakshi.sharma.12@gmail.com)

R. Kumar

e-mail: [dr Rajneeshgujral@mmumullana.org](mailto:dr Rajneeshgujral@mmumullana.org)

A. Jain

Virtualization Department, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand 248007, India

e-mail: [anurag.jain@ddn.upes.ac.in](mailto:anurag.jain@ddn.upes.ac.in)

# 1 Introduction

Cloud computing stands on virtualization, capable of providing software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS) over the Internet. This virtualization technology empowers the cloud environment by providing on-demand resource allocation to the consumers more flexibly and securely. A single host in a data center is capable to accommodate multiple virtual machines to satisfy varying consumer demands known as workloads. The unpredictable arrival of these variable workloads in cloud environs imbalance the load due to the irregular usage of resources by consumers. This results in fluctuating performance of standalone hosts which leads to performance degradation and SLA violations. This performance degradation can be overcome by a load-balancing mechanism so that no particular node is underloaded or overwhelmed. This balancing mechanism in cloud environs distributes the dynamic workload ideally across all the nodes. This increases resource utilization and incurs the stability in system performance in the cloud data center. Besides that, it also ensures the consumer satisfaction by providing consistent performance to the consumer in terms of response time. This load-balancing approach can be integrated at two levels in the cloud environment at the host level and VM level. Figure 1 shows the layered presentation of hosts, VMs, and applications.

At the lower layer, some hosts have real resources to provide such as memory, processing power, and storage. Above the host, there is a virtualization layer

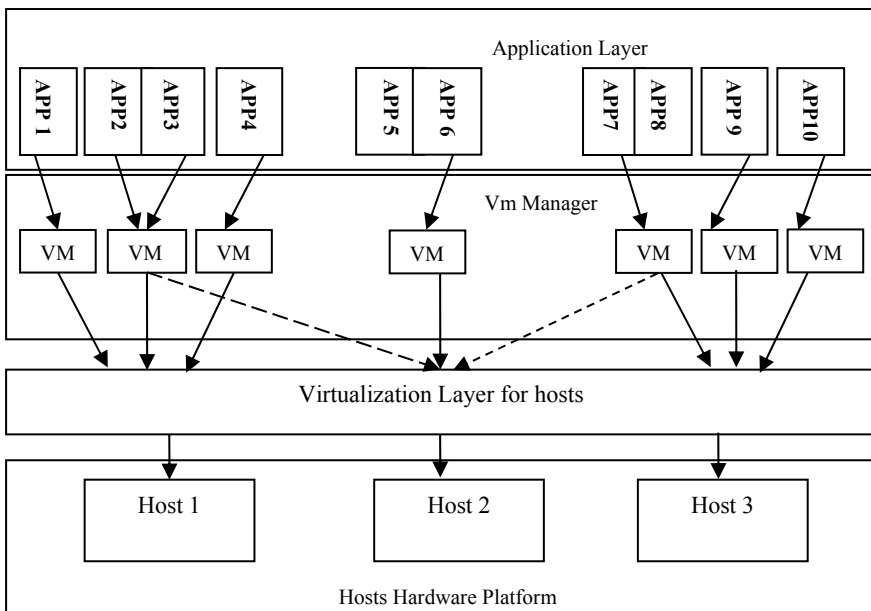


Fig. 1 Layered presentation of hosts, Vms, and applications

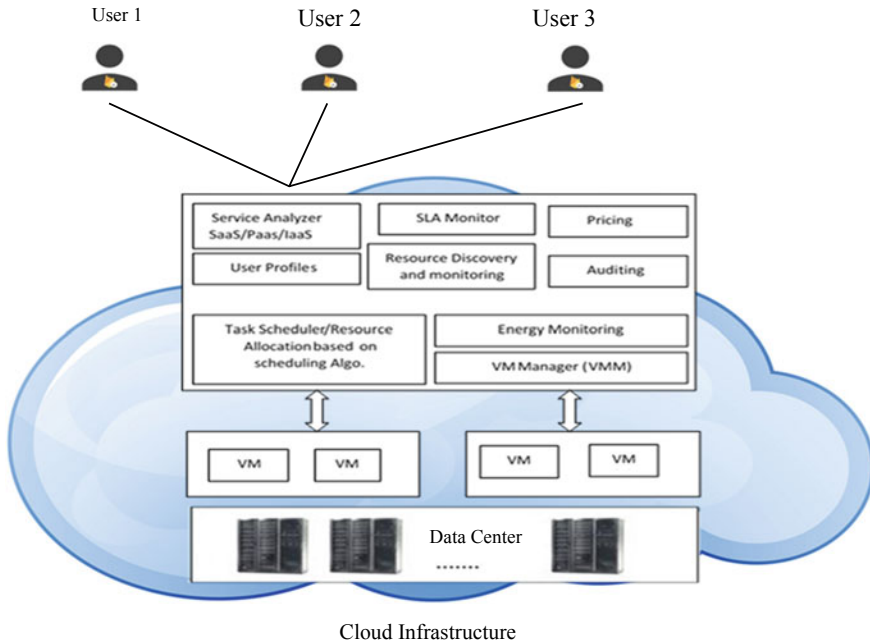
supported by hypervisor like Xen, Vmware, etc. This imitates the hardware resource virtualized in the form of virtual machines and manages these virtual machines hosted by hosts. Each host is capable to accommodate multiple virtual machines depends upon its hardware configuration. Applications are executed on these VMs which can handle multiple applications. So, load-balancing algorithms can be integrated at two levels that one with application scheduler at the application level and other with VM manager at VM level. To balance the load at VM level, a VM image can be migrated from one host to another which is a challenging issue while distributing the load dynamically. This paper focuses on load-balancing approaches integrated at the application level.

## 2 A Scheduling Model for Tasks in Cloud Computing

A data center in the cloud environment is not a single element rather it is a conglomeration of different physical elements that includes physical servers, storage subsystems, routers, networking switches, and firewalls that act as a primary repository for all kind of IT equipments. Each user request for execution in the cloud is directed to the nearby data center by the cloud service provider (CSP). The submitted request is interpreted by the SLA monitor to determine the QoS requirements shown in Fig. 2. Besides that, it also monitors the processing of the submitted request. These requests are processed on networked physical servers possessing a unique identity. Each physical server contained of different virtual machines that share different computing resources of the physical server with the help of hypervisor. These VMs contained within the server have the same hardware and software capabilities as that of a physical server. The user requests are directed to these VMs within the server for processing with the help of the scheduler if different resources are present there to compete based on a deadline [1]. After completion of the task these resources are released and available for the creation of new VMs for serving new user requests. In cloud computing task, scheduling is considered a difficult research problem as tasks should be allocated to an appropriate resource to optimize better performance. To achieve this researchers propose various strategies to overcome the following challenges in task scheduling.

- Cloud computing is based on the sharing of resources; so, there are various users compete for the resources in the environment.
- Resources used in the environment are heterogeneous so these may not perform identically for the given task [2].
- Task scheduler decomposes an application into various tasks so there may be dependencies between the sequence of execution and data communication [3, 4].

To satisfy huge user demands in a cloud environment, the need for load balancer arises. These user demands are heterogeneous and required heterogeneous resources to execute them. The general model of task scheduling shown in Fig. 2 consists of the following components. The task scheduler decomposes the user request into various



**Fig. 2** General model of task scheduling in cloud computing environment

task units  $\{T_1, T_2 \dots T_N\}$ . These task units are submitted to the queues of tasks in the cloud environment for processing. After the resource discovery and monitoring phase of resource management done by the cloud service provider (CSP), several resources are highlighted. The VM manager (VMM) has complete information about the active VMs, local task queue length on each server, and resource availability in different servers. VMM verifies the resource availability for given task units in the cloud environment. VMM sent these tasks to the task scheduler if there are sufficient active VMs to execute the tasks otherwise new VMs are created on servers where resource availability is present. If during the resource monitoring phase done by the service provider to optimize the resources, it is found that resources are underutilized or overprovisioned then the task scheduler will act as a load balancer that will result in better utilization of available resources also aids in energy management. Energy consumed by various resources is monitored by the energy monitor. The general task scheduling model is shown in Fig. 2.

## 2.1 Essential Factors Considered for Scheduling Decision

Following factors should be considered while making a scheduling decision by scheduler:

- **Different types of constraints**—Constraints can be from the consumer side and the cloud service provider side.
  - Constraints can be from the consumer side may be related to a deadline or any budgetary control.
  - Constraints can be enforced by the cloud service provider side regarding maximizing their resource usage to achieve maximum benefits.
- **Optimization criterion**—Optimization criterion is different from the constraints, as constraints are based on the estimated values such as the number of available resources in cloud infrastructure. Besides this in case of optimization criteria, the limits are specified by using the words minimum or maximum such as optimization criterion from the user side can be that task should be completed in a minimum amount of time. An objective function is used to express the optimization criterion, which helps in comparing the computed results with other tested results in the same field.
- **Quality of service (QoS)**—In cloud computing, QoS is used to represent the degree to which a set of inherent characteristics meet the requirements. In cloud computing environment, QoS can be based on various parameters such as performance, security, scalability, and reliability offered by an application and by the infrastructure or platform that hosts it. By invoking a QoS mechanism in the cloud computing environment allows the consumer to specify the requests such as performance, advanced reservation of resources, completion before a given deadline.

## 2.2 *Effect of Performance Metrics on Software Load Balancer*

In cloud environment software load balancers deal with virtual servers and virtual resources. These VMs are prone to performance-related issues due to the overprovisioning of virtual resources in a physical server. These VMs are easily manageable in comparison to physical servers that affects performance issues. Thus, to achieve better results, a load-balancing algorithm considers the following parameters for the stability of the approach. These parameters include minimization of response time, maximizing resource utilization, minimization of makespan, fault tolerance, throughput, etc [5–7]. To optimize these parameters for a load-balancing policy increases the quality of services (QoS) for an approach. Our literature is broadly classified based on some performance metrics that are important for the efficiency of load-balancing policy.

**Response Time(RT):** Response time is defined as the time lag between consumer request and its response. The sum of waiting time, service time, and transmission time are known as response time [8]. The value should be minimum for an effective load-balancing policy, thus the performance of a system inversely proportional to it. It can be calculated by the following formula



$$\text{Response Time } R(t) = \text{Finish Time } F(t) - \text{Submission Time } Sub(t)$$

**Resource Utilization(RU):** Maximizing resource utilization is another important aspect that is to be achieved from a cloud supplier perspective that may be contradictory to other performance parameters such as response time. Also, to achieve this, the usage of limited resources can be a bottleneck. The percentage of resource utilization estimates the cost of the system. The average utilization of resources for a scheduling policy can be calculated as [9]

$$\text{Average Resource Utilization} = \frac{\sum_{(i \in \text{VMs})} \text{Finish Time}(i)}{\text{Makespan } X \text{ number of VMs}}$$

**Makespan(MS):** User request is decomposed into task units for processing. So, makespan is the time when the latest task finishes related to a user request. It is independent of any particular execution order of tasks. It can be defined as [9]

$$\min_{s_j \in \text{Sched}} \left\{ \max_{j \in \text{Jobs}} |F_j| \right\}$$

Here,  $F_j$  notation represents the time when the task  $t$  finalizes, set of all schedules represented by *Sched* and Jobs represent the set of all jobs that are to be scheduled.

**Energy Consumption:** In cloud computing system, energy consumption is the total amount of energy absorbed by all the information and communication technology (ICT) devices interconnected in the system. The system include personal terminals, network components, and local servers. Proper resource management lowers energy consumption. In a cloud environment, each virtual resource has two states idle and active. An idle state of virtual resource consumes 60% of the energy consumption of the active state of that virtual resource [1]. The total energy consumption is the sum of the energy consumed during the active and idle state.

**Average Waiting Time(AWT):** It is the time spent in the waiting queue for a virtual machine to start the execution of the task. It can be defined as

$$\text{AWT} = \frac{\sum_{i=0}^n \text{Start Time } S(t) - \text{Submission Time } Sub(t)}{\text{Total number of tasks}}$$

**Throughput(THP):** It is based upon the number of tasks executed per unit time by a virtual machine. If the throughput value is high, then the system is assumed to have good performance. The value of throughput is inversely proportional to the makespan [10]. If tasks executed on a resource are of the same length, it can be calculated by the following formula (Table 1).

$$THP = \frac{\text{(Task length * number of tasks)}}{\text{Response Time}}$$

All the above research work included in literature considered different performance metrics as a QoS objective that is to be achieved during load balancing such

**Table 1** Load balancing approaches used in a cloud computing environment

Paper Ref.	Detailed approach with considered performance metric	Limitations	Platform
In 2011 Lu et al. [11]	The main objective of JIQ is to provide efficient load balancing for the large system without excessive communication overhead. It is a two-level load-balancing approach that communicates via a data structure named I-queue. The first level of load balancing in JIQ works towards providing an idle server for task assignment with communication off the critical path. On the other hand, the second level of load balancing deals with the allocation of idle servers to the I-queues. <b>Response time</b> is given prime concerned for this approach	<ul style="list-style-type: none"> <li>• Resource utilization is low at moderate load</li> <li>• At high load, there is no idle server, task is allocated to any random server</li> </ul>	The real platform, a TCP connection can be used for message passing between dispatchers and idle processors

(continued)

**Table 1** (continued)

Paper Ref.	Detailed approach with considered performance metric	Limitations	Platform
In 2016 Devi and Uthariaraj [12]	The load-balancing algorithm called improved weighted round-robin for non-preemptive dependent tasks considers some parameters such as interdependency of multiple tasks, task length of each requested task, processing capability of each VM in a data center. The algorithm works in three different stages that are static scheduler, dynamic scheduler, and load balancing. The objective is to minimize the overall <b>completion time</b>	There should be prior knowledge of task length before allocation to a VM	CloudSim
In 2018, Gawali and Shinde [13]	The proposed heuristic approach integrated the modified analytic hierarchy process (MAHP), longest expected processing time preemption (LEPT), bandwidth aware divisible scheduling (BATS) + BAR optimization and used divide and conquer methods to perform task scheduling and task allocation. <b>Response time and turnaround time</b> are minimized for the proposed approach	The proposed approach incorporates various techniques for scheduling the tasks, it is difficult to implement	CloudSim, epigenomics and cybershake scientific workflows used as input data

(continued)

**Table 1** (continued)

Paper Ref.	Detailed approach with considered performance metric	Limitations	Platform
In 2018, Wang et al. [14]	The proposed algorithm is a new variant of JIQ named JIQ-Pod. Scheduler in JIQ-Pod assigns the task to the least loaded server chosen after probing d servers uniformly at random by the scheduler when it finds its I-queue empty. Each dispatcher with an empty I-queue applies the Pod strategy. The author has also derived semi closed-form expressions for the delay in performance in JIQ. <b>Response time</b> is considered as a performance metric	At low load, many resources remain idle that consume more energy	CloudSim
In 2015 Pan and Chen [15]	The proposed approach considers the features of complex networks and utilizes it to establish a task—resource allocation model. The system model considers splitting of task scheduling approach into three layers viz. layer of task requirement, the layer of resource management, the layer of task execution. <b>Resource utilization</b> is considered a prime parameter	The proposed approach has not been analyzed for the population of different sizes	CloudSim

(continued)

**Table 1** (continued)

Paper Ref.	Detailed approach with considered performance metric	Limitations	Platform
In 2017 Kumari and Jain [16]	The proposed approach involves the scheduling of tasks based on the particle swarm optimization (PSO) approach using the max-min algorithm. The main consideration is given to makespan and CPU utilization as a QoS parameter. Experimental results show that there is an improvement of <b>makespan</b> by an average of 5.01% over the existing scheduling algorithm and <b>resource utilization</b> is improved by an average of 3.63% over the methods adopted previously	<ul style="list-style-type: none"> <li>• Homogeneous environment</li> <li>• Not tested for the large population</li> </ul>	CloudSim
In 2016 Ismail and Fardoun [17]	The objective of the proposed approach is to increase the efficiency of the application and to reduce the energy consumption of the virtual resources in the datacenter. It solves the non-linear programming model and takes scheduling decisions. Energy consumption is measured on different workloads on a real platform. The proposed approaches increase <b>resource utilization</b> and <b>manage energy</b> efficiently	Scalability is low	The experiments conducted on a real platform

(continued)

**Table 1** (continued)

Paper Ref.	Detailed approach with considered performance metric	Limitations	Platform
<p>In 2017 Yang et al. [18]</p>	<p>The proposed task scheduling approach considers the reliability of a balanced task based on game theory. Based on that, a balanced scheduling algorithm has been proposed that is applied to multiple computing nodes. The proposed algorithm finds the utility function by evaluating steady-state nodes ability, and then a cooperative game model is designed on task scheduling problems in the cloud environment. <b>Response time</b> and <b>reliability</b> of nodes are considered as a performance metric</p>	<ul style="list-style-type: none"> <li>• Homogeneous environment</li> <li>• Cost is not considered</li> </ul>	<p>Experiments conducted on real platform</p>
<p>In 2019 Lu and Sun [19]</p>	<p>The proposed algorithm considered the problem of energy consumption in green cloud computing of prime concern and used a resource-aware load-balancing algorithm to solve it. Problem formulation is done based on the combinatorial optimization problem that seeks to optimize the load balancing by minimizing energy consumption using resource utilization. The proposed approaches maximize <b>resource utilization</b> and <b>energy-efficient</b></p>	<p>Homogeneous environment</p>	<p>CloudSim</p>

(continued)

**Table 1** (continued)

Paper Ref.	Detailed approach with considered performance metric	Limitations	Platform
In 2013 Ramezani et al. [20]	It is a task-based scheduling technique based on particle swarm optimization (TBSLB-PSO). VMs are migrated for balancing the load. It minimizes the <b>response time and cost</b> . Memory is more efficiently used	<ul style="list-style-type: none"> <li>• Homogeneous environment</li> <li>• Not suitable for dependent tasks</li> </ul>	CloudSim
In 2016 Vasudevan et al. [21]	A dynamic load-balancing algorithm based on honeybee used to reduce the makespan in the heterogeneous environment of the cloud. The proposed approach minimizes the <b>makespan</b>	Scalability is low	CloudSim
In 2016 Ibrahim et al. [22]	In this algorithm, total processing power of all the available resources (VMs) and total requested processing power by the user's tasks has been calculated. The processing power of each VM is defined based on Amazon EC2 and Google pricing models. After that, each VM is allocated to a group of user's tasks based on the ratio of its needed power relative to the total processing power of all VMs. The proposed approach minimizes the <b>makespan and cost</b>	<ul style="list-style-type: none"> <li>• Homogeneous</li> <li>• Scalability is low</li> </ul>	CloudSim

(continued)

**Table 1** (continued)

Paper Ref.	Detailed approach with considered performance metric	Limitations	Platform
In 2016 Jain and Kumar [23]	The authors proposed a two-level scheduler by integrating join shortest queue approach and join idle queue approach. It has been stated that the scheduler is distributed in nature as it uses the characteristics of the two different schedulers. <b>Response time</b> is considered a prime parameter	<ul style="list-style-type: none"> <li>• Homogeneous environment</li> <li>• Communication overheads are more</li> </ul>	Cloud analyst

as response time, resource utilization, and energy management are of prime concern. The load-balancing approaches have been proposed in literature considering one performance metric or the other but to optimize these parameters in a multi-objective way is a challenging issue.

### 2.3 Multi-objective Optimization Load-Balancing Approaches

In the previous literature section, as it has been described that the scheduling of tasks in the cloud environment is multi-objective generally. So, while considering the optimization criteria for these approaches, it should be considered likewise that there is a good trade-off among these performance parameters positively [24]. Various approaches exist in multi-objective optimization to deal with multi-criteria problems. These are the following.

**Hierarchical approach:** This approach depends upon the type of application that is dealt with in the cloud environment. For instance, in the case of high-performance computing, one optimization criterion that is makespan is more important than the other criterion response time, or the situation is reversed if user requirements are more concerned. For example, if a set of optimization criteria is considered, i.e.,  $t_i, 1 \leq i \leq N$ , then for the hierarchical approach, these criteria are categorized by their priority. In such a way that if criterion  $t_i$  is less important than the criterion  $t_j$  then criterion value can not be varied for  $t_i$  while optimizing it for  $t_j$ . Such type



of approaches has the limitation that one should have prior establish the priority of criterion and at one time only one criterion can be optimized.

**Simultaneous Approach:** In such types of approaches, improvement in one performance criterion deteriorates the other performance criterion during optimal planning. Computation cost can be increased during dealing with many optimization criteria at the same time. Such types of problems can be addressed using the Pareto optimization theory [9]. The weighted sum approach is a type of Pareto optimization technique in which a single aggregated function is used after combining different optimization criteria.

### 2.4 Classification of Load-Balancing Algorithms

Load-balancing algorithms can be broadly divided into two categories based on the current state of the VM. If there is prior information regarding the current state of the VM after that load is distributed among the other nodes, then such type of allocation policy is known as a static strategy for allocating the load on the system. Besides that, if there is no information for distributing the load, it is distributed based on the current state of the system then it is called a dynamic strategy. Load-balancing algorithms are developed to balance the system at each node to make the system stable and reliable by assuring user satisfaction. The following Table 2 shows the different types of load-balancing algorithms used in the system.

**Table 2** Different types of load-balancing algorithm used in cloud environment

Algorithm name	Detailed approach	Type of strategy
Min-Min	In this approach task with least length having the minimum execution time is chosen to execute on a VM having the minimum completion time that is determined based on VM capacity. It means a task is given to the highest priority having the least execution time. It is used for a distributed system small in scale	Static
Max-Min	The approach is similar to Min-Min. The task with the largest length having the maximum execution time is given to the highest priority to execute on a chosen cloud resource having the minimum completion time. That is a task with maximum execution time is given the highest priority	Static

(continued)

**Table 2** (continued)

Algorithm name	Detailed approach	Type of strategy
Greedy Randomized Adaptive Search Procedure (GRASP)	It is a meta-heuristic search technique in which a solution occurs at each step and that solution is reserved for the final schedule. The termination of the GRASP is based on a certain condition and it gets terminated when the specified condition is satisfied	Static
Genetic algorithm	This approach uses biological techniques such as mutation, selection, inheritance, and crossover. These algorithms use some fitness value based on the chromosome to improve and maintain the population. A candidate solution is generated from a primary randomly generated population of chromosomes. The fitness of each chromosome is calculated within the population. Based on the fitness value, multiple people are selected from the primary population to generate a new population that is used for the second iteration. When a satisfactory fitness level is reached, algorithm is terminated	Dynamic
Ant Colony Optimization (ACO)	It is a nature-inspired search technique; it is commonly used for the large population. In the ACO algorithm. This technique is based on the random manner in which the ants search their food. During the food searching process whenever ants find a food source it evaluates the quantity and quality of food and taking back some part of the food to its nest. When ants get back to the same food source, the ant discharges a pheromone trail on the ground to give the information to the other ants about the food source	Dynamic
Particle Swarm Optimization (PSO)	It is a meta-heuristic search technique based on the social behavior of the particles. It optimizes a problem iteratively by finding a candidate solution. It makes no assumptions for the problem being optimized and searches for large spaces for the candidate solution	Dynamic

### 3 Conclusion

In this research paper, several aspects of cloud computing have been reviewed related to task scheduling problem in the cloud environment, that includes the concept optimization criteria, optimizing multi-objective load balancing approaches, different performance metrics. The literature reviewed in this paper concludes that these approaches are multi-objective, and there is a need to optimize these performance metrics more efficiently. Various solutions to these task scheduling problems have been discussed by using heuristic and meta-heuristic search techniques considering the nature and space of the problem. This paper aims to impart the complexity of scheduling a task to an appropriate resource during load balancing by considering various optimization criteria. As a future scope, more research efforts are required about modeling of these performance metrics to make it more predictive and responsive by changing the one or two decision variables.

### References

1. Mishra SK, Sahoo B, Parida PP (2018) Load balancing in cloud computing: a big picture. *J King Saud Univ-Comput Inform Sci* 32(2):149–158
2. Jain A, Kumar R (2017) Scalable load balancing approach for cloud environment. *Int J Eng Technol Innov* 7(4):292–307
3. Reddy VK, Reddy LSS (2012) A survey of various task scheduling algorithms in cloud environment. *Glob J Eng Appl Sci* 2(1):847–853
4. Sharma M, Kumar R, Jain A (2019) Implementation of various load-balancing approaches for cloud computing using CloudSim. *J Comput Theor Nanosci* 16(9):3974–3980
5. Sharma M, Jain A, Kumar R (2020) A proficient approach for load balancing in cloud computing—join minimum loaded queue. *IJDAR* 11(1):12–36
6. Sharma M, Kumar R, Jain A (2019) A system of distributed Join Minimum Loaded Queue (JMLQ) for load balancing in cloud environment. Patent application No 201911007589, pp 12780, India
7. Sharma M, Kumar R, Jain A (2019) A system of Quality of Service enabled (QoS) Join Minimum Loaded Queue (JMLQ) for cloud computing environment. Patent application No 201911039375, pp 51328, India
8. Milani AS, Navimipour NJ (2016) Load balancing mechanisms and techniques in the cloud environments: systematic literature review and future trends. *J Netw Comput Appl* 71:86–98
9. Khafa F, Abraham A (2008) Meta-heuristics for grid scheduling problems. In: *Metaheuristics for scheduling in distributed computing environments*. Springer, Berlin, Heidelberg, pp 1–37
10. Ashalatha R, Agarkhed J (2015) Dynamic load balancing methods for resource optimization in cloud computing environment. In: *Annual IEEE India conference (INDICON)*, IEEE, pp 1–6
11. Lu Y, Xie Q, Kliot G, Geller A, Larus JR, Greenberg A (2011) Join-Idle-Queue: a novel load balancing algorithm for dynamically scalable web services. *Perform Eval* 68(11):1056–1071
12. Devi DC, Uthariaraj VR (2016) Load balancing in cloud computing environment using improved weighted round robin algorithm for nonpreemptive dependent tasks. *Sci World J*. <https://doi.org/10.1155/2016/3896065>
13. Gawali MB, Shinde SK (2018) Task scheduling and resource allocation in cloud computing using a heuristic approach. *J Cloud Comput* 7(1). <https://doi.org/10.1186/s13677-018-0105-8>
14. Wang C, Feng C, Cheng J (2018) Distributed join-the-idle-queue for low latency cloud services. *IEEE/ACM Trans Networking* 26(5):2309–2319

15. Pan K, Chen J (2015) Load balancing in cloud computing environment based on an improved particle swarm optimization. In: 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), IEEE, pp 595–598
16. Kumari R, Jain A (2017) An efficient resource utilization based integrated task scheduling algorithm. In: 4th international conference on Signal Processing and Integrated Networks (SPIN), IEEE, pp 519–523
17. Ismail L, Fardoun A (2016) Eats: energy-aware tasks scheduling in cloud computing systems. *Procedia Comput Sci* 83:870–877
18. Yang J, Jiang B, Lv Z, Choo KKR (2017) A task scheduling algorithm considering game theory designed for energy management in cloud computing. *Future Gener Comput Syst*
19. Lu Y, Sun N (2019) An effective task scheduling algorithm based on dynamic energy management and efficient resource utilization in green cloud computing environment. *Cluster Comput* 22(1):513–520
20. Ramezani F, Lu J, Hussain FK (2014) Task-based system load balancing in cloud computing using particle swarm optimization. *Int J Parallel Prog* 42(5):739–754
21. Vasudevan SK, Anandaram S, Menon AJ, Aravindh A (2016) A novel improved honey bee based load balancing technique in cloud computing environment. *Asian J Inf Technol* 15(9):1425–1430
22. Ibrahim E, El-Bahnasawy NA, Omara FA (2016) Task scheduling algorithm in cloud computing environment based on cloud pricing models. In: *World Symposium on Computer Applications and Research (WSCAR)*, IEEE, pp 65–71
23. Jain A, Kumar R (2016) A multi stage load balancing technique for cloud environment. In: *International Conference on Information Communication and Embedded Systems (ICICES)*, IEEE, pp 1–7
24. Shi Y, Qian K (2019) LBMM: a load balancing based task scheduling algorithm for cloud. In: *IEEE Future of Information and Communication Conference (FICC)*, San Francisco

# An Energy-Efficient Routing Protocol Using Threshold Hierarchy for Heterogeneous Wireless Sensor Network



Ashok Kumar Rai and A. K. Daniel

**Abstract** Wireless sensor network have self-configured network which is consist of sensors and base station. The deployments of sensors are done in the required area to collect data and transmit it to the base station. The proposed energy-efficient routing protocol using threshold hierarchy for heterogeneous wireless sensor network (EEPH) protocol considers two parameters as distance and energy. The parameters are defined as the ratio of residual energy of current node to the total energy of all nodes. The distance is termed as the ratio of current node distance to the summation of distances of each node at that level from base station. The proposed protocol used three level of threshold hierarchy for different level of cluster-head selection process. The simulation result shows better performance and improved network lifetime compare to threshold-sensitive stable election protocol (TSEP).

**Keywords** Cluster head · Distance · Residual energy · Wireless sensor network · TSEP · Threshold hierarchy

## 1 Introduction

Wireless sensor network (WSNs) are widely used in various applications such as environmental monitoring, military application, agriculture sector, etc. They consist of thousands of sensor nodes deployed in sensing field and these sensors collects the data from field and transmits into the base station (BS). According to the requirement, the BS may be stable or moving. Different protocols have been proposed by the researchers to improve the lifetime of network and reduce the cost of network maintenance/overall cost of network. The rest of paper is organized as follows. Section 2

---

A. K. Rai (✉) · A. K. Daniel

Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh, India  
e-mail: [ashok7086@gmail.com](mailto:ashok7086@gmail.com)

A. K. Daniel

e-mail: [danielak@rediffmail.com](mailto:danielak@rediffmail.com)

explains the related work. Section 3 the proposed protocol. Section 4 simulation result and last section includes conclusion.

## 2 Related Work

It is essential in wireless sensor networks to enhance nodes residual energy and network lifetime by utilizing optimum energy conservation techniques. The area is divided in different regions on the basis of longitudinal distance to the sink. The regions constitute of cluster heads for transferring data to sink via single/multi-hop techniques. The static clustering is used to reduce overhead and enhances the life time of network. The selections of CHs are performed by using maximum residual energy of nodes in the given clusters. There are following related work are performed by researchers.

Ravindra singh et al. introduced a protocol in which position-based multiple approach is applied for the partition of network into different level. The protocol transmit data through the optimum route made by the position-based technique, and CH selection is done by using Fuzzy Logic [1].

Loscri et al. introduced a protocol in which the cluster is divided into two levels, i.e., level one and level two. These two levels of cluster have distinct threshold energy for the selection of CH (Cluster Head) which minimize the energy consumption required for the transmission of data to the BS [2].

Lindsey et al. introduced a protocol in which sensor nodes are forming a chain-based system for the data transfer from node to the sink. This protocol formation chain is done by connecting the nodes from its neighbor node. Here, global information of each node is maintained [3] to attain desired performance.

Vipul Narayan et al. introduced a protocol for overlapping problem of node in a given area is resolved. This protocol uses radius and residual energy of node which improve the coverage of network and its life time [4].

Chengfa Li, Mao Ye et al. introduced a protocol where energy consumption of the network is balanced. The nodes which are not involved in any cluster of the network are termed as isolated node [5] and it is not considered for the network energy analysis which helps to identifies the isolated nodes.

B. Manzoor et al. introduced a protocol where network is divided into four quadrants which give the better coverage areas, increase the stability period, and improves network lifetime [6].

Feng Ming Hu et al. introduced a protocol which further enhancement of PEGASIS. Unlike PEGASIS protocol, CH is not changed in every round which reduces the energy consumption for CH selection for each round. This is done in two phases as initialization phase and data collection and transmission phase [7].

Yaqiong Wang et al. introduced a case-based reasoning (CBR) protocol. The BS sends a message to all nodes and ask whether it is first round, if reply message

is 'yes' then clustering and CH selection is done otherwise CBR decide whether re-clustering or just new CH selection is done into the existing cluster [8].

M. Singh et al. introduced a protocol which uses two fuzzy variables as energy and residual energy. This protocol have a timer which is activated and broadcast the advertising message to all nodes. The node having maximum value of fuzzy becomes a CH [9].

Ashok Kumar Rai et al. introduced a protocol for homogeneous network in which both residual energy and distance of node from BS is considered. These two parameters reduced the threshold energy for selection of CH and increase the network life time [10].

Vipul Narayan et al. introduced a protocol for two-level heterogeneous network in which sleep and wakeup concept is applied on node which reduce the dissipation of energy and improves the lifetime of network [11].

Li Qing, Qingxin et al. introduced a protocol for two-level heterogeneous WSN. CHs selection for this protocol is depend upon the probability based on ratio of residual energy of each node and network average energy. With this approach, lifetime of network is increased [12].

Brahim Elbhiri et al. introduced a protocol for heterogeneous network in WSN which is based on efficient energy clustering scheme. In this protocol, both initial and residual energy of protocol are considered as a probability function for CH selection and to obtain the global knowledge of network [13].

Arati Manjeshwar et al. introduced a protocol which reduces the number of transmission for non-critical data. The protocol works on both types of protocol reactive and proactive by changing the threshold value. This approach improves the life time of WSN [14].

Yan Yu et al. introduced a protocol use heuristic approach for the selection of neighbor node to transmit the packets towards the target region. This approach balances the consumption of energy in the network which improves its lifetime [15].

Tokuya Inagaki et al. introduced a protocol where massive number of nodes is deploying in observed area, and only few numbers of nodes are selected for transmission of data. Other nodes which are not involved for the transmission of data are in sleep mode. In this way, this protocol saves the energy consumption of network [16].

Pooja Chaturvedi et al. introduced a protocol which has set of nodes in active state such that target area is monitored by at least one active sensor. Nodes with lower coverage probability are monitoring the trust level of neighborhood node [17]. In another work. the same author introduced a protocol in which node scheduling is done which is depend upon the trust value and coverage probability of node for target coverage area. It improves the coverage and connectivity of nodes [18]. Similarly, the reason for the sensor holes and its effect on the network in term of its life time and coverage probability [19] is discussed. A protocol for target coverage area which is done on the basis of the scheduling of node is introduced in literature [20]. Cluster mechanism of protocol is depending upon the three factors, i.e., residual energy, number of nodes selected for the CH and distance.

Vipul Narayan et al. introduced a protocol with multilevel clustering technique which give the solution for coverage and connecting problem due to cropping of animals [21].

Lou introduced multipath routing protocol where multiple nodes are involved to make a disjoint path towards the sink node. Through branch ware flooding technique, sink node generates a route update message for other nodes and built a minimum spanning tree. With this approach, the shortest route for the packet delivery from node to sink is discovered and transmission energy is minimized which improves the lifetime of network [22].

G. Smaragdavis et al. introduced a protocol in which two levels heterogeneity concepts is used. In first level nodes are called normal node and in second level nodes are advanced node. Probability to select CH in Normal node and advance node are distinct which reduce the consumption of energy [23].

A. Kashaf et al. introduced a protocol having three level of heterogeneity. First level nodes are called normal node, second are called intermediate node, and in third level are called advanced node. Like SEP protocol, it also has distinct threshold energy for distinct levels. Hard and soft threshold concept are also consider for this protocol [24].

### 3 Proposed Protocol

Let us consider a network and is divided into three zones as standard node area, intermediate node area headway node area. The nodes in standard level have low energy as compare to the nodes in intermediate and headway level. The nodes belong to intermediate level have low energy as compare to the headway node and less than standard node. The nodes belong to headway level have more energy as compare to both standard and intermediate node The protocol consider having two parameter for the threshold energy. First parameter is the ratio of residual energy of current node to the total energy of the level, and second parameter is the distance ratio of current node to the summation of distance of each node of the level from Base station.

#### Proposed Protocol for CH Selection

Let us consider a network of  $100 \times 100 \text{ m}^2$  having  $x$  nodes deployed in the given area. The total area is divided into three regions having different level of heterogeneity with distinct thresholds values. CH selection depends upon the threshold energy of different level with parameter  $E$  and  $D$  where  $E$  is residual energy of node and  $D$  is distance of node from BS. The following protocol is proposed.



<p>Given Parameter  <math>x</math> is total number of nodes  <math>z</math> is proportion of intermediate nodes to <math>x</math>  <math>y</math> is proportion of headway nodes to <math>x</math>                  Energy of intermediate nodes is <math>\mu</math> times more than energy of standard nodes                  Energy of headway nodes is <math>\alpha</math> times greater than energy of normal nodes                  Where <math>\mu = \alpha/2</math>                  Assign the level from 1 to <math>z</math> for all nodes  <math>N_s</math> = No. of Alive Nodes in Standard Level from 1 to <math>n</math>  <math>N_i</math> = No. of Alive Nodes in Intermediate Level from <math>n + 1</math> to <math>m</math>  <math>N_h</math> = No. of Alive Nodes in Headway Level <math>m + 1</math> to <math>z</math>  <math>G_s</math> = Set of standard nodes  <math>G_i</math> = Set of intermediate nodes  <math>G_h</math> = Set of headway nodes                  CH = CH                  CH.<math>E_s</math> = CH Energy in Standard level                  CH.<math>E_i</math> = CH Energy in intermediate level                  CH.<math>E_h</math> = CH Energy in headway level  <math>E_r</math> = Energy of Current Node  <math>E_i</math> = Energy of <math>i</math>th Node  <math>D_r</math> = Current node distance from BS.  <math>D_i</math> = <math>i</math>th Node distance from BS.    <math>P_{std}</math> = Probability of CH selection in Standard Level (Level 1)  <math>P_{int}</math> = Probability of CH selection in intermediate Level (Level 2)  <math>P_{hdw}</math> = Probability of CH selection in Standard Level (Level 3)                    Threshold Energy for level 1 (Standard level)  <math>T_s(n) = P_{std} / [1 - P_{std} \times (\text{mod}(r, 1/P_{std})) \times (E_r / \sum_{i=0}^m E_i) \times (D_r / \sum_{i=0}^m D_i)]</math>                  Where  <math>P_{std} = P_{opt} / (1 + y \cdot \alpha + z \cdot \mu)</math>                  Threshold Energy for level 2 (Intermediate level)  <math>T_i(n) = P_{int} / [1 - P_{int} \times (\text{mod}(r, 1/P_{int})) \times (E_r / \sum_{i=m+1}^m E_i) \times (D_r / \sum_{i=0}^m D_i)]</math>                  Where  <math>P_{int} = P_{opt} (1 + \mu) / (1 + y \cdot \alpha + z \cdot \mu)</math>                  Threshold Energy for level 3 (Headway level)</p>	<p><math>T_h(n) = P_{hdw} / [1 - P_{hdw} \times (\text{mod}(r, 1/P_{hdw})) \times (E_r / \sum_{i=m+1}^z E_i) \times (D_r / \sum_{i=m+1}^z D_i)]</math>                  Where <math>P_{hdw} = P_{opt} (1 + \alpha) / (1 + y \cdot \alpha + z \cdot \mu)</math>                  Depend upon Energy .Network is divided into Three levels                  If Alive nodes <math>&gt; 0</math> ,                      Then                          New CH<sub>s</sub> selection and re Clustering is done                      Else                          Exit                  End                  /* For standard Node*/                  For Node 1: <math>N_s</math>                      If <math>n \in G_s</math>                      if                          Residual energy of randomly selected node is greater than <math>T_s(\text{new})</math>                          It becomes CH<sub>s</sub>'                      Else                          /* Formation of clusters */                      Join nearest CH<sub>s</sub>                  End                  /* For intermediate Node*/                  For Node 1: <math>N_i</math>                      If <math>n \in G_i</math>                      if                          Residual energy of randomly selected node is greater than <math>T_i(\text{new})</math>                          It becomes CH<sub>i</sub>'                      Else                          /* Formation of clusters */                      Join nearest CH<sub>i</sub>                  End                  For headway Node*/                  For Node 1: <math>N_h</math>                      If                          <math>n \in G_h</math>                      if                          Residual energy of randomly selected node is greater than <math>T_h(\text{new})</math>                          It becomes CH<sub>h</sub>'                      Else                          /* Formation of clusters */                      Join nearest CH<sub>h</sub>                  End</p>
---	--

## 4 Simulation Result

Based on the analytical work and the proposed protocol is simulated using MATLAB and validate the execution of proposed protocol. Heterogeneous cluster size of WSN

is simulated in the field of dimension  $100 \times 100 \text{ m}^2$  and number of sensor nodes  $n = 100$  deployed in the given area for 10,000, 15,000 and 20,000 rounds of transmission of packets. Node deployment in different region are based on their energy parameter as shown in Table 1.

The three-level heterogeneous nodes are deploying in fixed zone to give proper utilization of whole area. The approach gives efficient utilization of power in the network. This concept improves the throughput and life expectancy of the network. The CH directly sends the data to BS.

The Fig. 1 shows the comparison between EEPH protocol and TSEP protocols in term of alive nodes for 10,000 rounds. The Fig. 2 shows the comparison between EEPH protocol and TSEP protocols in term of numbers of dead nodes for 10,000 rounds.

The Fig. 3 shows the comparison between EEPH protocol and TSEP protocols in term of packets send to BS for 10,000 rounds.

The Fig. 4 shows the comparison between EEPH protocols and TSEP protocols in term of Residual Energy for 10,000 rounds.

The Fig. 5 shows the comparison between EEPH protocols and TSEP protocols in term of Average Energy Consumed for 10,000 rounds.

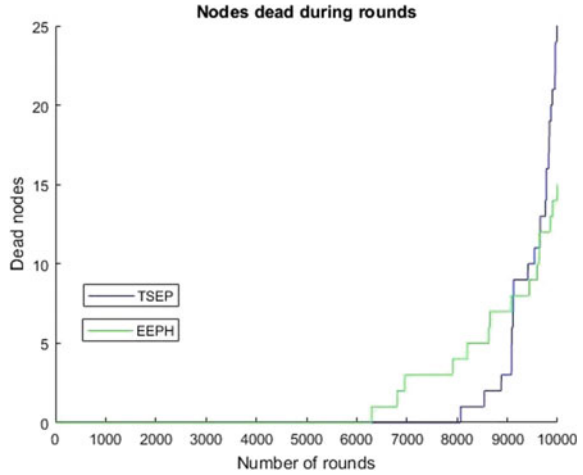
The alive nodes comparisons between EEPH protocol and TSEP protocol for 10,000 rounds shown in Table 2. At 8000 rounds the performance of TSEP protocol is better than EEHP but at higher rounds, i.e., more than 8000 round EEHP protocol performance is better.

The dead nodes comparisons between EEPH protocol and TSEP protocol for 10,000 rounds are shown in Table 3. At 8000 round the performance of TSEP protocol

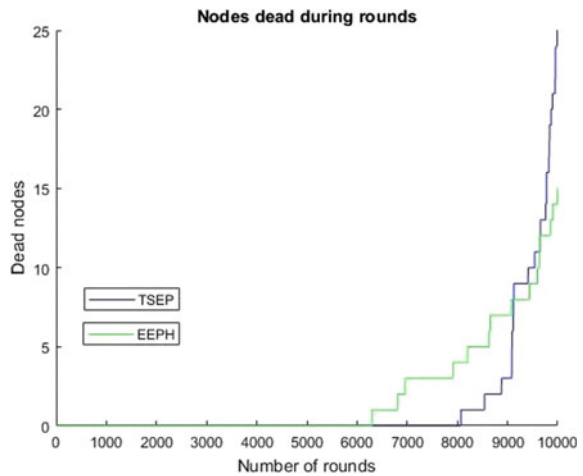
**Table 1** Parameter specifications

Parameter	Specification
Initial energy ( $E_o$ )	1.5
Number of nodes ( $x$ )	100
$E_{elec}$	$50 * 10^{(-9)} \text{ j}$
$P_{opt}$	0.1
Data bits	4000
Sink.x	100 m
Sink.y	100 m
Transmission energy (ETX)	$50 * 10^{(-9)} \text{ j}$
Energy for receiving data (ERX)	$50 * 10^{(-9)} \text{ j}$
EDA	$50 * 10^{(-9)} \text{ j}$
Efs	$10 * 10^{(-12)} \text{ j}$
Emp	$0.0013 * 10^{(-12)} \text{ j}$
A	1
Y	0.1

**Fig. 1** Number of alive nodes for 10,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Fig. 2** Number of dead for 10,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



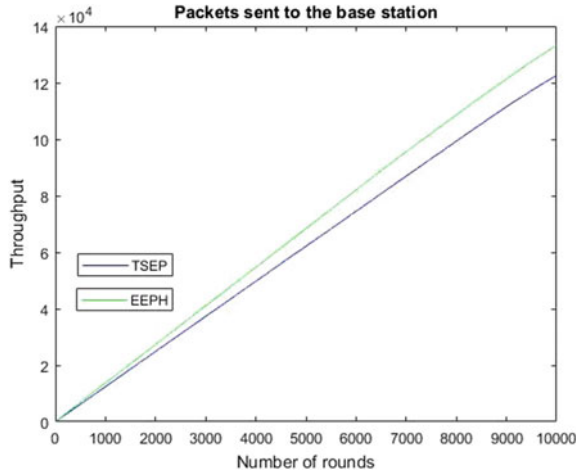
is better than EEHP but at higher rounds, i.e., more than 8000 round EEHP protocol performance is better.

The comparison between TSEP protocol and EEHP protocol in term of packets send to BS are more by EEHP protocol are shown in Table 4 for 10,000 rounds

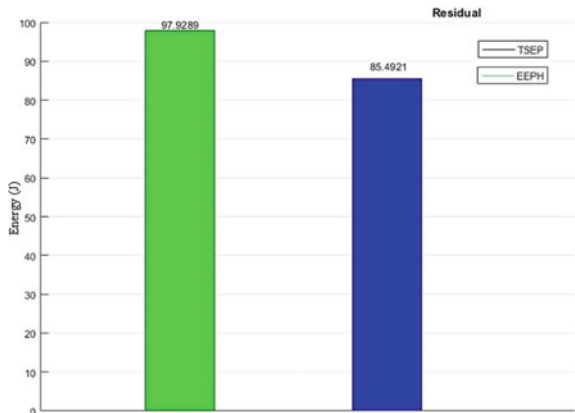
The comparison between EEHP protocols and TSEP protocols in term of numbers of alive nodes for 15,000 rounds is shown in Fig. 6; similarly, the comparison between EEHP protocols and TSEP protocols in term of numbers of dead nodes for 15,000 rounds is shown in Fig. 7.

The Fig. 8 shows the comparison between EEHP and TSEP protocols in term of number of packets send to BS for 15,000 rounds.

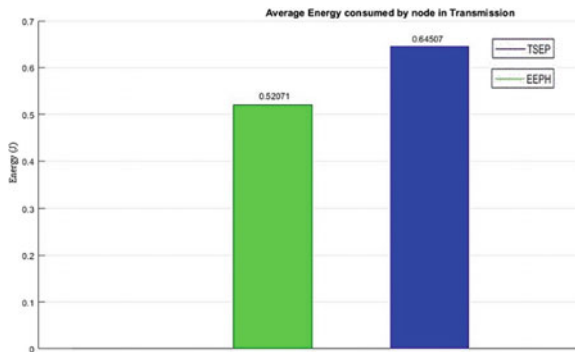
**Fig. 3** Packets send to base station for 10,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Fig. 4** Residual energy for 10,000 rounds  $(100 \times 100) \text{ m}^2$



**Fig. 5** Average energy consumed for 10,000 rounds in  $(100 \times 100) \text{ m}^2$



**Table 2** Number of alive nodes for 10,000 rounds

Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 10,000		
Number of rounds	TSEP	EEPH
	Number of alive nodes	Number of alive nodes
7000	100	97
8000	100	96
9000	97	93
10,000	76	86

**Table 3** Number of dead nodes for 10,000 rounds

Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 10,000		
Number of rounds	TSEP	EEPH
	Number of dead nodes	Number of dead nodes
7000	00	03
8000	00	04
9000	03	07
10,000	24	14

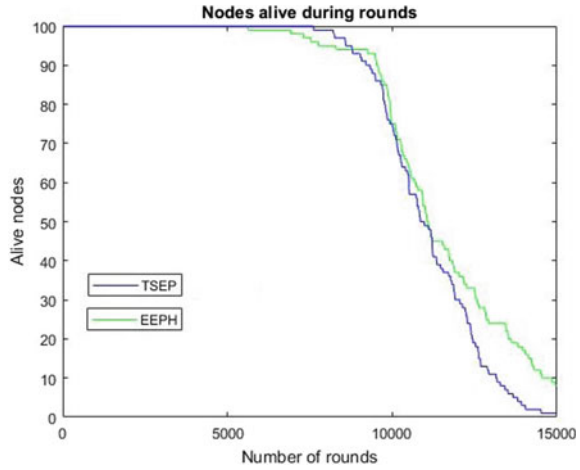
**Table 4** Number of packets send to BS for 10,000 rounds

Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 10,000		
Number of rounds	TSEP	EEPH
	Number of packets send to BS	Number of packets send to BS
7000	$8.5 \times 10^4$	$9.4 \times 10^4$
8000	$9.9 \times 10^4$	$10.95 \times 10^4$
9000	$11 \times 10^4$	$12.1 \times 10^4$
10,000	$12 \times 10^4$	$13.2 \times 10^4$

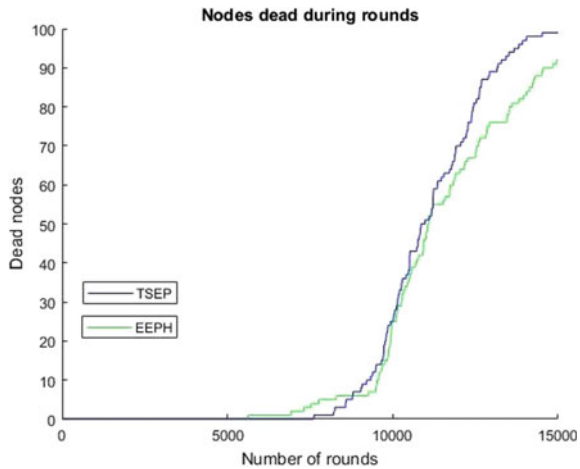
The comparison between EEPH and TSEP protocols in term of Residual Energy for 15,000 rounds is shown in Fig. 9.

The comparison between EEPH and TSEP protocols in term of Average Energy Consumed for 15,000 round is shown in Fig. 10.

**Fig. 6** Number of alive nodes for 15,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Fig. 7** Number of dead nodes for 15,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



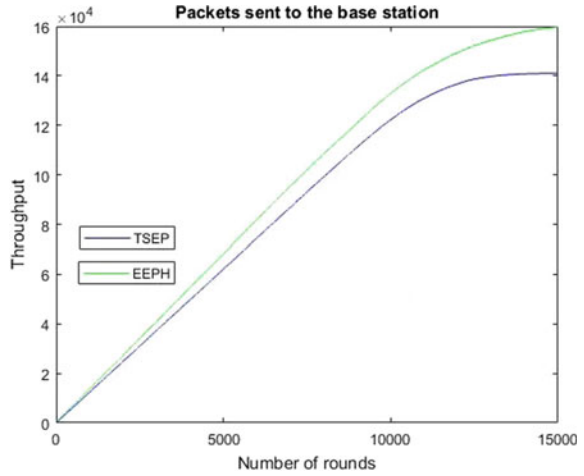
The alive nodes comparisons are done between EEHP protocol and TSEP protocol for 15,000 rounds shown in Table 5. At 7000 rounds the performance of TSEP protocol is better than EEHP but at higher rounds, i.e., more than 7000 round EEHP protocol performance is better.

The dead nodes comparisons are done between EEHP protocol and TSEP protocol for 15,000 rounds shown in Table 6. At 7000 round, the performance of TSEP protocol is better than EEHP but at higher rounds, i.e., more than 7000 round EEHP protocol performance is better.

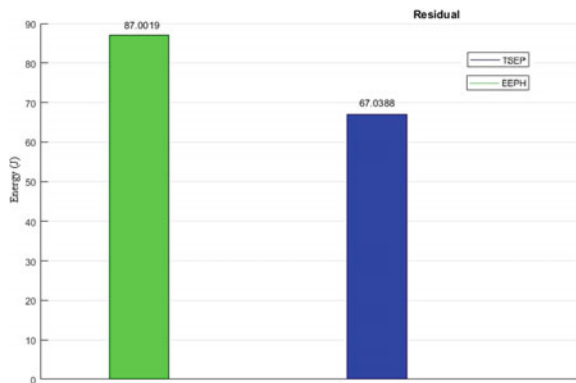
The comparison between TSEP protocol and EEHP protocol in term of packets send to BS are more by EEHP protocol shown in Table 7 for 15,000 rounds.

The comparison between EEHP protocol and TSEP protocols in term of a live node for 20,000 rounds are shown in Fig. 11 similarly the comparison between EEHP

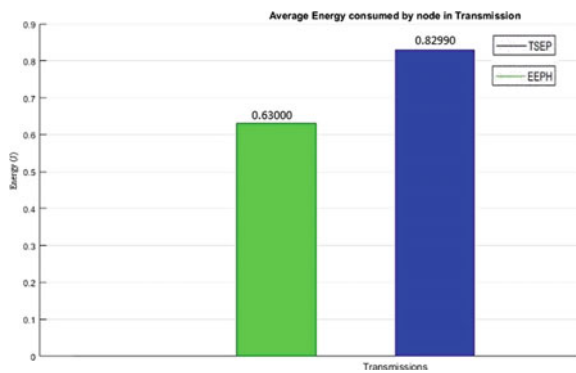
**Fig. 8** Packets send to base station for 15,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Fig. 9** Residual energy for 15,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Fig. 10** Average energy consumed for 15,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Table 5** Number of alive nodes for 15,000 rounds

Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 15,000		
Number of rounds	TSEP	EEPH
	Number of alive nodes	Number of alive nodes
7000	100	99
9000	96	95
11,000	50	54
13,000	12	25
14,000	01	19
15,000	00	09

**Table 6** Number of dead nodes for 15,000 rounds

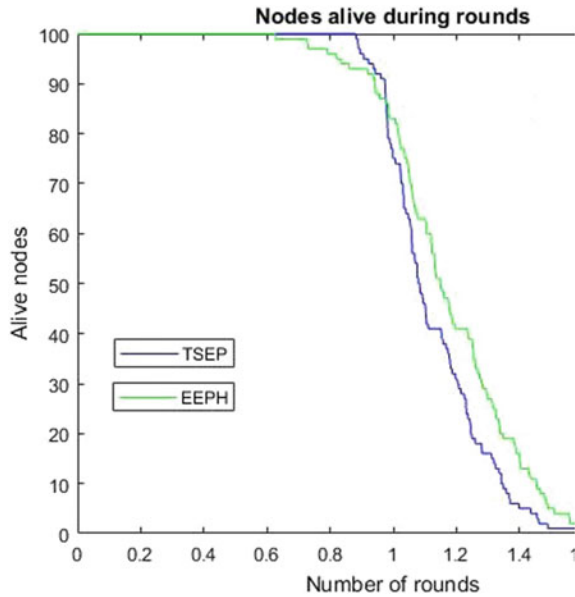
Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 15,000		
Number of rounds	TSEP	EEPH
	Number of dead nodes	Number of dead nodes
7000	00	01
9000	04	05
11,000	50	46
13,000	88	75
14,000	99	81
15,000	100	91

**Table 7** Number of packets send to BS for 15,000 rounds

Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 15,000		
Number of rounds	TSEP	EEPH
	Number of packets send to BS	Number of packets send to BS
7000	$8.4 \times 10^4$	$9.2 \times 10^4$
9000	$11 \times 10^4$	$12 \times 10^4$
11,000	$13 \times 10^4$	$14.1 \times 10^4$
13,000	$13.9 \times 10^4$	$15.1 \times 10^4$
14,000	$14 \times 10^4$	$15.8 \times 10^4$
15,000	$14.05 \times 10^4$	$16 \times 10^4$



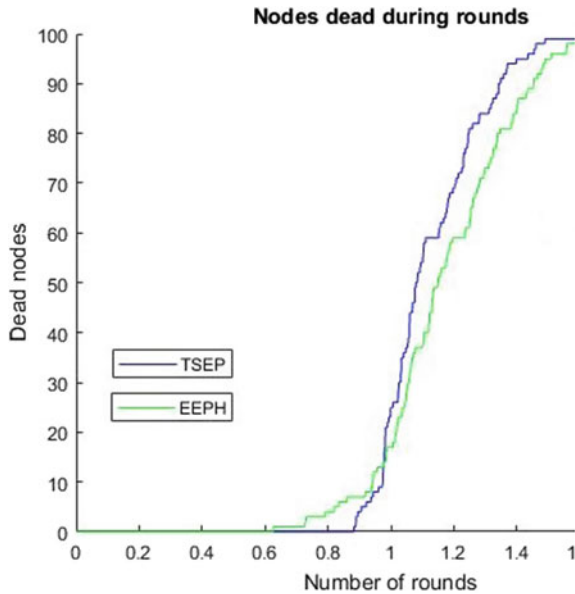
**Fig. 11** Number of alive nodes for 20,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



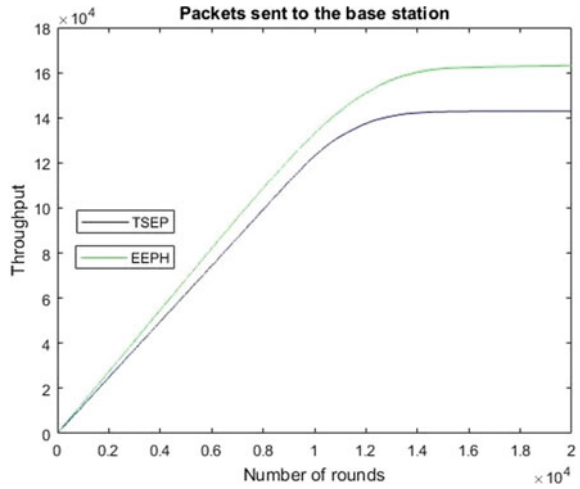
protocol and TSEP protocols in term of dead nodes for 20,000 rounds are shown in Fig. 12.

The comparison between EEPH protocol and TSEP protocols in term of number of packets send to BS for 20,000 rounds are shown in Fig. 13. The Fig. 14 shows

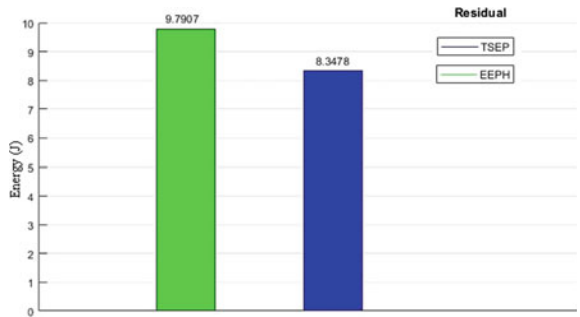
**Fig. 12** Number of dead nodes for 20,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



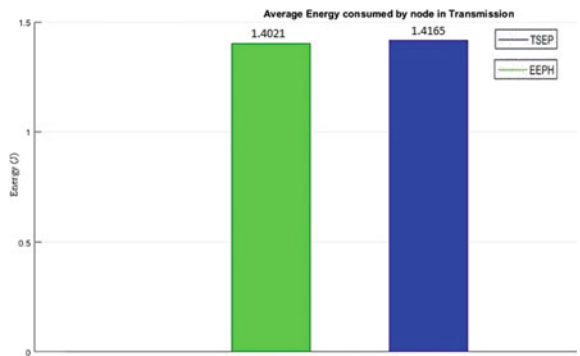
**Fig. 13** Packets send to base station for 20,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Fig. 14** Residual energy for 15,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



**Fig. 15** Average energy consumed for 15,000 rounds in  $(100 \times 100) \text{ m}^2$  Area



the comparison between EEPH and TSEP protocols in term of Residual Energy for 20,000 rounds.

**Table 8** Number of alive nodes for 20,000 rounds

Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 20,000		
Number of rounds	TSEP	EEPH
	Number of alive nodes	Number of alive nodes
8000	100	96
10,000	78	82
12,000	31	41
14,000	05	12
16,000	01	02
18,000	00	01
20,000	00	01

The Fig. 15 shows the comparison between EEPH and TSEP protocols in term of Average Energy Consumed for 20,000 rounds.

The alive nodes comparisons are done between EEPH protocol and TSEP protocol for 20,000 rounds shown in Table 8. At 8000 rounds the performance of TSEP protocol is better than EEHP but at higher rounds, i.e., more than 8000 round EEHP protocol performance is better.

The dead nodes comparisons are done between EEPH protocol and TSEP protocol for 20,000 rounds shown in Table 9 At 8000 round the performance of TSEP protocol is better than EEHP but at higher rounds, i.e., more than 8000 round EEHP protocol performance is better.

**Table 9** Number of dead nodes for 20,000 rounds

Area (100 × 100) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 20,000		
Number of rounds	TSEP	EEPH
	Number of dead nodes	Number of dead nodes
8000	00	04
10,000	22	18
12,000	69	59
14,000	95	88
16,000	99	18
18,000	100	99
20,000	100	99

**Table 10** Number of packets send to BS for 20,000 rounds

Area ( $100 \times 100$ ) m <sup>2</sup>		
Number of nodes: 100		
Total number of rounds: 20,000		
Number of rounds	TSEP	EEPH
	Number of packets send to BS	Number of packets send to BS
8000	$9.95 \times 10^4$	$10.8 \times 10^4$
0000	$12.1 \times 10^4$	$13.1 \times 10^4$
12,000	$13.9 \times 10^4$	$15.15 \times 10^4$
14,000	$14.15 \times 10^4$	$16.1 \times 10^4$
16,000	$14.15 \times 10^4$	$16.15 \times 10^4$
18,000	$14.15 \times 10^4$	$16.15 \times 10^4$
20,000	$14.15 \times 10^4$	$16.15 \times 10^4$

The comparison between TSEP protocol and EEPH protocol in term of packets send to BS are more by EEPH protocol shown in Table 10.

The simulation results of fixed area  $100 \times 100$  m<sup>2</sup> and 100 nodes on different rounds of transmission of packets in term of the energy saved in term of residual energy and average energy consumed after 10,000 rounds, 15,000 rounds and 20,000 rounds are shown in Tables 11 and 12.

**Table 11** Residual energy

Area ( $100 \times 100$ ) m <sup>2</sup>			
Number of nodes: 100			
Residual energy			
Protocol	10,000 rounds	15,000 rounds	20,000 rounds
TSEP	85.4921	67.0088	8.3478
EEPH	97.9289	87.0019	9.7907
(EEPH–TSEP)	12.4368	19.9931	1.4429

**Table 12** Average energy consumed

Area ( $100 \times 100$ ) m <sup>2</sup>			
Number of nodes: 100			
Average energy consumed			
Protocol	10,000 rounds	15,000 rounds	20,000 rounds
TSEP	0.64507	0.82990	1.4165
EEPH	0.52071	0.63000	1.4021
(TSEP–EEPH)	0.12436	0.1999	0.0144

The comparison between TSEP protocol and EEPH protocol in term of residual energy is shown in Table 11. The EEPH protocol have more residual energy saved as compare to TSEP protocol shown as difference of (EEPH–TSEP).

The comparison between TSEP protocol and EEPH protocol in term of average energy shown in Table 12. The EEPH protocol consumes less energy as compare to TSEP protocol as shown as the difference (TSEP–EEPH).

## 5 Conclusion

This research work proposed an energy-efficient protocol for heterogeneous network using two parameters as distance and energy. The parameters are defined as the ratio of residual energy of current node to the total energy at that level nodes. The distance is defined as the ratio of current node distance to the summation of distances of each node at that level from Base station. The proposed approach attains better energy-efficient performance to validate the superior performance of proposed energy-efficient protocol, conventional TSEP protocol is used to compare. From the experimental analysis, it is observed that proposed energy-efficient protocol attains better performance which minimizes the energy consumption and increases the network lifetime in heterogeneous wireless networks.

## References

1. Singh R, Gupta I, Daniel AK (2014) Position based energy-efficient clustering protocol under noisy environment for sensor networks using fuzzy logic technique. In: Science and information conference, London, UK, pp 843–848
2. Loscri V, Morabito G, Marano S (2005) A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH). In: 1999 IEEE vehicular technology conference, vol. 62. IEEE, New Jersey, No 3
3. Lindsey S, Raghavendra CS (2002) PEGASIS: power efficient gathering in sensor information systems. In: Aerospace conference proceedings, vol 3. IEEE, New Jersey
4. Narayan V, Daniel AK (2019) A novel protocol for detection and optimization of overlapping coverage in wireless sensor network. *Int J Eng Adv Technol* 8(S), ISSN :2249-8958
5. Li C, Ye M, Chen G, Wu J (2005) An energy-efficient unequal clustering mechanism for wireless sensor networks. In: IEEE international conference on mobile Adhoc and sensor systems conference
6. Manzoor B, Javaid N, Rehman O, Akbar M, Nadeem Q, Iqbal A, Ishfaq M (2013) Q-LEACH: a new routing protocol for WSN. In: International workshop on Body Area Sensor Networks (BASNet–2013) in conjunction with 4th international conference on ambient systems, networks and technologies (ANT 2013), Halifax, Nova Scotia, Canada, *Procedia computer science*, vol 19. pp 0926–931, ISSN 1877-0509
7. Hu FM, Kim YH, Kim KT, Park CW (2008) Energy-based selective cluster-head rotation in wireless sensor networks. In: International conference on advanced infocomm technology, Shenzhen, China

8. Wang Y, Baek KW, Kim KT, Youn HY, Lee HS (2008) Clustering with case-based reasoning for wireless sensor network. In: International conference on advanced infocomm technology, Shenzhen, China
9. Singh M, Gaurav, Soni S, Kumar V (2016) Clustering using fuzzy logic in wireless sensor network. In: International conference on computing for sustainable global development (INDIACom), New Delhi, India
10. Rai AK, Daniel AK (2020) An energy efficient routing protocol for wireless sensor network. *Test Eng Manag* 12556–12563, ISSN:0913-4120
11. Narayan V, Daniel AK, Rai AK (2020) Energy efficient two tier cluster based protocol for wireless sensor network. In: International Conference on Electrical and Electronics Engineering (ICE3), IEEE
12. Qing L, Zhu Q, Wang M (2006) Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Comput Commun* 29(12):2230–2237 (Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands)
13. Elbhiri B, Saadane R, El fldhi S, Aboutajdine D (2010) Developed Distributed Energy-Efficient Clustering (DDEEC) for heterogeneous wireless sensor networks. In *I/V communications and mobile network (ISVC)*, 5th international symposium, 30 Sept–2 Oct 2010
14. Manjeshwar A, Agrawal D (2002) A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. *IPDPS*, 15–19 April 2002
15. Yu Y, Govindan R (2001) Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor networks
16. Inagaki T, Ishihara S (2009) HGAF: a power saving scheme for wireless sensor network. *J Inf Process* 17:255–266
17. Chaturvedi P, Daniel AK (2015) An energy efficient node scheduling protocol for target coverage in wireless sensor networks. In: Fifth international conference on communication systems and network technologies, pp 138–142
18. Chaturvedi P, Daniel AK (2016) Trust based target coverage protocol for wireless sensor networks using fuzzy logic. In: International conference on distributed computing and internet technology, pp 188–192
19. Chaturvedi P, Daniel AK (2014) Recovery of holes problem in wireless sensor networks. In: International Conference on Information Communication and Embedded Systems (ICICES2014), pp 1–6
20. Chaturvedi P, Daniel AK (2017) A novel sleep/wake protocol for target coverage based on trust evaluation for a clustered wireless sensor network. *Int J Mobile Network Des Innov Inderscience Publishers (IEL)*
21. Narayan V, Daniel AK (2020) Multi-tier cluster based smart farming using wireless sensor network. In: 3rd International Conference on Big Data and Computational Intelligence (ICBDICI)
22. Lou W (2005) An efficient N-to-1 multipath routing protocol in wireless sensor networks. In: Proceedings of the 2nd IEEE international conference on Mobile Ad-hoc and Sensor System (MASS '05), Washington, DC, USA, pp 672–680
23. Smaragdakis G, Matta I, Bestavros A (2004) SEP: a stable election protocol for clustered heterogeneous wireless sensor networks. In: second international workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004)
24. Kashaf A, Javaid N, Khan ZA, Khan IA (2013) TSEP: threshold-sensitive stable election protocol for WSNs. In: International conference on frontiers of information technology

# Performance Analysis of Fuzzy-Based Relay Selection for Cooperative Wireless Sensor Network



Nitin Kumar Jain and Ajay Verma

**Abstract** The application of cooperative relaying is to limit the error in transmission and enhance the network throughput of nodes in the wireless sensor network. Intelligent relay selection scheme can be employed to improve the system performance further. This paper presents a performance comparison of proposed and existing relay selection schemes for cooperative wireless sensor networks. Conventional random relay selection scheme is compared with the proposed Fuzzy-based Relay Selection (FRS) scheme concerning frame error rate and network throughput. Simulation results depicted the effectiveness of the scheme providing better network throughput performance while limiting the error with improving the error rate performance in wireless sensor networks.

**Keywords** Cooperative communication · Wireless sensor networks · Fuzzy-based relay selection · Random relay selection · Signal-to-Noise Ratio (SNR)

## 1 Introduction

The advancement of micro-sensors development technologies in Wireless Sensor Network (WSN) has made it possible to implementation of smart, tiny and low-cost sensors. In addition, these sensor networks have opened many perspectives of applications and services in different areas such as industry, health, agriculture as well as home automation. Networks can have a number of nodes and some sensors

---

N. K. Jain (✉)

Department of Electronics and Communication IPS Academy, Institute of Engineering and Science, Indore, India

e-mail: [engg.nitin81@gmail.com](mailto:engg.nitin81@gmail.com)

A. Verma

Department of Electronics and Instrumentation, Institute of Engineering and Technology, DAVV, Indore, India

e-mail: [ajayrt@rediffmail.com](mailto:ajayrt@rediffmail.com)

to reach hundreds or even thousands. On the other hand, sensors can serve, where the human presence becomes very risky such as military zones or volcanic eruption zones [1]. Despite this great success of sensor networks, several problems remain open. For example, the first problem of sensors is their limited energy capacity. To lower manufacturing costs, several types of sensors have integrated and non-separable batteries. These sensors are no longer usable when their batteries are empty. The first challenge is to design layers protocol that saves energy and extends the life of the sensor [2]. In addition, a sensor has a processor and a low-capacity memory which imposes additional constraints on the developer. The sensors use a wireless channel to communicate, which most often is not reliable, and it is frequently disturbed by phenomena such as noise, reflections and distortions resulting in packet and energy losses. To combat the instability of the radio channel, some solutions have been proposed which uses multiple antennas to create spatial and temporal diversity. Though, the limitation of the size of the sensors prevents the installation of several antennas.

Cooperative communication [3–5] is an intermediate solution to this problem. It proposes to share the antennas of the neighbouring nodes for creating spatial diversity. Having information on the Channel State Information (CSI); the nodes lend themselves to each other antennas by transmitting the packets of their neighbours on channels of the best quality. This favourable strategy avoids the problems of unstable channels while increasing the link capacity.

Historically, cooperative techniques have emerged at the layer level, and many proposals have been made to this effect. On the other hand, the level of the upper layers, taking into account these cooperative techniques is still a subject on which many studies must be conducted.

## 2 Related Work

Cooperative communication overcomes the effect of multipath fading in a wireless network [3]. The proposed concept of relaying could be utilized for getting diversity on the node which receives the data, and which is achieved by the cost of slightly increment in complexity, more consumption of energy and more bandwidth utilization. To overcome the problem and enhance the spectral efficiency, space–time coding was proposed in [4]. Relaying technique could be minimizing the excessive utilization of bandwidth by proper selection of relaying and relay selection schemes [5]. In [6], authors proposed the different types of application of relay selection schemes. In [7] investigated the Channel State Information (CSI) based relay selection scheme, where the node with the best CSI is selected for relaying the data. In [8] presented a new relay selection strategy based on Fuzzy Inference System (FIS) which is targeted to improve packet delivery ratio and network lifetime, and authors further improved their work is presented in [9]. In [10] proposed a duty cycle-based cluster head selection scheme for WSN to enhance the lifetime of the WSN. Further improved similar work is presented by the same authors in [11]. In [12] investigated



a relay selection for wireless sensor network. In [13] proposed a combined study of relay selection and power allocation based on energy pricing. In [14] presented a new relay selection based on a hybrid neuro-fuzzy model to improve the network lifetime. A relay selection based on energy consumption is investigated in [15] and [16].

In this paper, the FRS scheme is proposed for improving the performance of the WSN by considering residual energy of nodes and path loss between the links as input parameters for the selection of the appropriate relay node.

### 3 System Model

For cooperative diversity, consider an area of  $100 \times 100$  square metres as depicted in Fig. 1, where  $d$  denotes sink (destination) node, and  $s$  denotes source node and a group of relay nodes ( $r_1, r_2, \dots, r_i$ ). For the simulation, at any particular time, relay node can perform a single task, either receive or send the information.

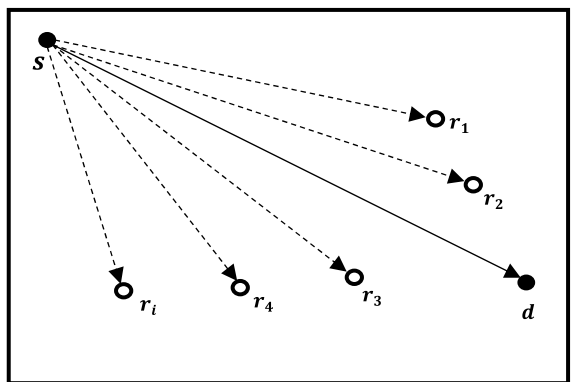
It is considered that every single node of the network consists of an alone RF antenna for transmitting and receiving the information. For the simulation, assumed that no one to one communication is possible between transmitting (source) and receiving (sink) node. The transmission of the data from the transmitting node mainly comprises two phases;

1. Broadcast phase
2. Relaying phase.

Let assume  $l$  is the length of data obtained from the BPSK modulation scheme. The data received at the sink node and relay nodes during the broadcast phase are given by:

$$Y_{s,d} = \sqrt{P_a} \cdot l \cdot h_{sd} + \eta_{sd} \tag{1}$$

**Fig. 1** Cooperative wireless sensor network scenario



$100 \times 100 m^2$

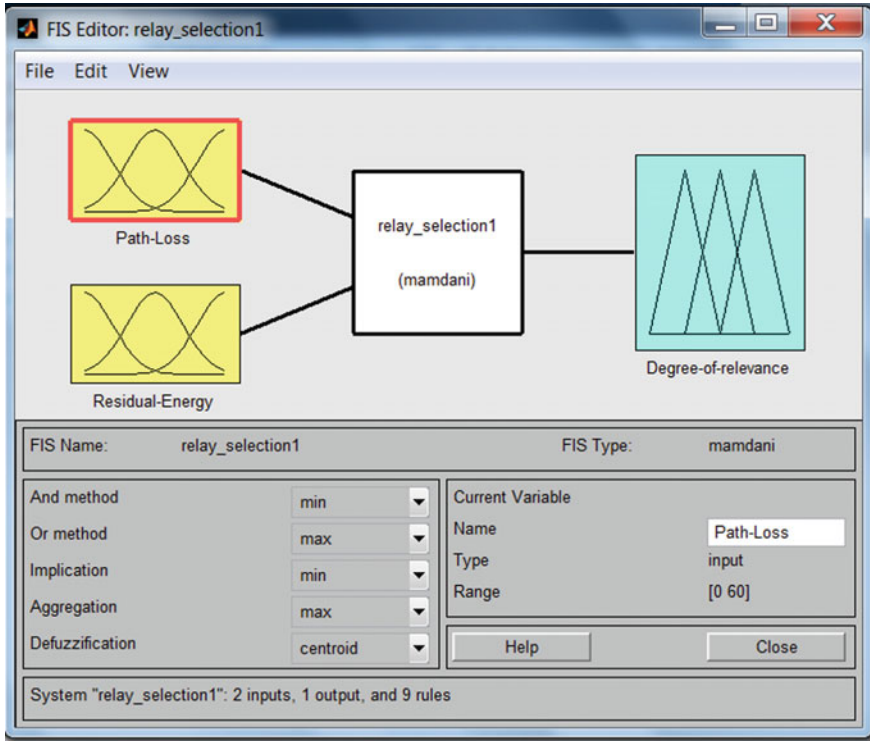


Fig. 2 FIS architecture for FRS scheme

$$Y_{s,r_i} = \sqrt{P_a} \cdot l \cdot h_{sr_i} + \eta_{sr_i} \quad \text{for } i = 1, 2, \dots, n \tag{2}$$

where  $h_{sd}$  and  $h_{sr_i}$  represent the channel coefficient between source–destination and source–relay node, while  $\eta_{sd}$  and  $\eta_{sr_i}$  represent additive white Gaussian noise with  $N_0$  variance.  $P_a$  denotes the power available during transmitting the data at the source node.

Once data received by sink node, then sink node check the data, whether data is corrupted by channel noise or not. If corrupted data is received, then the sink node executes the relaying phase. In this phase, the relay node is chosen for retransmitting the information to the sink node d is given by:

$$Y_{r_i,d} = \sqrt{P_r} \cdot l \cdot h_{r_i,d} + \eta_{r_i,d} \tag{3}$$

where  $h_{r_i,d}$  represents the channel coefficient between relay and destination node, while  $\eta_{r_i,d}$  represent AWGN.  $P_r$  denotes the power available during transmitting the data at the relay node.

The distance between relay and sink node can be computed by the formula given in Eq. (4).

$$\text{Dist}_{r_i,d} = \sqrt{(X_{r_i} - X_d)^2 + (Y_{r_i} - Y_d)^2} \tag{4}$$

And, by using Eq. (4), respective path loss can be computed using the formula given as:

$$\text{Path Loss} = 10 \cdot \alpha \cdot \log_{10}(\text{Dist}_{r_i,d}) + C. \tag{5}$$

where  $C$  denoted the constant which depends on the system losses.

$$\text{Energy}_{r_i} = \frac{E_0(r_i)}{e_{Tl-rl}}, \tag{6}$$

where  $E_0$  represents the initial energy given to the node before starting the simulation, and  $e_{Tl-rl}$  denotes the energy consumed by each node during transmitting and receiving the message.

### 3.1 Fuzzy-Based Relay Node Selection

#### i. Fuzzification

Fuzzification is the method of mapping of crisp input set to the fuzzy set by using Membership Functions (MFS). The Residual energy (Re) and Path loss (Pl) are considered as input parameters. The output of FIS is represented by Degree of Relevance (DoR). Figure 2 represents the input and output MFS-based FIS architecture of proposed FRS scheme. The input and output variables are defined in terms of linguistic values which are given as:

$$Pl = (\text{Weak(W)}, \text{Average(A)}, \text{Strong(S)})$$

$$Re = (\text{Low(L)}, \text{Medium(M)}, \text{Full(F)})$$

$$DoR = (\text{very bad}, \text{bad}, \text{intermediate}, \text{good}, \text{very good})$$

**Table 1** Fuzzy rules for proposed FRS scheme

S. No.	Pl	Re	DoR
1	W	L	Bad
2	W	M	Good
3	W	F	Very good
4	A	L	Bad
5	A	M	Intermediate
6	A	F	Good
7	S	L	Very bad
8	S	M	Bad
9	S	F	Intermediate

ii. Inference

As per states of MFS for the inputs (Pl and Re), nine rules have been created, which are worked as an inference of FIS. Table 1 represents a total of nine rules for FIS.

Relay with a maximum degree of relevance computes channel gain and forward the amplified version of information received in broadcast phase to the destination. In such a scheme, the relay node amplifies the data received from s with available transmission power  $P_b$  with gain factor  $\beta$  given as:

$$\beta = \frac{\sqrt{P_b}}{\sqrt{P_a |H_{sr_i}|^2 + N_0}} \tag{7}$$

The received signal from selected relay to the destination node can be presented as:

$$Y_{r_{\text{selectedtoDes}}} = \beta \cdot H_{r_{r_{\text{selectedtoDes}}}} \cdot Y_{sr_{\text{selectedtoDes}}} + \eta_{r_{\text{selectedtoDes}}} \tag{8}$$

where  $H_{r_{\text{selectedtoDes}}}$  represents channel gain between selected relay and destination node, and  $\eta_{r_{\text{selectedtoDes}}}$  represents additive noise with the same parameters as  $\eta_{sd}$  and  $\eta_{sr_i}$ .

Finally, replacing the values in Eq. (8), received signal at destination node will be:

$$Y_{r_{\text{selectedtoDes}}} = \frac{\sqrt{P_a P_b}}{\sqrt{P_a |h_{sr_i}|^2 + N_0}} h_{r_{\text{selectedtoDes}}} \cdot h_{sr_i} \cdot l + n'_{r_{\text{selectedtoDes}}}, \tag{9}$$

where  $n'_{r_{\text{selectedtoDes}}} = \beta h_{\text{selectedtoDes}} \eta_{sr_i} + \eta_{r_{\text{selectedtoDes}}}$  represents equivalent noise. It is assumed that  $\eta_{sr}$  and  $\eta_{r,d}$  are independent; therefore,  $n'_{r_{\text{selectedtoDes}}}$  is a zero-mean AWGN noise with variance given by:

$$\left( \frac{P_b |h_{r_{\text{selectedtoDes}}}|^2}{P_a |h_{sr_i}|^2 + N_0} + 1 \right) N_0$$

In our system, model assumed that relay selection is performed by the sink node so that all the information related to channels like channel coefficients are available at sink node only. The sink node uses Maximal Ratio Combining (MRC) techniques to combine the data received from the source and relay nodes. Therefore, the output of MRC at the sink node is given as:

$$Y = w_1 * Y_{sd} + w_2 * Y_{rd} \tag{10}$$

where parameters  $w_1$  and  $w_2$  known as link weights are calculated in such a way that signal strength of MRC output signal is maximum. The link weights may be calculated as  $w_1 = \frac{\sqrt{P_a h_{sd}^*}}{N_0}$  and  $w_2 = \frac{\sqrt{P_b h_{r_{\text{selectedtoDes}}}^*}}{N_0}$ . Assume that average energy of  $l$  is unity, the instantaneous SNR at the destination node using maximum ratio combining is given by:

$$\gamma = \frac{P_a |h_{sd}|^2 + P_b |h_{r_{\text{selectedtoDes}}}|^2}{N_0} \tag{11}$$

### 3.2 Bit Error Rate Analysis

The average BER probability for the system using binary phase-shift keying modulation given as [17]:

$$P_E = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \psi_\gamma \left( \frac{q}{\sin^2 \phi} \right) d\phi \tag{12}$$

where  $\phi_\gamma(s) = \left( 1 + \frac{s\bar{\gamma}}{m} \right)^{-m}$ , replacing this value in the above equation, it can be rewritten as:

$$P_E = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \psi_\gamma \left( 1 + \frac{q \bar{\gamma}}{m \sin^2 \phi} \right)^{-m} d\phi \tag{13}$$

For binary phase-shift keying modulation scheme  $q = 1$ .

To solve Eq. (13), let us consider,  $t = \cos^2(\phi)$ ,  $\cos(\phi) = t^{\frac{1}{2}}$  and  $\sin(\phi) = (1 - t)^{\frac{1}{2}}$  also.

$dt = -2\cos\phi\sin\phi d\theta$  or  $d\phi = -\frac{dt}{2\cos\phi\sin\phi}$  put these values in Eq. (13), to have

$$P_E = \frac{1}{2\pi} \int_0^1 \left( \frac{m\sin^2\phi + q\bar{\gamma}}{m\sin^2\phi} \right)^{-m} \frac{dt}{\cos\phi\sin\phi} \tag{14}$$

$$P_E = \frac{1}{2\pi} \int_0^1 \left( \frac{m(1-t) + q\bar{\gamma}}{m(1-t)} \right)^{-m} \frac{dt}{(t)^{\frac{1}{2}}(1-t)^{1/2}} \tag{15}$$

$$P_E = \frac{\psi_\gamma(q)}{2\pi} \int_0^1 (t)^{-\frac{1}{2}}(1-t)^{m-\frac{1}{2}} \left( 1 - \frac{t}{1 + \frac{q\bar{\gamma}}{m}} \right)^{-m} dt \tag{16}$$

where  $\psi_\gamma(q) = \left( 1 + \frac{q\bar{\gamma}}{m} \right)^{-m}$ , final expression for average BER probability can be given

$$P_E = \frac{\psi_\gamma(q)\Gamma(m + \frac{1}{2})}{2\sqrt{\pi}\Gamma(m + 1)} \left[ {}_2F_1 \left( m, \frac{1}{2} : m + 1; \frac{1}{1 + \frac{q\bar{\gamma}}{m}} \right) \right] \tag{17}$$

where,  ${}_2F_1(a, b; c; x)$  is known as Gaussian hypergeometric function.

**Frame error rate** characterizes the quality of the channel; it is the ratio of total corrupted frames that have been received to the total number of frames transmitted.

**Random relay selection** is the conventional method of relay selection in which randomly node being selected as a relay node. It does not consider any parameter to select or check the quality of the channel; it results in poor network throughput and error rate performance.

## 4 Results and Discussion

This research work is focused on the implementation of FRS scheme to improve the performance of cooperative WSN. For this purpose, a MATLAB-based simulation scenario has been presented which consists of a single antenna source, destination and a set of relaying nodes. The simulation environment is considered to be modelled with Nakagami fading environment. To evaluate the system’s performance, frame error rate and network throughput are considered as reference parameters.

Simulation results of our model are based on the performance comparison of proposed FRS and random relay selection strategies. The performance comparisons of these strategies are based on Frame Error Rate (FER) and network throughput. The results of the simulation are depicted in Figs. 3, 4, 5 and 6 for the 10, 20, 30 and

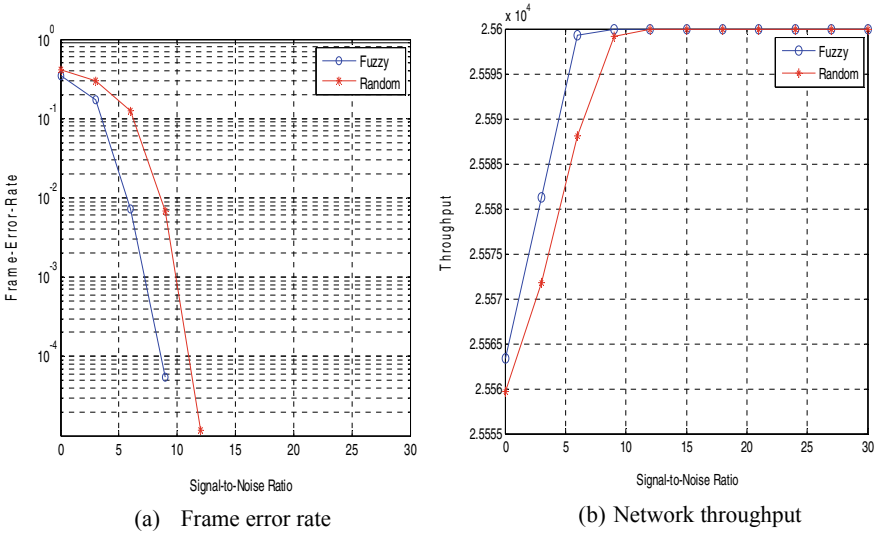


Fig. 3 Comparison of random and proposed FRS schemes for 10 relays

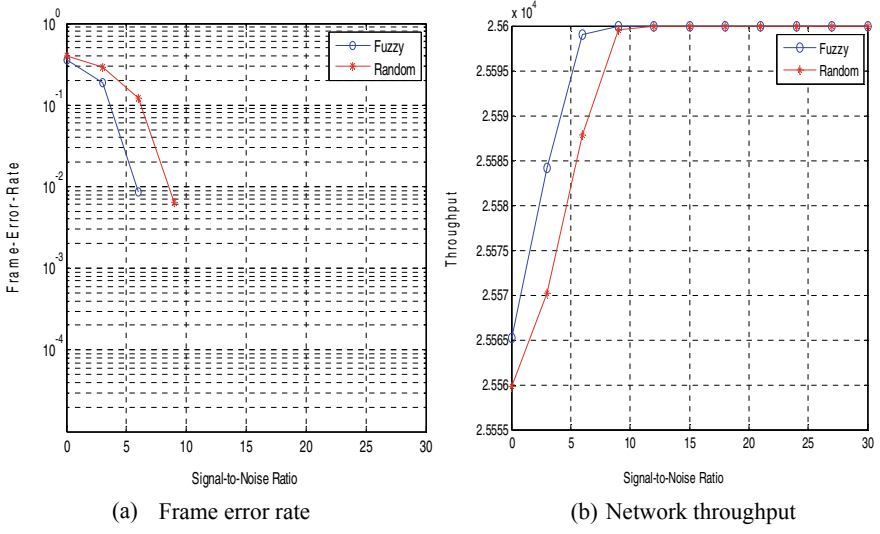


Fig. 4 Comparison of random and proposed FRS schemes for 20 relays

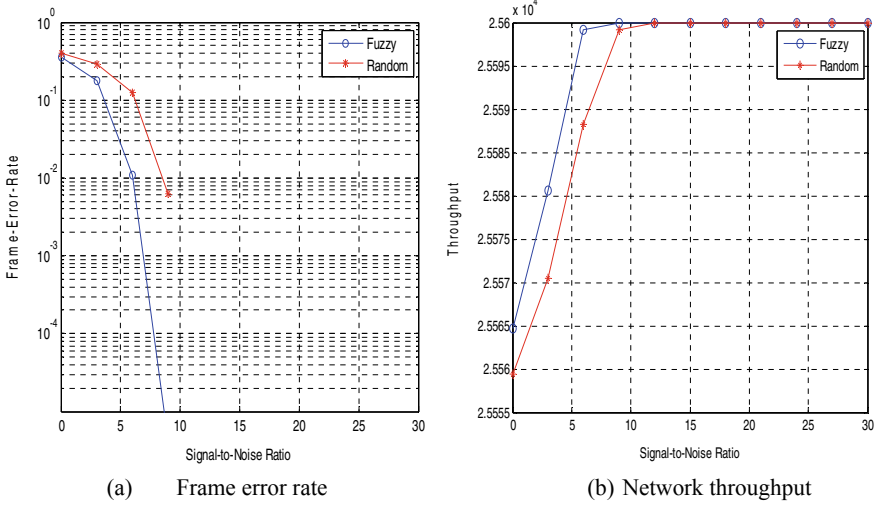


Fig. 5 Comparison of random and proposed FRS schemes for 30 relays

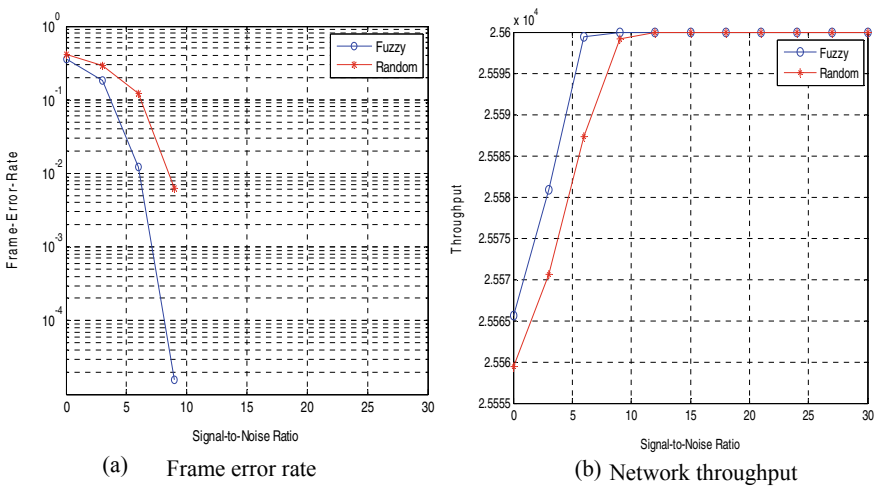


Fig. 6 Comparison for random and proposed FRS schemes for 40 relays

40 relay nodes deployed in WSN, respectively. On observing these results, it is clear that the proposed FRS scheme provides better bit error performance than random relay selection scheme. The network throughput is also better for the proposed FRS scheme.



## 5 Conclusion

Cooperative diversity techniques are the potential solution to spectrum inefficiency problems. Several advance communication techniques including Cognitive Radios (CRs), Wireless Sensor Networks (WSNs), Ad-Hoc Networks, VANETs use the application of cooperative relaying to enhance system performance. There are some deficiencies in cooperative relaying techniques too. The most important among which is the selection of optimal relay selection methodology. In this regard, the FRS technique for cooperative wireless sensor networks is proposed. The proposed scheme uses PI and Re to calculate a degree of relevance to select most suitable relay node, which cooperates with the source node. Simulation results based on error rate performance and network throughput are presented and compared with random relay selection scheme for Nakagami fading environment. On observing the simulation results, it can be concluded that the proposed scheme gives better network throughput and provides error-free transmission at low SNR values as compared to a random relay selection scheme.

## References

1. Akyildiz IF et al (2002) Wireless sensor networks: a survey. *Comput Netw* 38(4):393–422
2. Lewis FL (2004) Wireless sensor networks: smart environments: technologies protocols and applications, pp 11–46
3. Laneman NJ (2002) Cooperative diversity in wireless networks: algorithms and architectures. Dissertation Massachusetts Institute of Technology
4. Laneman NJ, Gregory W (2002) Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks. In: Global telecommunications conference (GLOBECOM'02). IEEE
5. Jain NK, Dongariya A, Verma A (2017) Comparative study of different types of relay selection scheme for cooperative wireless communication. In: IEEE international conference on information, communication, instrumentation and control, Indore
6. Laneman NJ, David T, Gregory W (2004) Cooperative diversity in wireless networks: efficient protocols and outage behaviour. *Inf Theor IEEE Trans* 50(12): 3062–3080
7. Bletsas A et al (2006) A simple cooperative diversity method based on network path selection. *Sel Areas Commun IEEE J* 24(3):659–672
8. Jain NK, Verma A (2019) Relay node selection in wireless sensor network using fuzzy inference system. *J Commun* 14(6):423–431
9. Jain NK, Yadav DS, Verma A (2019) A fuzzy decision scheme for relay selection in cooperative wireless sensor network. *Int J Commun Syst* 32:e4121
10. Shah IK, Maity T, Dohare YS (2020) Algorithm for energy consumption minimisation in wireless sensor network. *IET Commun* 14(8):1301–1310
11. Shah IK, Maity T, Dohare YS (2020) Weight based approach for optimal position of base station in wireless sensor network. In: International conference on inventive computation technologies (ICICT), Coimbatore, India, pp 734–738
12. Zarifi K, Abuthinien M, Ghayeb A, Affes S (2009) Relay selection schemes for uniformly distributed wireless sensor networks. In: Proceedings of IEEE WCNC, pp 1–6
13. Ke F, Feng S, Zhuang H (2010) Relay selection and power allocation for cooperative network based on energy pricing. *IEEE Commun Lett* 14(5)

14. Kaiser SM (2010) Neuro-Fuzzy (NF) based relay selection and resource allocation for cooperative networks. In: IEEE international conference on electrical engineering/electronics computer telecommunications and information technology
15. Yang W, Cai Y, Xu Y (2009) An energy-aware relay selection algorithm based on fuzzy comprehensive evaluation. In: International conference on networks security, wireless communications and trusted computing, IEEE
16. Jain N, Dongariya A, Verma A (2019) Energy efficient relay node selection in WSN using MPEGASIS protocol. In: Proceedings of recent advances in interdisciplinary trends in engineering & applications (RAITEA), Elsevier
17. Shin H, Hong LJ (2004) On the error probability of binary and M-ary signals in Nakagami-m fading channels. IEEE Trans Commun 52(4):536–539

# Rational Against Irrational Causes of Symptom Recognition Using Data Taxonomy



S. M. Meenaatchi and K. Rajeswari

**Abstract** The rational factor generally discharges the exact requirements from the mindset to highlight the questions as it is with the actual cause of nature. Irrespective, the irrational factors will flow from the mindset based on their individual scares and fearfulness of their subconscious thoughts of the mind. In general sequence, symptoms are the real facts to find out the cause for their solution to detect from the known values of attributes. Opinion mining is essential to find the data values and taxonomy to roll out the irrational factor on symptoms recognition. The research methodology is applied to find the new insight, which identifies the actual symptoms towards the disease belongings in a quick manner. The expected working hypothesis of outcomes may vary from the rational as dynamic issues against irrational, as unfounded factors, because the examination category depends on an expressive manner. The research work planned to expose and estimate roughly the rational and irrational factors of symptoms for the diseases. The paper presents features to detect the symptoms for the specific two diseases with positive or negative identification to correlate them in future survival.

**Keywords** Data analyses · Opinion mining · Symptoms recognition · Supervised learning factor

## 1 Introduction

Information retrieval is the main process to gather the relevant data reasonably. There may be two cases of approach which will be happening in a hospital environment like, defining all the complaints by the patients or all the parameters has to be detected by the doctor both by questioning and with some reports and lab testing. Defining

---

S. M. Meenaatchi (✉) · K. Rajeswari  
PG & Research Department of Computer Science, Tiruppur Kumaran College for Women,  
Tiruppur, Tamil Nadu, India  
e-mail: [meenaatchiniit@gmail.com](mailto:meenaatchiniit@gmail.com)

the causes may be the highlighted portion, which mainly supports to push the data into the next level of understanding as well as for the treatment also. If the patient rationally defines the symptoms, then it would be trouble-free to follow the hospital procedure quickly. Otherwise, the irrational factors will expose to time delay and made uncertainty to detect on highlighting symptoms parameter.

Symptoms detection is the very first step of analyzing and derives to predict the stages of the diseases easily and quickly with the importance of clinical audit's [1]. Many survivals on data mining are sustained with deliverables with intellectual and operational manner [2]. Unfortunately, the irrational cause factor is also defined along with the required illustration in general cases. Knowledge data discovery (KDD) is required to fine-tune the parameters reasonably to avoid the multiple subsequences to fulfill the end-user and to determine the feather consequences to prolong in the future. Many case studies and application features use a rough set approach based on the parameters like age and lower extremity meter score to detect the class value, to find the walking parameter with its possibility [3]. Data dependency and classification are defined based on object orientation and service-orientation based on the research proposal [4]. While reviewing the relevant research papers, many proposals are classified and cataloged towards the illustration and mainly hold with their attribute dependency with the functional approach. Impact smells of the research work are moved tremendously based on the coding factors either by service approach or in a descriptive language code.

Detection is the main factor in analyzing the data and processing with its level of usage. Making a study over that, fault detection is applied in many major applications like in bank transactions, electrical circuit designing, and even in human life, many false approaches is made towards the device. If the fault is detected, the usage of the appliances and applications made life easy and in a high-flying way. Fault detection is studied with the basic terminology like about the fault, failure, malfunction, and types of fault[6]. The fault detection approaches are also applied with their functionality like:

- Data methods and signal methods
- Process model-based methods
- Knowledge-based methods.

All the functional features define the development mainly towards validating the model with the parameters as studied. Diabetes-based symptoms are identified [4] and survived in statistical analysis, based on attributes like age, gender, race, BMI, annual household income, and household member size. Cardiovascular disease and diabetes are studied and reviewed by researchers based on their pressure and lipids and their research proposed with the scientific approach to analyze the parameters based on diseases like microvasculature, microvascular disease, inflammation, oxidative stress, heart failure, and so on. The research view is proposed to analyze and help the symptoms checkers as below:

- Emergency care
- Non-Emergency care reasonable

- Self-care reasonable.

One of the research work revolves around the symptoms checker, through web analysis, based on context-aware with disease details, significance on symptoms detection and provides the final predict disease as a result [3]. This paperwork describes a survey on data mining with its challenges, features, and algorithms techniques in detail.

## 2 Problem Statement Towards Research Objective

Many issues are analyzed towards the detection of the symptoms to profile the needs of human life. Research objectives revolve to sustain from the basic level of understanding to the risk factor of prediction through symptoms detection with rational and irrational factors for the human to analyze the disease. The main requirements to analyze the proposal are deep knowledge-gathering over them to stand on the prediction as an opinion mining in an ideal manner. The problem statement is designed and focused on the recognition of the symptoms with information gathered and used to prolong the implementation. The unique dependency towards the research objective of the problem statement is to identify the symptoms to predict the diseases in an easy and quick approach. The data mining process elevates with the steps to develop as new insights as problem definition, data collection, analyzing the data, plan the model, and build the model. (As per Fig. 1).

Figure 2 states the research process which highlights to underlying the development effort are:

1. Uniquely identify the issues. (i.e., not interrupt with other approaches.)
2. Plan the implementation strategy which involves the research work. (Study time, data collection, originate the problem, data analyses, and soon).
3. Design the model. (i.e., the application).
4. Execution to test the values for prediction.
5. Train and test the model to build new insights.

The main focal point of research objectives is contributed to analyses and study the features like:

1. Identify the problem definition towards disease identification
2. What are the dependent symptom factors for a rational and irrational cause?



**Fig. 1** Steps involved in the data mining process

**Fig. 2** Steps towards the research process



3. What is the cross relational covers between symptoms on heart issues and diabetes?

### 3 Implementation Methodology

Features are patterned after the deep study applied to the research process and with the attributes, the next stage to prolong is with the implementation of data taxonomy. The implementation featured along with the patient records to recognize the symptoms are detected to highlight the rational cause in actual ways. Formulating the research process with the implementation, the following steps are planned as follows:

1. Define the problem statement
2. Data collection and analyze the attribute
3. Feature extraction to conceal the formulation
4. Data visualization.

**Problem Statement:**

The first level of implementation is to define and analyze the problem statement and predict the symptoms with the high irrational cause parameter compared to the rational cause features. The planned hypothesis towards the data collection has to reach a minimum amount of the rational cause as 1 with the maximum amount of irrational cause 9 on symptoms code with the target range from 1 to 10.

**Data Collection:** Data collection is handled from the patient’s records based on the complaints registered from the hospital. The collection requires an initial stage to recognize the symptom either knowledgeable or vice versa.

Data collection is done from the inquiry and the answering format from the individual and executed.

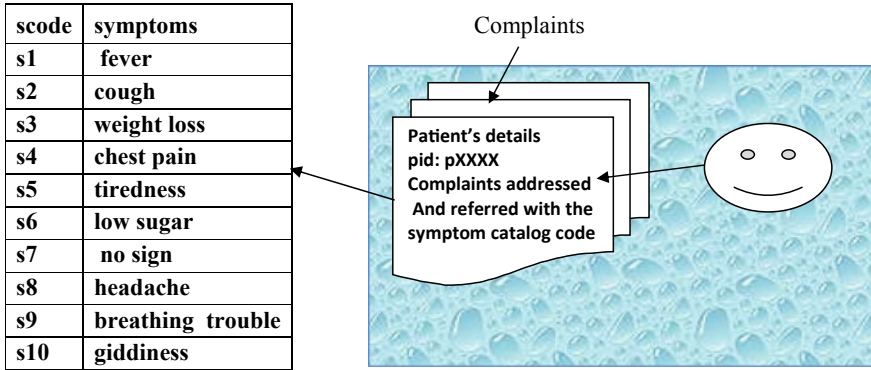


Fig. 3 Illustration towards patients monitoring process

**Feature Extraction:** Feature extraction is applied based on the dependency between the attributes as defined in the above Fig. 3. The extraction of attributes is very essential to correlate them with the limited parameters, which will support to define the problem statement appropriately.

**Data Visualization:** To display the results based on data sampling method, and applied by comparing rational against irrational symptoms recognized using positive or negative states for each and every individual to suggest the opinion poll easily.

Rational cause parameters are defined as an exact or accurate syndrome defined for the disease from the previous studies in the real world. It will be relevant to the diseases as well as relevant to other symptoms also.

The exact or actual symptoms of rational cause parameters are defined as H1 and D1, where H1 represents the heart diseases and D1 denotes diabetes. The causes which are related to heart disease (H1) are fever, chest pain, no sign, breathing trouble with score are s1, s4, s7, and s9. Similarly, the cause features of diabetes (D1) are weight loss, tiredness, low sugar, giddiness which is represented as score are s3, s5, s6, s10.

- For heart issues:
  - H1 = {fever, chest pain, no sign, breathing trouble}
  - i.e., H1 = {s1, s4, s7, s9}
- For diabetes issues:
  - D1 = {Weight loss, tiredness, low sugar, giddiness}
  - i.e., D1 = {s3, s5, s6, s10}.

Irrational cause parameters are irrelevant to the disease syndromes and do not coincide with the other symptoms. Unfortunately, irrational cause parameters are described based on the unawareness of the disease knowledge and symptoms. The irrational attribute is denoted as IR1 with key values like cough and headache, as below:

- For irrational cause factor for heart disease
  - IR1 = { cough, headache }
  - i.e., IR1 = {s2,s 8} (Figs. 4, 5 and 6).

Behavior features between the two rational and irrational patterns mainly depend on the understanding of the particular knowledge with their mindset parameters like stress, family history and obesity, and so on. Even though the disease is unique, it can be sustained to define the rational factors easily. The dataset attributes defined as patient code and symptom detection features with yes or no state values. Predictive opinion mining is required to catalog the issues when both the multiple factors are gathered based on the disease. The sampling analysis of the data of the proposed model is shown below:

- For diabetes as major and with heart issues.

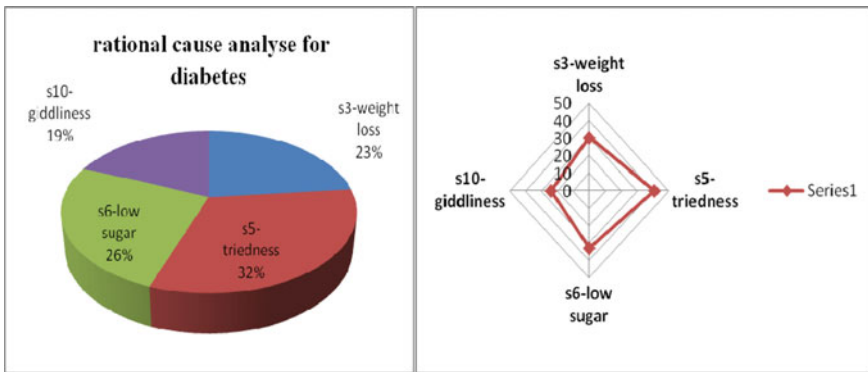


Fig. 4 Rational cause parameters for diabetes

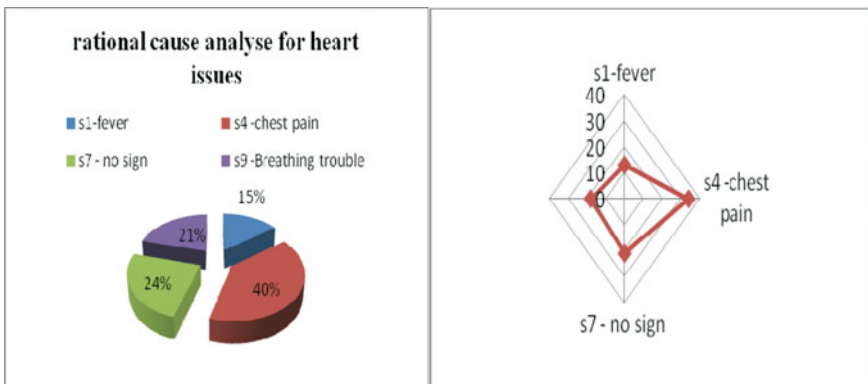
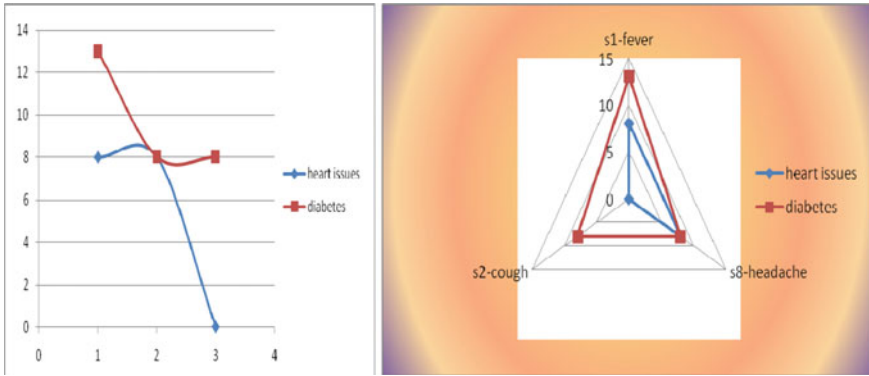


Fig. 5 Rational cause parameters for heart issues

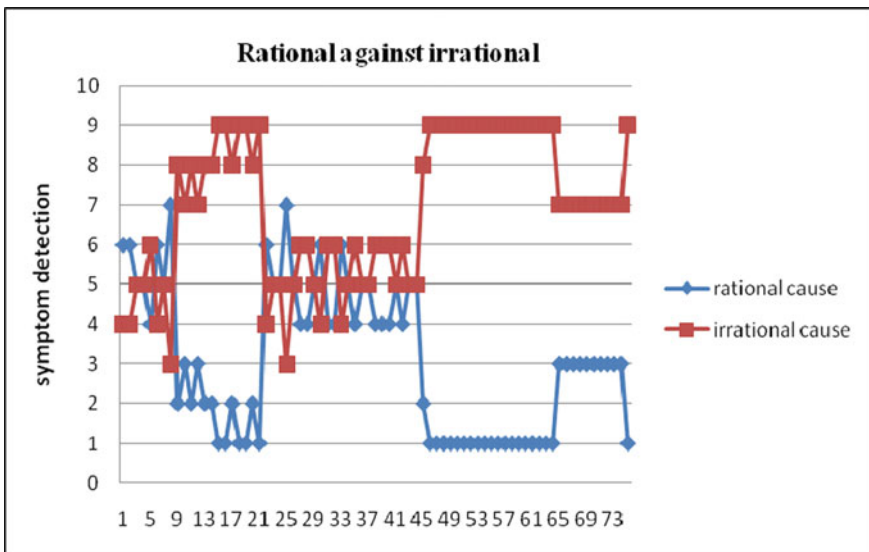




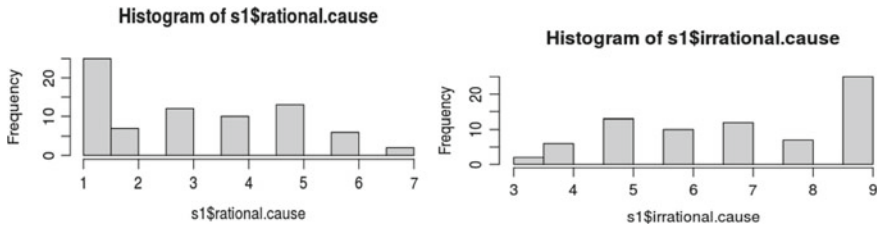
**Fig. 6** Irrational cause for heart diseases and diabetes with its symptoms codes

- $D1 = \{ \text{Weight loss, chest pain, tiredness, low sugar, no sign, giddiness} \}$ .
- i.e.,  $D1 = \{ s3, s4, s5, s6, s7, s10 \}$ .
- For heart issues as major and with diabetes.
  - $H1 = \{ \text{fever, chest pain, no sign, low sugar, breathing trouble} \}$ .
  - i.e.,  $H1 = \{ s1, s4, s6, s7, s9 \}$  (Fig. 7).

To define the resultant factor, the records are traced with more than 500 values, and by using a random sampling feature, some of the records are picked up and



**Fig. 7** Symptoms detection is applied to analyze the rational against the irrational cause for the diseases



**Fig. 8** Histogram representation of rational and irrational cause

traced the complaints to detect the rational or irrational parameters. Absolutely, among the results, the rational values are very low compared in some sequences only; i.e., symptoms are defined correctly. Whenever the irrational values are high, the symptoms recognized and registered as complaints will also be high, with unknown features and it will be included to the state the disease (Fig. 8).

## 4 Conclusion

Unique identification of the features is required to highlight the cause of symptoms in an accurate way, but exactly the main focal point depends upon the disease with their stages. The rational cause can be easily shortlisted toward the unreasonable source belongings, to make some opinion tuning easily. Sufficiently, the knowledge transplanted with the elementary indication and final execution from the support of deep crawling on the application of subjective analyses. In the foremost development of the research work, many attributes are planned to survive train the model to provide the resultant causes dominantly.

## References

1. UmaPermal, The importance of clinical audit in India. *Curr Med Res Pract*. <https://doi.org/10.1016/j.cmrp.2020.05.010>
2. Joseph (2019) Survey of data mining algorithm's for intelligent computing system. *J Trends Comput Sci Smart Technol (TCSST)* 1(01):14–24
3. Shi Y, Japa A (2017) Multi-dimensional analysis on data sets for information retrieval. In: *International conference on intelligent sustainable systems (ICISS 2017)*, 7–8 Dec, 2017
4. Kalaivani R (2017) A study of rough sets theory and it's application over various fields. *J Appl Sci Eng Methodologies* 3(2):447–455
5. Sabir F (2018) A systematic literature review on the detection of smells and their evolution in object-oriented and service-oriented systems. *Softw Pract Expert*
6. Zagreb, Fault detection methods: a literature survey
7. Clark NG (2007) Symptoms of diabetes and their association with the risk and presence of diabetes. *Diabetes Care* 30(11)

8. Walter FM, Rubin G, Symptoms and other factors associated with time to diagnosis and stage of lung cancer: a prospective cohort study. *Br J Cancer*. <https://doi.org/10.1038/bjc.2015.30>
9. Kao H-C, Tang K-F, Chang EY (2008) Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. In: The thirty-second AAAI conference on artificial intelligence (AAAI-18)
10. Walter FM, Rubin G (2015) Symptoms and other factors associated with time to diagnosis and stage of lung cancer: a prospective cohort study. *Br J Cancer* 112(Suppl 1)
11. Microsoft SQL Server BI Toolkit. <https://www.microsoft.com/sqlserver/en/us/solutions-technologies/SQL-Server-2012-business-intelligence.aspx>
12. [https://en.wikipedia.org/wiki/Mental\\_disorder](https://en.wikipedia.org/wiki/Mental_disorder)
13. <https://towardsdatascience.com/data-science/home>
14. <https://www.engpaper.com/data-mining-2019.html>
15. <https://www.w3.org/2012/06/pmod/opinionmining.pdf>
16. <https://www.sciencedirect.com/topics/computer-science/data-mining-research>
17. <https://www.journals.elsevier.com/current-medicine-research-and-practice/recent-articles>

# Multimodal Emotion Analytics for E-Learning



J. Sirisha Devi and P. Vijaya Bhaskar Reddy

**Abstract** E-learning is one of the modes of learning which is gaining popularity nowadays. To analyze the effectiveness of the E-class, there are many ways like feedback at the end of the lecture. But analyzing at the end will not serve the purpose. By analyzing the feedback of the listeners during the class improves the quality of the E-class. This can be achieved by analyzing the face and acoustics of the listeners. In a few cases, the data may not be obtained from both modalities. So, a novel method is proposed which will continue to perform and successfully recognize emotions even when one modality is absent. Product trends can be predicted by using this model for market analysis and ads can be improvised to reach their end target. Thereby, the manufacturers will receive relevant and better product feedback which leads to a rise in product sales. Studying how users view these multimodal posts will be of interest to sociologists and psychologists. The suggested project will provide a platform for publishing research papers in reputed journals, provides an opportunity for manpower training.

**Keywords** E-learning · DBN · LSTM · Swarm intelligence

## 1 Introduction

In E-learning, webcams and microphones are used as input devices. They are used to capture the listener's facial expressions and voice for continuous monitoring of the listener's emotions. This offers a remarkable improvement in the quality and effectiveness of E-learning [1, 2].

---

J. Sirisha Devi

Department of Computer Science and Engineering, Institute of Aeronautical Engineering, JNTU(H), Hyderabad, Telangana, India

P. Vijaya Bhaskar Reddy (✉)

School of C&IT, REVA University, Bengaluru, India

e-mail: [bhaskarreddy.pv@reva.edu.in](mailto:bhaskarreddy.pv@reva.edu.in)

Across time and nations, the level of commonness of serious mental issues does not contrast excessively. In any case, the rate predominance of basic mental issues will change contingent upon the mental climate. The scope of treatment for mental disorders in most places is very restricted [3]. The focus is too much on medication. The proposed project gives scope to everyone to monitor their mental health condition to avoid severe consequences like committing suicide [4].

Emotions are what drive the audience to purchase. Most successful initiatives and marketing campaigns focus on the emotions of the end user [5]. Initiatives like Swachh Bharat, Swasth Bharat by the Government of India were able to reach citizens of our country as they were able to relate themselves to the schemes emotionally. Ads like Dove's "Real Beauty," Nike's "Just Do It," attempted to move the viewer's heart, and then rely on the experience to build a lasting bond with the brand. Predicting, influencing, and regulating human behavior are the tools of the legal system. Emotions play a vital in decisions, institutions, and legal rules.

Emotions play a vital role in the learning process and handling a person. Therefore, it is expected that emotions influence how individuals learn and that a few emotions can help or frustrate the learning procedure. In the conventional learning setting the educator fills in as a facilitator between the understudy and his adapting course, he effectively sees the understudied perspective and changes the encouraging procedure to the understudy's needs and conduct. In internet learning conditions, this constant part does not exist. A conceivable answer for this issue can be the expansion of instruments that empower PCs to recognize and meddle when the understudied requires help or inspiration to continue with progress. Consequently, the proposition of another design that connected to learning stages will attempt to close to the hole between the understudies and web-based learning stages regarding emotional impact. There comes the next problem to be solved, feature selection. The concept of multimodality increases the dimensional of the feature set. The process of selecting relevant features that improve classifier's accuracy is known as feature selection. In real-world scenarios, to train a classifier, a few tens to hundreds of features are considered. So, undoubtedly, there is a need for a technique that removes the unwanted and irrelevant features and thereby, gives an optimal feature set to the classifier. The focus of the proposed work is on designing a feature selection technique that inherits the properties of swarm intelligence for optimization of the feature set. Finally, the multimodal emotion recognition system will be able to detect the emotion of a subject to improve the efficiency of E-learning. The proposed architecture will evaluate the emotional state of the student in an online learning environment. Based on the result, immediate remedial action will be taken by the tutor to enhance the learning results.

## 2 Origin of the Proposal

In literature, it is shown that there is a relationship between the learning process and human emotions. Despite the fact, this association is far from being basic and direct.

It is acknowledged that positive and negative emotional states can cause various types of reasoning and can have an impact on the learning viewpoint.

The primary question was:

**“In an interactive E-learning environment, is it possible to recognize and measure a student’s emotion?”**

In general, feedback, i.e., the emotion of a student about a lecture will be collected either in written or oral at the end of the session which may not give a possibility to rectify the pitfalls during that particular session. The quality of a lecture can be refined only when the tutor gets knowledge about the student’s emotions on the spot. There have been numerous examinations utilizing different single modalities like facial expressions.

Another question was:

**“Will the learning process of a student increase with the analysis of his emotions?”**

In noting emphatically to the central question, this inquiry is halfway replied. As results appeared, the understudies’ learning results can be enhanced by adding a passionate segment to a learning stage. Usually, speech data and facial data are the two most prominent modalities that are used for emotion recognition. A few cases of multimodal emotion databases can be found in (Zhang et al. 2016). It is widely accepted by the psychological theory that the emotions of humans can be classified into six general categories happiness, anger, sadness, fear, disgust, and surprise (Sumpeno et al. 2011). When the multimodal data is considered, the dataset will become huge and much likely to have redundant and irrelevant features in it. So, it is required to remove the irrelevant features which do not contribute to raising the accuracy of the classification process [6]. The focus of the proposed work is on designing a feature selection technique that inherits the properties of swarm intelligence for optimization of the feature set. Finally, multimodal emotion recognition system will be able to detect the emotion of a subject to improve the efficiency of E-learning [7].

A very important aspect of multimodal learning that is missed in these approaches is the ability to automatically learn the features of various modalities, to effectively integrate multiple modalities, and to inherently associate different modalities [8]. All these can be easily achieved by utilizing deep learning methods, as they are capable of extracting task-specific features from the data and learning the relationship between modalities through a hard representation. This shared representation of the data which reveals the association between different modalities makes the trained structure a generative model [9, 10]. That is, the model would be able to make the best use of the complementary information of different modalities, and handle missing modalities as long as the relationship between the absent modality is efficiently learned by the model. The importance of dealing with missing modalities is two-folds [11, 12]. First, a modality is available during the training phase but might be missing or corrupted during the online recognition phase. Second, a certain modality may not be useful for real applications but it is advantageous to train a more robust model with this modality together with others that would be always available. In this proposed work, deep learning models are used on multimodal data to learn the association between modalities.

### 3 Proposed Methodology

#### 3.1 Methodology

Block diagram of the proposed model is shown in Fig. 1.

##### Phase I: Data Preparation

In this segment, features are extracted from different modalities and they are configured.

##### Images:

2D static images or video sequences are the two different approaches to gather facial expressions. In this research, the 2D method is adopted for emotion recognition from facial expressions. There are mainly five steps for emotion recognition from facial expressions: detection, tracking, feature extraction, feature selection, and classification as shown in Fig. 2.

The extraction of facial features assumes a critical part in feeling acknowledgment from outward appearance. Among the facial features, eyes, nose, and lip are the most critical features. Different methodologies have been done to extricate facial focuses, (for example, eyes, and nose) from pictures and video succession of appearances.

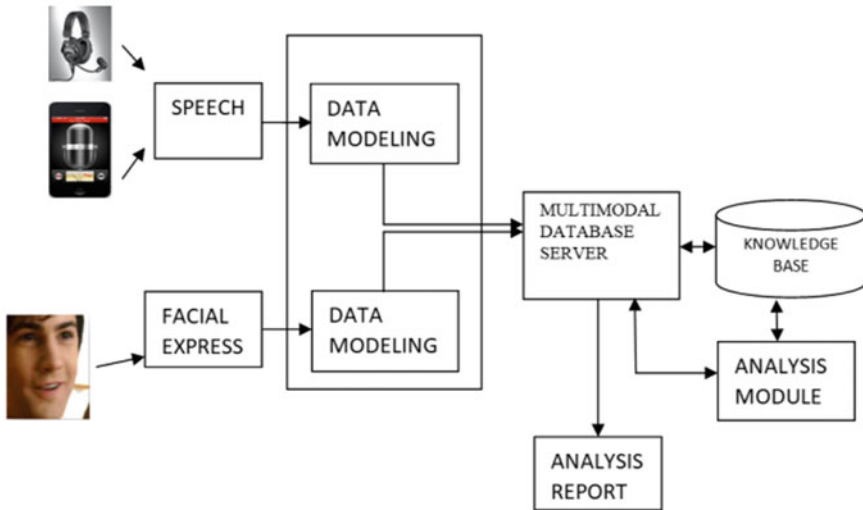


Fig. 1 Block diagram of multifaceted emotion recognition



Fig. 2 Facial emotion recognition

From each image, 2048 features are extracted and placed in a vector. These features vectors are fed as input to long short term memory (LSTM) networks in a sequence and after an entire sequence, an analysis report of the results is prepared. For a given sequence of n frames  $X_i \in \{X_1, \dots, X_n\}$ , the last  $X_n$  frames are used for target prediction as shown in Fig. 3.

The length of the sequence of an image is predefined; if any sequence is shorter than the predefined value, then the blanks in the image are filled by black frames at the beginning. These black frames do not contribute any valuable features to the system, so the bio-inspired optimization algorithm will extract features from the rest of the frame.

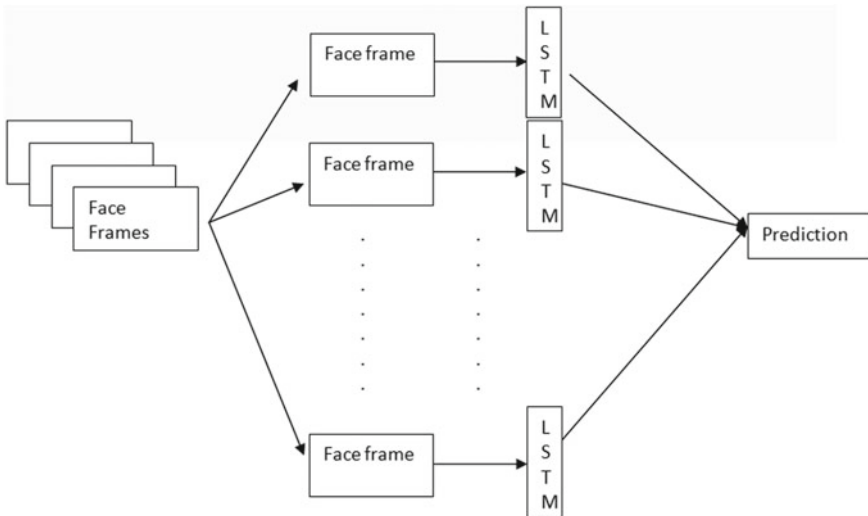
The optimization algorithm that will use for the optimization is as follows.

The parameters of employee ants are fitness and probability. The maximum number of iterations will be initialized first. Based on the number of features, construct a fully connected graph. The sum of employee ants and onlooker ants gives us the total numbers of ants that run for optimization.

The first iteration begins with an initial pheromone value. K-fold cross-validation method is used to calculate the predictive accuracies at each node. The objective function used is given by Eq. 1.

$$f_i = \left( 1 - \frac{\sum_{i=1}^n S_i}{n} \right) \tag{1}$$

Employee ants placed on each node will travel towards their nearest neighbor using random proportional rule, Eq. 2.



**Fig. 3** Feature extraction using DBN and prediction using LSTM



$$p_{ij}^k(t) = \frac{[T_{ij}]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [T_{il}]^\alpha [\eta_{il}]^\beta}; \quad \text{if } j \in N_i^k \tag{2}$$

By the end of each iteration, the predictive accuracy of each feature subset is evaluated using k-fold cross-validation.

Local pheromone is updated, fitness, and probability of subset of features are evaluated using Eqs. 3, 4, 5, respectively.

$$T(i, j) = (1 - \rho)T(i, j) + \rho T(0) \tag{3}$$

$$\begin{aligned} \text{Fit}_i &= \frac{1}{1 + f_i}; \quad \text{if } f_i \geq 0 \\ &= 1 + \text{abs}(f_i); \quad \text{if } f_i < 0 \end{aligned} \tag{4}$$

$$P_f = \frac{\text{fit}_i}{\sum_{i=1}^m \text{fit}_i} \tag{5}$$

Generate a new set of employees and onlooker ants for the next iteration.

These optimized feature vectors are fed as a single input to deep belief networks (DBN). ROI net is used to convert each image frame into a feature vector of length 2048, thereby the input dimension to DBN from ‘M’ frames is  $M * 2048$ , where ‘M’ is the number of images.

**Voice:**

Voice signals captured are preprocessed, and thus, obtained filtered signal is divided into frames. Hamming window of size 20 ms with an offset of 10 ms is used. A feature set of 39 mel frequency cepstral coefficients (MFCCs) are considered from each frame, as shown in Fig. 4.

These frame vectors are fed as input to LSTM. Frames belonging to a speech signal are concatenated into a single vector and fed as input to DBN. In case the speech signal is sparse, the acoustic holes is filled with silence. As the silence part cannot contribute any features, the proposed bio-inspired swarm intelligence optimization algorithm will extract the optimal feature set from the rest of the frame.

**Phase II: Feature Selection**

Feature subset selection serves as a pre-processing component for the classification process. The classification process then uses these selected relevant features to classify a given data set and compute results. The higher the relevance of selected features the higher will be the accuracy of the classifier. A straight approach of



**Fig. 4** Speech signal processing

feature subset selection is that every possible feature subset is tested and finalizing a subset that is most suitable for a given domain problem. This approach involves an algorithm that leads to cost extensive search of space which is certainly not optimal from the performance point of view.

Till now in research for finding optimal feature set was either classifier or data set dependent. These selection techniques either provide redundant or irrelevant features within the space limits which reduce the accuracy of the classifiers. Still, an open research problem is the identification of optimal feature set selection techniques for a large data set.

Many techniques including non-heuristic and heuristic have been applied to solve feature subset selection problems. These algorithms range from univariate to multi-variate where univariate techniques consider one feature at a time and evaluate the features based on some defined statistics [13]. On the other hand, the multi-variate techniques consider the overall subset evaluation to give classifier accuracy. Due to this reason, the multi-variate problem gives a better solution even when features are dependent on the data set of the problem domain. The multi-variate techniques are further divided into a filter, wrapper, and hybrid approaches. Filter techniques do the feature subset selection without the interruption of classifiers while wrapper approaches include the classifier as well. This makes the wrapper approaches a better candidate for providing a solution.

Many bio-inspired algorithms like genetic algorithm, particle swarm optimization, ant colony optimization, artificial bee optimization, crow optimization, cockroach optimization, shuffled frog leaping algorithm, etc., have been applied to give a solution of feature selection technique using certain benchmark data sets. The genetic algorithm is a biologically inspired natural algorithm and can get high classification accuracy of support vector machine (SVM) classifier. However, the algorithm is itself a computationally very expensive algorithm leading to less optimized solutions. On the other hand, the PSO also provides the FS solution strategies. Also, the bat algorithm has been applied on multiple techniques thereby giving better results than GA and PSO algorithms [14, 15].

Thus, a bio-inspired optimization algorithm is been proposed to evaluate the problem and comparison will be made with the previous techniques that are related to human emotions from multifaceted data. It increases the efficiency of the system by identifying the emotion of the end-user before decision making.

By analyzing the objectives of the proposed system, false acceptance and false rejection are the two measures used for estimating the accuracy of emotion recognition. These two errors are traded-off against each other. The equal error rate (EER) is the working point where the two sorts of blunders happen similarly. This is demonstrated by the crossing point of the working bend with the straight line. Parameters in which commitments for the assessment of the anticipated calculation are recorded underneath in Table 1.

### **Phase III: Build the model for emotion recognition in intelligent systems**

After feature selection, the optimal set of features is used for emotion recognition with DBN and LSTM as classifiers. For experimentation, the seven basic emotions will be considered (angry, happy, deprived, disgust, sad, neutral, joy). From the

**Table 1** Parameters for bio inspired optimization algorithm implementation

$\alpha$	$P$	$\delta$	$\beta$	$C_{miss}$	$P_{target}$	$C_{FA}$
1	0.2	0.8	0.1	10	0.01	1

dataset chosen for experimentation, 60% of the samples from each modality are randomly chosen for training, and for testing, the remaining 40% is used. A sequence of 5 s from each modality is fed to the proposed model, and the emotion is identified by a classifier. As the objective is to classify emotions but not subjects, from the same dataset, the process of randomly dividing the data for training and testing will be repeated for ‘ $n$ ’ number of times and an average of these results is taken which improves the accuracy of the system. Tests will be performed on all modalities per segment to obtain the accuracies of DBN and LSTM. Consistency of the emotion is verified by considering every segment of emotion against other segments, including it.

## 4 Experimentation and Results

Relevant literature on E-learning and emotion recognition will be collected from the internet, visiting R & D, labs and discussing with experts in this area. Required hardware and software equipment to conduct the research will be purchased. Audio and facial data are collected from various people by giving assurance to them that their data will be secured [9]. Data will be stored in the database.

Features are extracted from speech and face gestures. Then, using the proposed algorithm, optimal feature set is selected. Thus, gathered information from speech and facial expressions are sent to the analysis module where deep neural networks will get trained for analysis. A deep neural network is used to analyze the emotional condition of a subject. The results are used to evaluate the total system followed by testing of the final system and field test. The proposed work plan is depicted in the following chart (Fig. 5).

The developed model will continue to perform and successfully recognize emotions even when one modality is absent. Product trends can be predicted by using this model for market analysis and ads can be improvised to reach their end target. Thereby, the manufacturers will receive relevant and better product feedback which leads to a rise in product sales. Studying how users view these multimodal posts will be of interest to sociologists and psychologists.

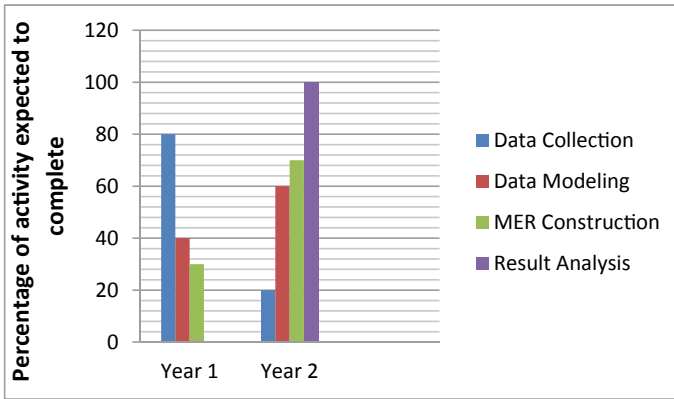


Fig. 5 Proposed work plan

## 5 Conclusion

About the utilization of the research outcome of this project, it will be recommended to various agencies working on autoresponders and chatbots. The developed algorithm can be utilized to derive an optimal feature set for acquiring the emotions from persons with mental illness and can be helpful for the doctors to assess the state of the patient. Secure cloud framework protecting patient privacy and the confidentiality of their data. Finally, there is a scope to get patents on this tool. The suggested project will provide a platform for publishing research papers in reputed journals, provides the opportunity for manpower training.

## References

1. Alepis E, Stathopoulou IO, Virvou M, Tsihrintzis GA, Kabassi K (2010) Audio-lingual and visual-facial emotion recognition: Towards a bi-modal interaction system. In: Proceedings-international conference on tools with artificial intelligence, ICTAI, vol 2, pp 274–281. 10.1109/ICTAI.2010.111
2. Cheong C, Cheong F, Filippou J (2013) Using design science research to incorporate gamification into learning activities. In: PACIS 2013 Proceedings, pp 1–14
3. Faria R, Almeida A, Martins C, Goncalves R (2015) Emotions in online learning. In: Advanced science and technology letters, pp 16–21
4. Vogt T (2010) Real-time automatic emotion recognition from speech. Doctoral dissertation. Retrieved from <http://pub.unibielefeld.de/luur/download?func=downloadFile&recordId=2301483&fileId=2301486>
5. Schuller B, Batliner A, Steidl S, Seppi D (2011) Recognizing realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge. *Speech Commun* 53(9–10):1062–1087
6. Sirisha Devi J, Yarramalle S, Nandyala SP, Vijaya Bhaskar Reddy P (2017) Optimization of feature subset using HABC for automatic speaker verification. In: At 2017 IEEE international conference on electrical, computer and communication technologies

7. Baler JT et al (2016) Toward expert systems in mental health assessment: a computational approach to the face and voice in dyadic patient-doctor interactions. *JMIR Publications* 2(1)
8. A Review of Emotion-Aware Systems for E-learning in Virtual Environments. Available from: [https://www.researchgate.net/publication/304108417\\_A\\_Review\\_of\\_Emotion-Aware\\_Systems\\_for\\_E-learning\\_in\\_Virtual\\_Environments](https://www.researchgate.net/publication/304108417_A_Review_of_Emotion-Aware_Systems_for_E-learning_in_Virtual_Environments). Accessed 3 Aug 2018
9. Deshmukh SP (2018) Feedback based real time facial and head gesture recognition for E-learning system. In: *CoDS-COMAD '18 proceedings of the ACM India joint international conference on data science and management of data*, pp 360–363
10. Saini AK, Khatri P, Raina K (2018) Towards understanding preference of use of emoticons for effective online communication and promotion: a study of national capital region of Delhi, India. In: Saini A, Nayak A, Vyas R (eds) *ICT based innovations. Advances in intelligent systems and computing*, vol 653. Springer, Singapore
11. Srivastava P, Singh MM (2017) Job satisfaction among healthcare professional in public and private healthcare setup in India. *Indian J Res*
12. Juutinen S (2011) Emotional obstacles of e-learning. *Computing*. University of Jyväskylä
13. Eslaminejad A, Safa M, Ghassem Boroujerdi F, Hajizadeh F, Pashm Foroush M (2017) Relationship between sleep quality and mental health according to demographics of 850 patients with chronic obstructive pulmonary disease. *J Health Psychol SAGE J*
14. Patel V, Weobong B, Weiss HA, Anand A, Bhat B, Katti B, Dimidjian S, Araya R, Hollon SD, King M, Vijayakumar L, Park A-L, McDaid D, Wilson T, Velleman R, Kirkwood BR, Fairburn CG (2017) The healthy activity program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomized controlled trial. *Lancet* 389(10065):176–185
15. Sirisha Devi J (2016) Speaker verification with optimized feature subset using MOBA. In: *2016 IEEE distributed computing, VLSI, electrical circuits and robotics (DISCOVER)*

# Intrusion Detection: Spider Content Analysis to Identify Image-Based Bogus URL Navigation



S. Ponmaniraj, Tapas Kumar, and Amit Kumar Goel

**Abstract** Every user wants to access their expected contents from legitimate web pages on the Internet by entering key phrases into search engines. Fortunately, crawler or spider (also known as web robots) retrieves the precise contents along with more number of vulnerable data, due to the crawler's page navigations system. Search engines are simply bringing their responsive web pages based on assigned indexing values for all the visited links, without identifying hidden intrusion contents and web links on all the malicious pages. In the previous attacking era, spammers used (Spam) text for their attacks; later, many text-based filtering tools came into the picture to analyze spam text. Since images have very complicated features to extract its contents, after spam text identification, hackers nurtured their attacking methods based on the images, and they started to embed their spam contents into the image and spread it on victim's web pages or email id. These images are called spam images. Images are scanned for its text extraction by Optical Character Recognition (OCR) system, and then, the mined contents are matched with spam text databases for spam text identifications. Based on threshold values of matched contents, spam images are identified and applied to the remedial actions. Subsequent failures of said techniques, hackers started to attack targeted victim's data and system with the help of non-spam images. In which, they simply imbed their malicious web links and erroneous contents into the images, and those images are placed at the legitimate web pages in the form of some advertisements or stimulating user's desires. Navigating to these intruded sites causes to DoS attacks, data and security breaching of a victim's system. This paper is going to discuss web robots and its architecture, web content analysis and identification of intrusive substances of bogus images on web page contents.

---

S. Ponmaniraj (✉) · T. Kumar · A. K. Goel  
School of Computing Science and Engineering, Galgotias University, Greater Noida, India  
e-mail: [ponmaniraj@gmail.com](mailto:ponmaniraj@gmail.com)

T. Kumar  
e-mail: [tapas.kumar@galgotiasuniversity.edu.in](mailto:tapas.kumar@galgotiasuniversity.edu.in)

A. K. Goel  
e-mail: [amit.goel@galgotiasuniversity.edu.in](mailto:amit.goel@galgotiasuniversity.edu.in)

**Keywords** Classifier · Crawler · Harvest ratio · Intrusion detection system (IDS) · Optical character recognizer · Page indexing · Parser · Web robots

## 1 Introduction

Internet connects the entire world through search engines. Crawling web pages from various sites are extremely difficult in progress. Before the initialization process, the crawler has to parse the input key phrase for identifying the mandatory items to be searched. Once the input is given, then the parser analyzes that input terms then moving toward to get the relevant pages based on its scoring values and index values [1]. Fetching of entered key terms from a range of websites initiated with a single-server URL; then, it traverses along with various URLs till system stops the digging operation through web search engines with the help of crawling functions. Crawler classified into many types based on its searching progress such as a generalized crawler, focused crawler, profound crawler, and augmented crawler.

If any crawler doesn't speedily process the searching operations, then it leads to creating a path for security compromising of a targeted user system, and even some times it will act as a malicious crawler to intrude the system's activities. This kind of behaviors brings privacy issues of the legal and sensitive data which are being communicated from one URL to another. This could happen even if the search engine's crawler meets up with middle-size web pages. It affects the commotion of Internet industries healthy enhancement [2].

Ji et al. [3] analyzed web search engine's activities to identify its general functional operations, and they delivered their views on crawler activities in a few parts as follows; key phrase collectors; indexer for assigning value to the visited links; retriever for input queries from various URLs; user interface activities to pass and retrieve data. Crawl does the web search for content or code from the Internet; then, visited pages are given with index terms and values for further updates. Finally, the rank will be provided for all the visited web pages based on the threshold values assigned by page relevancies for solving the user queries [4].

Rashmi et al. [5] and Vijay et al. [6] focused on the problems, where user-facing it in mid of searching operations due to lack of security implementation over crawlers on search engines. They expressed their views on intrusion (attacks) as if any unwanted and unauthorized access of targeted victim's information and system's security functions over networks via floating of several advertisements and bogus links embedded with legal sites. Recognizing those intrusions using a set of protocol mechanisms is known as Intrusion Detection System (IDS). Recently, hackers started to create intrusions attacks by adding their malicious links with some legal sites. Due to this, spiteful links and codes victim will get continuous unwanted responses from some links or phishing mode of breaches. This continuous progress from any unwanted and unfaithful links causes to the Denial of Services (DoS) attacks.

Recently, attackers used images for deploying their bogus links on well indexed and reputed web pages which glide on the Internet. Forthcoming chapters of this

research paper discuss web robots—crawler’s architecture and functions, parsed web content analysis to find out weblink intrusions.

## 2 Intrusion Detection System

Basics of cryptographic concept explain different types of attacks done by hackers such as active and passive attacks. Active attacks harm to the data which are communicated between endpoints. This type of attack either performs modification of data’s originality or makes the data unavailable to the users when it is required, whereas the passive mode of attacking data will not alter the original content, and simply this attack used to listen to the conversation for getting content when it is being communicated. Both attacks are the key ways to amend the modification, harms to the data, affecting the system resources, etc. Masquerade, session replay, denial of services, traffic analysis, and release of message contents are the example for both active and passive attacks [7].

In most cases, above-said attacks are happening at network path and communications. Such unwanted commotions are soaking up on the system and network resources toward handle the legal and sensitive data. Any unwanted attacks or activity toward harming the data over the network are called intrusions. Proper checking and monitoring will help in avoid intruders to hack the system resources and data. A proactive mechanism for network intrusion is to get ideas on intruders attacking methods, observing the network paths in between end-users and responses for the identified attacks. Two important methods are used in intrusion attacks such as Intrusion Detection System (IDS) and Intrusion Prevention System (IPS). In which, IDS tools are used to detect the intrusion type of attacks and alert the users to get protected, whereas IPS used to protect users from intrusions, and it works based on found intrusion attacks and applying concerned rules against that attacks. IDS won’t protect the user’s data and resources from the intrusions. It simply generates an alert signal to end-users regarding attacks when it analyzed over a communication path.

### 2.1 Network Intrusion Detection System Architecture

Figure 1 shows that architecture of the network intrusion detection system, in which users are connecting with the World Wide Web through routers and gateways. All users are connecting themselves with the outside world via those routers that is why hackers choosing IP addresses to pass their malicious codes for attacking the victim systems and data. Network Intrusion Detection System (NIDS) uses a pattern recognition system for matching the attacking models to identify intrusion attacks. It generates a signal to the users if any pattern got matched on communication links.

There are many ways to attack a victim’s data through network intrusions. The following section explains a few attacking methods.



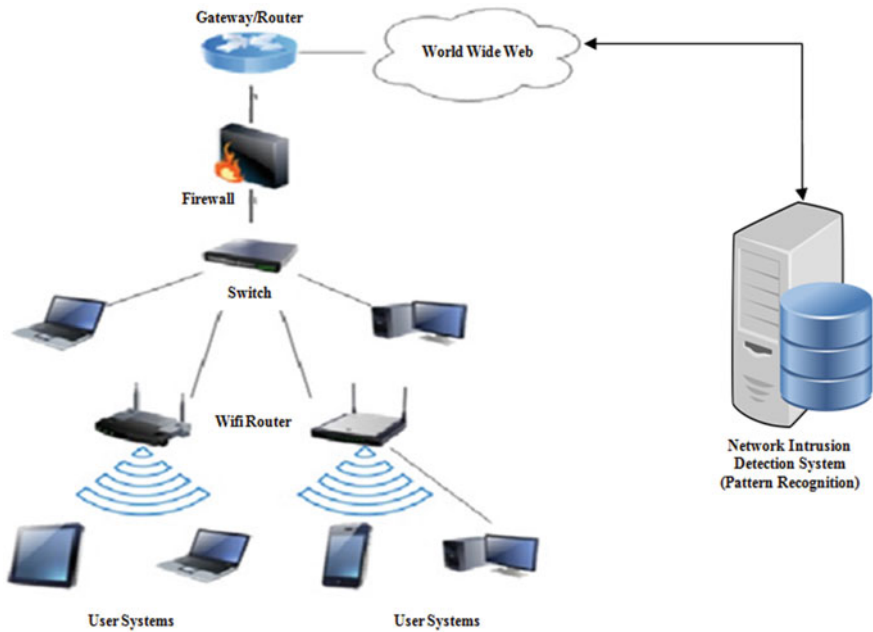


Fig. 1 Network Intrusion Detection System (NIDS) model architecture

**Source level attack:** Attacker uses an operating system-based security utilities to compromise system functionalities. In this attacking method, hackers are developing legitimate utilities for the operating systems, and in the back-end progress, they are attacking the victim’s system through loopholes of that deployed utility.

**Routing diversions:** In this method, the attacker uses numerous routing paths to reach the targeted victim system. Since hacker uses a variety of routing directions, detection or monitoring tools are unable to trace the malfunctioned IP addresses.

**Content overwriting:** This is the process of modifying the original content of transferred data. Attackers passing slow commands to the targeted systems which rewrite the normal data based on the commands written by attackers.

**Gateway script:** Commonly attacker writes the gateway script to fetch server details and file systems which are stored along with “..” symbol or “|” pipe symbol. Gateways are the routing place, where all the servers are reaching and passing through. Once the server got hacked, then it is very easy for the attacker to moving through servers toward victims system. To compromise server systems and files, the hacker uses gateway script.

**Protocol-based attacks:** IP spoofing is the best example of protocol-based attacks. By means of traffic analysis, hackers can spoof the victims IP addresses.

**Flooding:** Denial of Services (DoS) is the exemplar of a flooding attack. The continuous flow of request to the targeted server or the victim system causes the system’s resources to be busier at the moment. Due to resources unavailability,

requested data can't be shared with an opponent at the moment, and it will be looking like a crash or damaging of data, system functions as well.

The above-said methods are the little examples of network-based intrusion attacks, and few mitigations are there to avoid intrusions. The following matters are simple ways to protect systems and data from intrusion attacks.

- Formatting/deleting log files information
- Deploying well-secured encryption functions on departing data
- Analyzing the system files through Rootkit software
- Monitoring of every incoming and outgoing packet on a network path
- Firewall and security function deployment on network security infrastructure
- Update anti-virus software and system file patches
- Reading of warning information from system security functions before accepting any cookies and unwanted installations, etc.

### 3 Web Robots—Crawler Architecture

Web search engines are energetic through several important modules such as user interfaces (data acquisition), crawling (key phrase searching), parsing, assigning index and score value, page ranking (identification of relevancies). The above-mentioned functional modules are carrying smooth operations on finding the most relevant pages for user-requested queries. Firstly, the user has to enter key terms which they like to search on the Internet. This action will be done through user interface applications. Once the input passed onto searching functions, then crawler activated parser to extract the required data on Internet pages. From a million pages, the parser does analyze the contents of each page and returns the matched web pages to the user. Based on the index values, priority of a web page preferences will be assigned. If any pages not reflecting their score and index value, then it will be updated with the index value for future access and relevancy manipulation. Every crawler must be designed based on "Politeness Policies" to limit their traffic toward any web server for better performance [8]. Figure 2 represents the overall model for the web crawling functions to bring the relevant pages to the web robot.

In a traditional web search engine, users can possess only a few legal contents, and most of the searching contents for the user's request are digging out from imperceptible URL link sources, and they are called as deep net or deep web [9]. As per the networking research announcements, 80% data used by today's Internet databases are delivered by this hidden deep web URLs.

### 4 Web Crawler—Operational Methods

Web search engine depends on the process of crawler to access the targeted user contents. The crawler needs to get an input URL from initiator; then, it has to collect

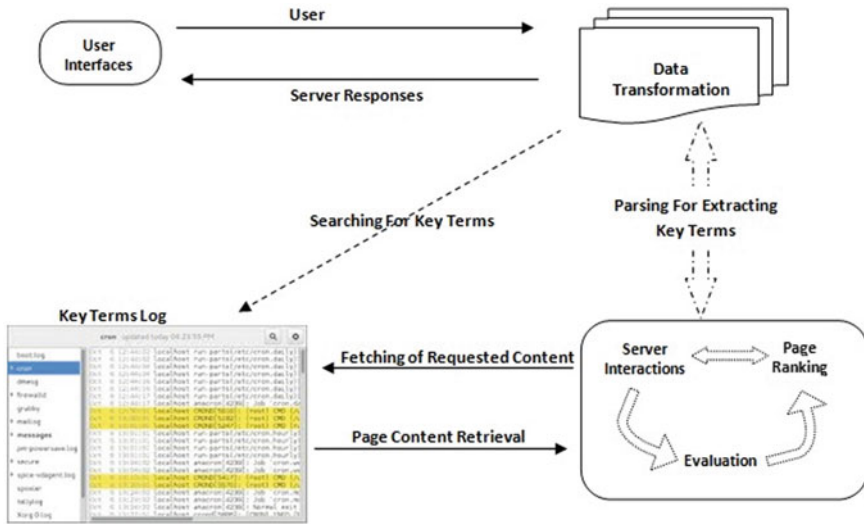


Fig. 2 Model architecture for web robot—crawler searching operation

the URL seeds to crawl the key phrases all over the Internet pages. Web robots and crawler comprises crawler, parser, and classifier modules to bring the relevant page contents as per entered search items from user interfaces [10]. In a modern web page designs, all pages are implemented with ease of access tag elements called semantic tag elements. Those tags are used to identify the location, where the problem arises. The subsequent semantic tags are imported all websites designed through HTML5 languages. `<header>`,

`<nav>`, `<figcaption>`, `<figure>`, `<section>`, `<aside>`, `<article>`, `<summary>`, `<time>`, etc.

Classification models are applied to identify the semantic data of any web sites, and these data are used as a training model dataset for matching the extracted content. If any crawler wanted to be automated, then these semantic elements must be configured as a supervised learning dataset a label for giving training to the searching elements and its patterns [11]. All the tag elements are used to find the navigation URL to the next page based on its empirical values to calculate their score value. This empirical value has been calculated by the means of “How often the pages are visited?,” “How much relevancy that visited page contains?.” To get the most relevant page for any key phrases, then all possible attribute properties of “contain tags” need to compute for finding their score values.

### 4.1 Parser

The parser does an analysis of key phrase items from the downloaded web pages by the crawler to retrieve the relevant page values and documents. In modern technology, so many evaluation schemes came to perform their assessments and now ‘‘SemEval (Semantic Evaluation)’’ method is applied to evaluate the identified key phrases for semantic elements [12, 13]. From the crawler parser identifying the labels *header, title, permLink, (H)Ref, Str, action, nav, and some meta data* related to search items. A simple example of defining parser is as follows *Parse. Parser = parser(body\_cont, Link Handler, Option.)*. The variance from multiple resource contents is analyzed using (Co) variances and eigenvalues of training set data. Spread out of any classes of values with the same covariance is computed as follows;

$$\text{Covariance} = \frac{\sigma_{\text{between}}^2}{\sigma_{\text{within}}^2} = \frac{(\bar{\omega} \cdot (\mu_1 - \mu_0))^2}{\bar{\omega}(\sum 0 + \sum 1)} \tag{1}$$

where

- S—Significant differences
- $\sigma$ —Covariance of different dataset
- $\mu$ —Mean value for the datasets
- $\omega$ —Linear combinations of features and means

$$\text{Relevancy}(\text{Page, Key}) = \frac{\sum_{k \in p \cap d} f_{kd} f_{kp}}{\min(\sum_{k \in p} f_{kp}^2) \cdot (\sum_{k \in p} f_{kp}^2)} \tag{2}$$

where

- ‘‘K’’ is search term frequencies
- ‘‘d’’ is key terms in a domain
- ‘‘p’’ is a page under exploring.

Use (1) and (2) for variance identification between different datasets with searching terms on all the web pages and relevancy of the key terms on the searching web pages. Once the relevancy value is found, then the page is assigned with indexing value for easy access in future.

### 4.2 Classifier

Classifications of searching entity with their related terms are very tricky and complicated. For an example, if any user wanted to search for mobiles, then it brings millions of pages with related items, and in parallel, it brings some unwanted matters like shapes and colors or the peripherals of the mobile devices or some other related terms used for mobiles. Such kind of behaviors wasting user’s time and network resources usage. Classifier makes this work simple and returns most pertinent matters to the

user queries. Quality of classifier is depending on the process of segregating the index terms and the segregation falls on two important categories; training processability and intrinsic severance of different datasets [14].

Classification leads to giving precision and recall value for the search engines to confirm page relevancy [15]. Classification for the web contents is different types based on the specific contents passed to search on the Internet. They are classified into topic classification, purposeful classification, attitude classification, etc. The classification is computed for the similar data items which are done by,

$$\text{Score}(d) = \sum_{t \in q} (tf(t \text{ in } d) * idf(t^2) * boost(t.\text{field in } d)) * \text{LengthNorm}(t.\text{field in } d) * \text{coord}(q, d) * \text{QueryNorm}(q) \tag{3}$$

where

$t$  = terms

$tf(t \text{ in } d)$  = Term frequency factor in document ( $d$ )

$\text{lengthNorm}(t.\text{field in } d)$  = Normalization field value

$\text{coord}(q, d)$  = Coordination factor of a query terms in document ( $d$ )

$\text{queryNorm}(q)$  = Query normalization value.

Relevancy of the classified data from a document depends on four different categories such as true positive (Relevant documents only present), true negative (Identifying non-relevant documents), false positive (System correctly identifying non-relevant documents), and false negative (Omitting relevant documents instead selecting them) [16]. For an example, relevance between two different datasets (A) and (B) is derived as follows based on searching occurrences from one document value on another. After precision and recall manipulation, F1—measure must be taken into the place to balance the dataset differences. The terms precision, recall, F1—Measure and Accuracy are described as follows;

$$\text{Precision} = \frac{|A \cap B|}{|B|} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FP}|} \tag{4}$$

$$\text{Recall} = \frac{|A \cap B|}{|A|} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FN}|} \tag{5}$$

$$F_1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Accuracy} = \frac{|\text{TP}| + |\text{TN}|}{|\text{TP}| + |\text{TN}| + |\text{FP}| + |\text{FN}|} \tag{7}$$

In (4), (5), (6), and (7),

A = Relevant dataset B = Irrelevant dataset

TP = True positive value FP = False positive value TN = True negative

FN = False negative value.

### 4.3 JSON Classifier

For a crawler, to identify intrusion attacking terms and protocol overhead parameters, it has to analyze crawled web page contents with custom classifiers. Custom classifiers are created with the help of XML tags, JavaScript Object Notation (JSON), and Comma Separated Values (CSV) formats. A crawler needs to call custom classifiers to manage imbalanced of built-in classifiers to return the recognized dataset. JSON format defines the data structure with name-value pairs or order of list values. The dot or bracket notations are used in JSON classifiers to define a path object to create table schema. Dot and bracket notations in JSON format is used to specify a child field of a JSON object. “\$” symbol represents for root elements. “\*” symbol is used to mention numeric or name-value presences in the JSON path [17]. The following statements are the example for JSON scripts,

```
[
  {
    "type": "Wikipedia",
    "id": "wikipedia/org/wiki/Main_Page",
    "name": "Main_Page"
  },
  {
    "type": "Wikipedia",
    "id": "wikipedia/org/index/title_Main_Page/action=edit",
    "name": "Wikipedia's View Source"
  }
]
```

In the above example type, id and name label for the classifying items using JSON classifier format. When JSON uses \$[\*] notation, it means that all the records will be maintained in the array under JSON object. The root notation for the above statements are defined as follows,

```
root
|-- record: array
```

And inside of this root record array contains,

```
root
|-- type: string
|-- id: string
|-- name: string
```

When a crawler goes for built-in JSON classifier, it gives the schema at the end as follows;

```

root
|-- type: string
|-- id: string
|-- name: string
|-- identifiers: array
|  |-- element: struct
|  |  |-- scheme: string
|  |  |-- identifier: string
|-- other_names: array
|  |-- element: struct
|  |  |-- lang: string
|  |  |-- note: string
|  |  |-- name: string

```

“String” in JSON finds the label for the record of crawled pages. Each record maintains the anchor text and many reference links to be visitor being visited. All those reference links and names will be stored in the record to access the relevant page contents.

#### 4.4 Open-Source Crawler’s Functionalities

The behavior of a crawler depends on different policies used in fetching and visiting referenced links from the source URL seed. The outcome of every crawler follows any of the following policy methods [18];

- Selection policy—which pages to be downloaded
- Revisit policy—Checks for changing the web pages and its contents
- Politeness policy—Avoids overloading of traffics from various sources
- Parallelization policy—Coordination with distributed server system for better performance

*Selection policy:* This policy works on the base principle of Online Page Importance Computation (OPIC) like page ranking algorithm for selecting useful and related pages. Partial page rank computation, backlink, breadth-first search algorithms are working well on this policy progresses.

*Revisit policy:* Revisiting policy is working for identifying two important cost factors such as freshness and age. This cost of a crawler varies on searching page’s qualities to find outdated copies or accuracy of the local copy. At time factor  $T$ , the freshness of any pages  $P$  is defined as follows,

$$F_P(T) = \begin{cases} 1 & \text{If } P \text{ is equal to the local copy at time } T \\ 0 & \text{Otherwise} \end{cases} \quad (8)$$

In a repository, the age of a page  $P$  at time  $T$  is computed to identify the outdated characteristics of the page  $P$ . It is defined as,

$$A_P(T) = \begin{cases} 0 & \text{If } P \text{ is not modified at time } T \\ T & \text{Otherwise} \end{cases} \quad (9)$$

*Politeness policy*: Mainly, this policy is applied to the crawlers to avoid collisions and traffic from multiple sources at the server-side. The weakness of any poorly coded crawler is, it simply get crash servers or routers when it tries to download heavily loaded page contents. Server overload will be the result yielded by the poorly designed crawler.

*Parallelization policy*: To avoid duplication of downloading same page contents from any website, parallelization policy is emerging with crawlers. It supports running multiple processes in parallel, and due to this copying same contents, many times will be avoided.

## 5 Implementation of Intrusion Analysis of Parsed Web Contents

For intrusion analysis from downloaded web documents, first, the crawler has to separate the keywords and index terms. Then, from the classifier's output database files, key terms need to pass through for relevancy comparison. Matching an element from a single web document doesn't a matter of complex operations but when the keywords are searching from thousands of pages took much complex operation and that moment crawler must be activated for automated classifications for clustering the index terms to easy search [16, 19, 20]. The searching key terms from the Internet works as follows;



**Implementation Algorithm:**


---

*Visit* a child node  $a^*$  (Starting value shouldn't be 0)

*Check* for the relevancy

*If*(child node  $a^* > 0$ )      $\forall a^*$  is a relevant page

*If*  $a^*$  is a relevant page then add it to the tail (left side) of an initiator

*Make*  $a^*$  as a current node

*Search* for anchor text  $atx^*$  for an image

*Let* anchor text  $atx^*$  be the contents of *child node, reference, text data,*

*anchor text for something else*

*Compute*  $atx^*$  relevancy, if ( $atx^* >$  relevancy factor (Greater than 0))

*Set*  $atx^*$  as current node as a child of its existed node

*Else*

*Move*  $atx^*$  to the irrelevant list (*Intrusion links*) at right side

*Move* next  $atx^*$  value for identifying relevant items

*Continue* process till system stops

*Add* it to the right side of the parent tree where irrelevant pages (*Intrusion Pages*) stored already

*Continue* for the next items to search relevant

*Stop* the process after identifying all the relevancies and indexing the visited links.

---

After identifying the entire irrelevant page documents, those lists will be stored into the intrusion labeled database list to stop irrelevant web link access. Then, intrusion database list must be eliminated for future protections, and it will be done by a harvest ratio value for eliminating irrelevant pages. Harvest ratio value must be high for the better throughput in terms of removing the irrelevant pages; otherwise, crawler may spend a lot of time for reduction of extraneous web documents. Harvest ratio value is calculated by,

$$\text{Harvest\_Ratio} = \frac{\text{Relevant\_Pages}}{\text{Pages\_Downloaded}} \quad (10)$$

Harvest rate is the measurement for finding the relevant searched pages on the Internet by the crawlers. High harvest rate means that a high level of true positive values for the searching terms. Simply harvest rate satisfies the crawling target pages CTP among the crawled pages CP. The ratio value of a harvest rate must be high for better throughput; otherwise, a crawler looks for more time to eliminate non-relevant pages. In such situations, new crawlers will be used instead of an older one.

Few open-source crawler functionalities are analyzed based on four assessment values such as average, maximum, weighted, and eight-level data with a set of parameters text link, menu link, string link, redirections, statistic and references, scripts links, etc., and the open-source crawlers are searching for the relevant web documents. Table 1 shows that summary comparison for open-source crawlers to identify the relevant web documents from Internet [21].

Form the above-mentioned table, analysis WebCopierPro crawler shows the best performance on all the attributes compared with other open-source crawlers.

### **WebCopierPro Features**

**Table 1** Open-source crawler’s summary comparison

Link possibilities	Heritrix (%)	Nutch (%)	Pavuk (%)	Teleport (%)	Web2 desk (%)	WebCopier Pro (%)	WebHTTrack (%)
Text link	2–100	2–100	2–100	2–100	2–100	2–100	2–100
Href = Java	6–50	3–25	0–0	4–33	5–42	12–100	3–25
Menu link	6–50	3–25	0–0	3–25	4–33	12–100	2–17
Flash link	0–0	0–0	0–0	0–0	0–0	0–0	2–50
Applet link	2–50	0–0	0–0	0–0	0–0	0–0	2–50
Redirects	6–60	6–60	2–20	6–60	4–40	0–0	6–60
Class or java link	0–0	0–0	0–0	0–0	0–0	2–50	0–0
Ajax link	0–0	1–50	0–0	0–0	0–0	0–0	0–0
Links with a	2–50	2–50	0–0	3–75	2–50	0–0	2–50
VbScript link	3–75	3–75	0–0	3–75	3–75	0–0	2–50
Static string link	26–62	19–45	4–10	8–43	22–52	16–38	22–52
Concatenated string link	6–50	2–16	0–0	1–8	1–8	12–100	1–8
Special	1–6	2–13	0–0	5–31	1–6	12–75	0–0
String link tests passed	33–47	23–33	4–6	24–34	24–34	40–57	23–32

WebCopierPro is used as a crawler for search engine progress, and it is used to download the web pages when it is in online or offline. It also monitors the web page contents before downloading from servers. Some of the main features of WebCopierPro tool are listed below [22];

- It is an offline browser and allows users to browse their requested pages without an Internet connection.
- Drag and drop project feature are available in this tool.
- It supports for a larger volume of web page content downloads (Offline downloading also possible).
- Downloaded web contents are exported in different formats such as MHT, CHM, and TXT.
- This tool allows users to switch in between project list, contents tree, and log file windows.

## 6 Conclusion

Web search engines are utilizing crawlers to search and satisfy the user-requested queries over the Internet using searching key phrases in all the navigating URL

documents. Before navigating toward child URL link, a query image or an image which contains anchor text for referring outbound links must be analyzed for its spam contents by using OCR tools to avoid continuing the travel around, and from the source location itself, a system can partially conclude that whether to take move on to the next link or not. This kind of spam protection avoids intrusion access at the beginning of crawling operations. Still, crawler of any search engines or spiders has to parse and analyze an alluded image URL links to identify and avoid intrusive attacks from all over Internet-connected systems. In future, this kind of intrusive and false web navigation by any non-text-based notion such as images, GIF animation, styled and attractive buttons is monitored and controlled by automated web crawler IDS implementation.

## References

1. Sun L-W, He G-H, Wu L-F (2010) Research on the web crawler. *Comput Knowl Technol* 6(15):4112–4115
2. Quan B, Gang X et al (2014) Analysis and detection of bogus behavior in web crawler measurement. *Procedia Comput Sci* 31:1084–1091 Elsevier
3. Ji X (2009) Research on web service discovery based on domain ontology. In: 2009 2nd IEEE international conference on computer science and information technology. <https://doi.org/10.1109/iccsit.2009.5234756>
4. <https://moz.com/beginners-guide-to-seo/how-search-engines-operate>
5. Rashmi R et al, IDS based network security architecture with TCP/IP parameters using machine learning. Electronic ISBN: 978-1-5386-4491-1, (PoD) ISBN: 978-1-5386-4492-8
6. Vijay R et al, A novel survey on intrusion detection system and intrusion prevention system. *IJSTR*. ISSN 2277-8616, 2019
7. <https://www.rsaconference.com/industry-topics/blog/network-intrusion-methods-of-attack#:~:text=A%20network%20intrusion%20is%20any%20unauthorized%20activity%20on%20a%20computer%20network.&text=In%20most%20cases%2C%20such%20unwanted,network%20and%20for%20its%20data>
8. MARC NAJORK, Web Crawler Architecture, Microsoft Research, Mountain View, CA, USA
9. Agrawal S, Agrawal K (2013) Deep web crawler: a review. *IJIRCST*. 1(1). ISSN: 2347- 5552
10. Akilandeswari J, Gopalan NP (2008) An architectural framework of a crawler for locating deep web repositories using learning multi-agent systems. In: 2008 Third international conference on internet and web applications and services. <https://doi.org/10.1109/icwi.2008.94>
11. <https://medium.com/teleporthq-io/understanding-the-web-parsing-web-pages-semantically-805ef857854d>
12. <https://en.wikipedia.org/wiki/SemEval>
13. Tharwat A et al (2017) Linear discriminant analysis: a detailed tutorial. IOS Press. <https://doi.org/10.3233/aic-170729>
14. [http://www.ra.ethz.ch/cdstore/www6/Posters/725/Web\\_Search.html](http://www.ra.ethz.ch/cdstore/www6/Posters/725/Web_Search.html)
15. Zhu Z, Improving search engines via classification. <https://www.dcs.bbk.ac.uk/site/assets/files/1025/zheng.pdf>
16. Chen X, Zhang X (2008) HAWK: a focused crawler with content and link analysis. In: 2008 IEEE international conference on e-business engineering. <https://doi.org/10.1109/icebe.2008.46>
17. <https://docs.aws.amazon.com/glu/latest/dg/custom-classifier.html>
18. [https://en.wikipedia.org/wiki/Web\\_crawler](https://en.wikipedia.org/wiki/Web_crawler)

19. Chakrabarti S, van den Berg M, Dom B (1999) Focused crawling: a new approach to topic-specific web resource discovery. In: Proceedings of the 8th international world wide web conference
20. Hersovici M, Jacovi M, Maarek YS, Pelleg D, Shtalhim M, Ur S (1998) The shark-search algorithm. An application: tailored web site mapping. *Comput Netw ISDN Syst* 30(1–7):317–326. [https://doi.org/10.1016/s0169-7552\(98\)00038-5](https://doi.org/10.1016/s0169-7552(98)00038-5)
21. Yadav M, Goyal N, Comparison of open source crawlers—a review. *IJSER*. ISSN 2229-5518, 2015
22. <http://www.filefacts.com/webcopier-pro-info>

# Plant Species Identification Using Convolutional Neural Network



Harsha H. Ashturkar and A. S. Bhalchandra

**Abstract** The plant plays a vital role in human survival, so the preparation of the plant database and its identification system is essential for maintaining biodiversity. From the past two decades, plant identification based on leaf classification is on the main focus of researchers. It is proposed to build the convolutional neural network model for plant identification. However, overfitting is a significant issue in such a network. A regularization strategy like dropout is a good approach for addressing this issue. Along with this, transfer learning and parameter fine-tuning have come out with a good result on the same database. In the proposed system, performance of convolutional neural network is measured in terms of obtained validation accuracy.

**Keywords** Convolutional neural network · Overfitting · Pre-trained network · AlexNet · Classification

## 1 Introduction

Plant identification is rudimentary to agriculture, forest management, rural medicine, and other trading or commercial applications. Furthermore, automating plant identification is the most encouraging solution toward bridging the gap in Botany and Computer Science community. Plant species identification can be done through several geometrical, and its aspect attributes exhibited in its stem, roots, flowers, leaves, and fruits. Nonetheless, for many of these attributes, the inter-species dissimilarities are often indistinct and insignificant. However, plant leaves accommodate rich details for species identification because mainly leaves stay connected to plant

---

H. H. Ashturkar (✉) · A. S. Bhalchandra  
Department of Electronics and Telecommunication Engineering, Government College of Engineering, Aurangabad, Maharashtra, India  
e-mail: [harshaashturkar17@gmail.com](mailto:harshaashturkar17@gmail.com)

A. S. Bhalchandra  
e-mail: [asbhalchandra@gmail.com](mailto:asbhalchandra@gmail.com)

for a longer duration compared to fruits and flowers as they are not seasonal, and secondly, leaves have reasonably distinguishing property features like texture and shape. Plant identification is a quite tough job also for skilled botanists. It is also unfeasible for the ordinary people due to large digit of plant species accessible on earth. In computer vision, many researchers have worked on plant identification using leaf using the conventional method which depends on hand-crafted attributes like shape [1–3], local binary features [3, 4], histogram-oriented gradient [4–6], SIFT [7], and their combinations and trained them on SVM classifiers [3, 5] or various neural networks [8].

Although various promising techniques of plant species identification are proposed, the capability of an attribute or feature learning is not completely utilized. Besides, it is uncomplicated to realize low-scale features. Whereas, achieving mid-scale and high-scale features are fairly tough without implementing learning procedures. Convolution neural networks (CNNs) can grasp common features from data. The capability of convolutional neural networks is utilized by numerous researchers to exceed traditional techniques based on hand-picked features.

The simple outline of a convolutional neural network with some overfitting minimizing techniques is competitive to conventional perspective. Along with this, the employment of a pre-trained network like AlexNet to extract more discriminative features from plant leaf images is done.

The remaining paper is arranged as follows: a review of related work in Sect. 2. A detailed explanation of the used convolutional neural network along with details of layers in Sect. 3. The report of results obtained during the experiment is in Sect. 4. The conclusion is drawn which is in Sect. 5.

## 2 Related Work

In the past two decades or more, the work of plant species identification is mainly based on plant organs like leaves, fruits, flowers, bark, etc., among all, leaves that are comfortably available throughout the year. Arunpriya et al. [1] extracted digital morphological features of leaves, classify it using the support vector machine (SVM), and tested it on the Flavia dataset. They compared the result with the K-NN classifier and observed that the SVM approach outperform the K-NN approach. Beghin et al. [9] extracted shape and texture characteristics as a feature, and the results of this method were combined using incremental classification algorithm, and they acquired good accuracy compared to the use of single feature characteristics. Raut et al. [6] extracted parameters of the digital signature of leaf shape for plant recognition, where Harsha et al. [4] concatenated HOG and LBP features and acquired better accuracy than that of a single feature.

All these methods represent conventional feature extraction. In recent years, deep learning methods like CNN showed excellent performance in object detection and classification. The above conventional methods are useful for getting low-level

features and not for mid-level or high-level features as a learning method is not introduced.

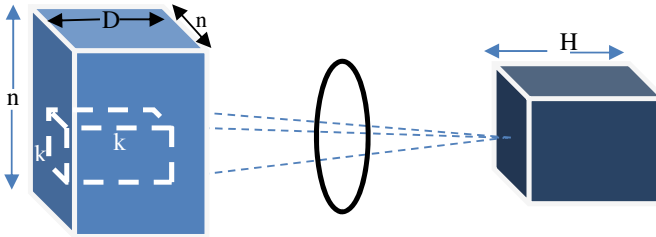
Lio et al. [10] have used the CNN because of its strong ability of feature extraction and tested their model on 32 species of Flavia database and achieved overall high accuracy. Ying [11] described various overfitting reducing techniques that help to generalize the model. The author stated concisely about overfitting from the perspective of causes and its solutions. Almisreb et al. [12] used a transfer learning approach for solving a classification problem with fewer amounts of data. They used a pre-trained Alexnet convolutional neural network to classify their data for ten classes. Some fine-tuning of the network has to be done to adjust the network according to data and concluded that the transfer learning approach gives good results even on a small amount of data.

### 3 Proposed Method

#### 3.1 Convolutional Neural Network

The convolutional neural network model was originally designed more than two decades ago to operate several arrays of information such as signals, color images, and videos even. Now, due to the handiness of extensive datasets and with development of technology such as graphic processing units (GPUs), CNN is presently a commodity in the computer vision area. In a conventional outline of pattern recognition, feature descriptors are hand crafted. It can correctly represent highly precise properties that are useful for classification or prediction. Convolutional layers are capable of drawing local features since they limit the receptive field or area of hidden layers to be local. CNN has appeared as a fruitful structure for expressing features and specifications in image processing. It does not require to evaluate features engineering which is longstanding and effort consuming. The generalization of the process forms its empirical and adaptable perspective to the many application issues of classification and recognition.

**Convolutional layer:** Different attributes of the input image are extracted by the convolutional function. Lower measure attributes are pullout by initial convolutional layers, whereas higher scale layers withdraw high-measure attributes. Figure 1 describes tasks performed by 3D convolution used in CNNs. Let,  $n*n*D$  is input size, and  $k*k*D$  is the size of each kernel. The convolutional function is computed with input and  $H$  kernels individually and generates separate output attributes, so for  $H$  number of kernels, separately  $H$  features are generated. The kernel is evaluated on the whole image from the left-top corner to the right-bottom corner. Corresponding to this location, each feature in output carry  $(N - K + 1) * (N - K + 1)$  elements. If the size of the convolution kernel is tiny, then it is not capable to reflect the input image position feature. On the conflict, the difficulty of the feature may outstrip the expressiveness of a broad convolution kernel.



**Fig. 1** Convolutional process

**Pooling layer:** Pooling minimizes the bulk of the feature map. It makes the features strong against distortion and noise. Pooling restores the outcome of node at a particular position using a summary statistic of adjacent location. Pooling can be of various types: average pooling, max pooling. In both types, the input is split into non-overlapping two-dimensional spaces. Max pooling indicates to the maximum output inside rectangular neighborhood while average pooling indicates to average output. Pooling assists to make output roughly invariant to little translation as well as down samples total feature map independently shrinking width and height, keeping the depth unharmed (Fig. 2).

**Activation function:** Linear functions are simple to compute however is limited due to their complexity and that is why non-linear functions are employed. Activation function supplies the elementary unit that can be used frequently in a higher dimension of the network model so that, incorporating with an attenuation matrix to change the weights of signal from layer to layer to fit an arbitrary and complex function.

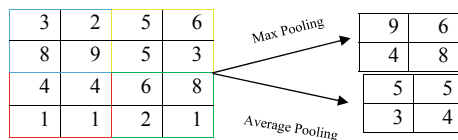
Rectified linear unit (ReLU) permits for rapid and more productive training by plotting negative numbers to zero and maintaining positive numbers as it is.

$$\text{ReLU} = \max\{0, x\} = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (1)$$

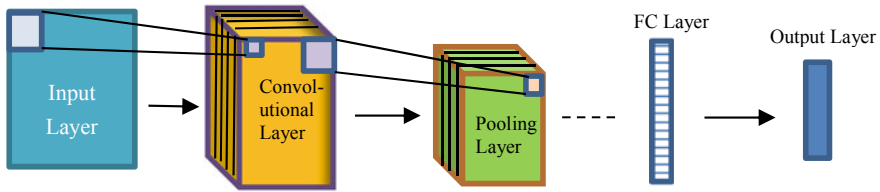
The above three operations are replicated over many layers, with every layer learning to expose different features.

**Fully connected layer:** These are frequently utilized as finishing layers of a CNN. FC mostly used to plot the particular features to image labels. Mathematically, these

**Fig. 2** Pictorial representation of max and average pooling with window size 2\*2 and stride of 2







**Fig. 3** General architecture of convolutional neural network

layers sum a weighting of the preceding layer of attributes, showing the accurate merge of elements to decide a certain target output result.

**SoftMax layer:** SoftMax logistic regression is used for multiclassification. A SoftMax layer is normally the final layer used in the neural network model. The SoftMax layer should have the exact number of nodes as that of output later. SoftMax allocates decimal probabilities to every class in a multi-class problem.

The paper proposed three convolutional layers and three max pooling layers as an indigenous neural network. The activation function used is ReLU. The learning rate was fixed as the reducing index function so that the model can better converge. The general CNN structure is shown in Fig. 3.

With varying kinds of images and with different input image sizes, the architecture of convolutional neural network varies. In this work, the input image size is considered to be 227\*227 pixels. Architecture is described in Table 1. Every convolutional layer is followed by a ReLU activation function, and the max pooling method is used. Finally, a fully connected layer has n inputs that correspond to the n categories of leaf data, and at last, SoftMax loss is assigned.

After training the convolutional neural network, it was observed that the model worked well on labeled training data but not on unseen testing data. That means generalization capability acquired by the model is not satisfactory. Overfitting occurs when a model performs well with training data and fails to perform well on test data. Precisely, the model learns the noise patterns which are present in the labeled training data; hence, it created a gap between the training and validation accuracy. To reduce overfitting, multiple solutions are available. Early stopping strategy, network reduction strategy, data augmentation strategy, regularization strategy, etc.

**Table 1** Proposed CNN architecture

Characteristic	L1		L2		L3	
	Conv	Pool	Conv	Pool	Conv	Pool
Filter size	5*5	2*2	3*3	2*2	3*3	2*2
Stride	2	2	2	2	2	2
Number of filters	100	100	250	250	250	250
Output size	112*11	56*56	27*27	13*13	6*6	3*3

**Dropout layer:** Dropout is the most well-known and helpful approach in case of overfitting on CNN. The main point of dropout is to randomly drop units and related connections during training. This stops units from co-adapting. The usual procedure is as follows [11]:

1. Drop arbitrarily half of the hidden neurons to build a new straightforward network
2. Train the decreased network utilizing stochastic gradient
3. Reinstate the detach neurons
4. Randomly separate half of hidden neurons from the reinstate network to build a new thinned network
5. Redo the operation above until getting an ideal set of parameters.

### 3.2 Pre-trained Network

Another approach that is implemented by using a pre-trained network called AlexNet with modifications in the last few layers according to application. AlexNet is a convolutional neural network designed by Alex Krizhevsky in 2012 [13]. Transfer learning gives the potential to use the pre-trained networks by tweaking it with application-specific data. The main part of transfer learning is to reuse knowledge obtained in a prior training process, to enhance the learning policy in a new or extra complex mission. That is, to say, transfer learning provides a proper solution for speeding up the learning mechanism in image recognition, image classification.

The output size of any convolution layer depends upon factors like stride ( $s$ ), filter size ( $f$ ), size of the input ( $n$ ), and padding amount ( $p$ ), that is, given by

$$f(z) = ((n * 2p - f/s) + 1) \quad (2)$$

If the output is not an integer, then it is rounded down the nearest integer using the floor function. The dataset used in the proposed system is much minor and different than that of the ImageNet database which is used to train the original AlexNet, and hence, changes in some last layers are required to get the required class labels at the output. Since the dataset is small, overfitting may occur. Fine-tuning of a higher level of the network is to be done as beginning layers are designed to extract generic features only (Fig. 4; Tables 2 and 3).

The earlier layers of AlexNet CNN are retained as a fixed feature extractor for leaf dataset. In the first move of transfer learning, the last three layers of the pre-trained network are replaced with the set of layers which can classify 20 classes to classify the 20 subjects. To adopt the new output (20 subjects), a fully connected layer of filter size 128\*128 is added. Then, the rectified linear unit (ReLU) layer is connected to upgrade the non-linear problem-solving capability moreover; ReLU avoids vanishing gradient effect, and it also trains the model at particular times faster. To get 20 neurons at the output, one more fully connected layer is added to classify the 20 categories.

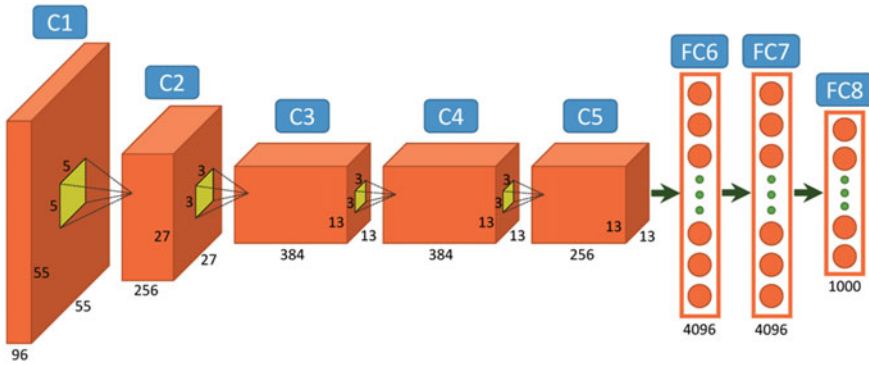


Fig. 4 AlexNet architecture with output size of each convolution [15]

### 4 Experimental Results

Plant species identification is implemented by two methods. The first one is by designing a CNN model along with regularization techniques to avoid overfitting and improve generalization. To generalize model more accurately, another methodology, i.e., pre-trained AlexNet by fine-tuning its parameters is implemented.

The proposed method is applied to leaf samples from the leafsnap dataset [14]. Leafsnap database contains images captured by the mobile camera in controlled light, and another source is the high-quality laboratory of pressed leaves. Leafsnap database considered a total of 185 plant species from the northeastern United States with varying numbers of images per species. Leaf images of plant species that are considered in the proposed model are as shown in Fig. 5. Experimentation is performed by both the techniques of CNN on 20 random species of plants taken from field images. The experiment is performed using deep learning toolbox in MatLab 2018a programming environment on a computer with Windows 10, 64-bit operating system.

Training is performed on a single GPU with a constant learning rate. In all 1000 images are used for training and testing purposes; out of which, 70% of images are used for training, where each class contains the varying number of leaf images. Training options used while training specifies the parameters and their corresponding values as shown in Table 4. The convolutional part requires very a smaller number of parameters than that of the fully connected layers. The optimizer used is stochastic gradient descent with momentum. The minimum batch size is set low, to boost up training procedure and keep the model away from low memory issues. The learning rate is set at least to slow the learning pace so it can catch up with the previous layers.

The performance of the model is monitored and measured by using classification related metrics like accuracy, precision, recall, F\_score. The confusion matrix is the primary requirement for calculating all these metrics.

Accuracy is the simplest and most commonly used form of metrics.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (3)$$

**Table 2** AlexNet architecture

Layer	Layer type	Layer detail	Name
1	Image input	227 × 227 × 3 images with “zero-center” normalization	Input
2	Convolutional	96 11 × 11 × 3 convolutions with stride [4 4], padding [0 0 0 0]	Conv1
3	ReLU	ReLU	Relu1
4	Normalization	Cross-channel normalization with 5 channels per element	Norm1
5	Max pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]	Pool1
6	Convolutional	256 5 × 5 × 48 convolutions with stride [1 1], padding [2 2 2 2]	Conv2
7	ReLU	ReLU	Relu2
8	Normalization	Cross-channel normalization with 5 channels per element	Norm2
9	Max pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]	Pool2
10	Convolutional	384 3 × 3 × 256 convolutions with stride [1 1], padding [1 1 1 1]	Conv3
11	ReLU	ReLU	Relu3
12	Convolutional	384 3 × 3 × 192 convolutions with stride [1 1], padding [1 1 1 1]	Conv4
13	ReLU	ReLU	Relu4
14	Convolutional	384 3 × 3 × 192 convolutions with stride [1 1], padding [1 1 1 1]	Conv5
15	ReLU	ReLU	Relu5
16	Max pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]	Pool5
17	Fully connected	4096 fully connected layer	Fc6
18	ReLU	ReLU	Relu6
19	Fully connected	4096 fully connected layer	Fc7
20	ReLU	ReLU	Relu7
21	Fully connected	1000 fully connected layer	Fc8
22	Softmax	Softmax	Prob
23	Classification	crossentropyex	Classification

In many cases, accuracy is not a good indicator for performance, especially when class-specific performance is to be measured in an unbalanced data. Precision is used in such cases. It is calculated as

$$\text{Precision} = \text{True\_Positive} / (\text{True\_Positive} + \text{False\_Positive}) \quad (4)$$

**Table 3** Fine-tuning of pre-trained AlexNet network

Layer	AlexNet		Fine-tuned AlexNet	
	Layer type	Layer detail	Layer type	Layer detail
21	Fully connected	1000 fully connected layer	Fully connected	128 fully connected layer
22	Softmax	Softmax	ReLU	ReLU
23	Classification	Crossentropyx	Fully connected	20 fully connected layer
24			Softmax	Softmax
25			Classification	Crossentropyx

**Fig. 5** Leaf species used from leafsnap database**Table 4** Training options

Epochs	30
Learning rate	0.0001
Training function	SGDM
Validation frequency	50
Iterations per epoch	45

Recall is defined as a fraction of samples from a class which are correctly predicted by the model.

$$\text{Recall} = \frac{\text{True\_Positive}}{\text{True\_Positive} + \text{False\_Negative}} \quad (5)$$

In some applications, both recall and precision are important; F1-Score is a combined way of these two in a single metric, mathematically expressed as Eq. 6.

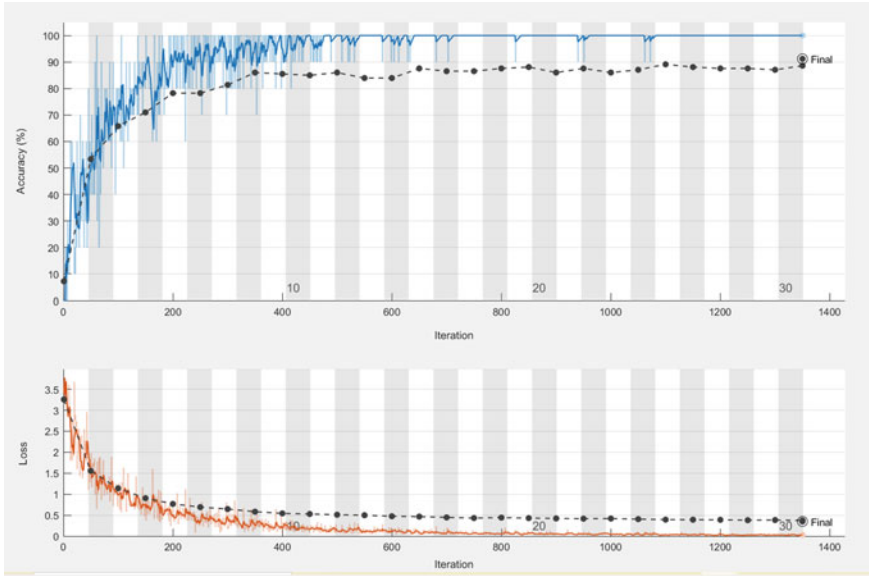


Fig. 6 Training and validation results with loss rate of designed convolutional neural network

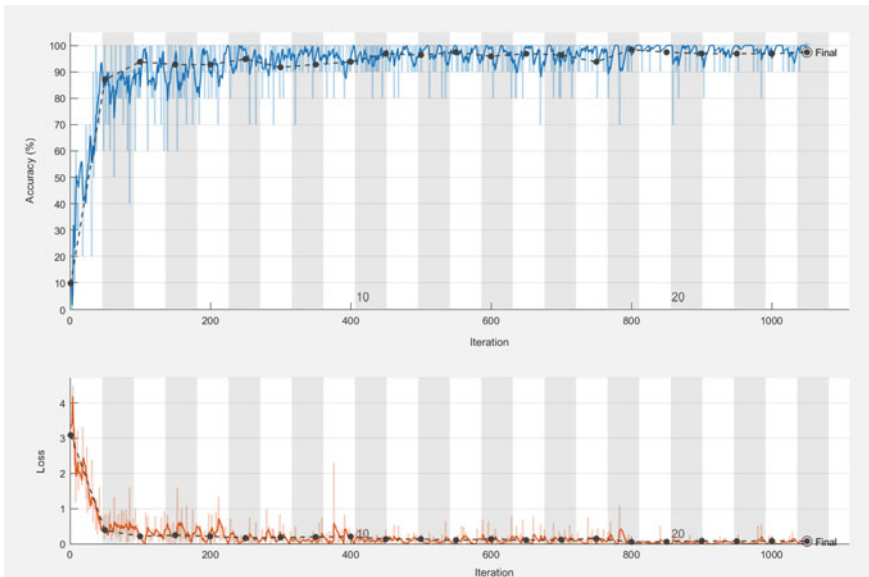


Fig. 7 Training and validation results with loss rate of transfer learning

**Table 5** Performance metrics

S. No.	Method	Accuracy	Precision	Recall	F1_score
1	Designed CNN	91.19	88.90	91.57	92.03
2	Fine-tuning of pre-trained network	97.41%	97.58	97.95	98.53

$$F1\text{-score} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall}) \quad (6)$$

Table 4 shows the accuracy and other performance metrics results obtained by both models on the validation set. Both models are trained on the same parameters mentioned in Table 3 but the AlexNet has met the validation criteria after 1100 epochs. Though the dropout method greatly improved the accuracy of the CNN network designed from scratch, still as compared to the results of training the model from the scratch; AlexNet fine-tuned model has performed better with a classification accuracy of 97.41%. Due to a greater number of layers and parameters, pre-trained network required more time for computation than that of network designed from scratch.

## 5 Conclusion

The paper proposed a deep learning method that outperforms conventional machine learning techniques even in the presence of a relatively smaller number of training samples. Our previous work based on concatenated hand-crafted features acquired an accuracy of about 80%. The design of an improved neural network by using the dropout technique reduces overfitting. Motivated from the fact that the transferred CNN performs better compared to the CNN trained from the scratch; hence, pre-trained AlexNet model with application-oriented fine-tuning is used, and the accuracy rate achieved is about 97%.

However, generalizing ability can be improved by using a larger database. In the future, an attempt will be made to recognize leaves that are attached to the branches without having the plain background to develop an automated plant identification system including its implication in the open-set recognition task.

## References

1. Priya CA, Balasaravanan T, Thanamani AS (2012) An efficient leaf recognition algorithm for plant classification using support vector machine. In: International conference on pattern recognition, informatics and medical engineering (PRIME-2012), Salem, Tamilnadu, pp 428–432. <https://doi.org/10.1109/icprime.2012.6208384>

2. Chathura Priyankara HA, Withanage DK (2015) Computer assisted plant identification system for Android. In: Moratuwa engineering research conference (MERCCon), Moratuwa, pp 148–153. <https://doi.org/10.1109/mercon.2015.7112336>
3. Ren XM, Wang XF, Zhao Y (2012) An efficient multi-scale overlapped block LBP approach for leaf image recognition. In: Proceedings of the 8th international conference on intelligent computing theories and applications (ICIC'12). Springer, Berlin, pp 237–243. [https://doi.org/10.1007/978-3-642-31576-3\\_31](https://doi.org/10.1007/978-3-642-31576-3_31)
4. Ashturkar H, Bhalchandra AS, Behare M (2020) Enhanced HOG-LBP feature vector based on leaf image for plant species identification. In: Presented in recent trends in image processing and pattern recognition 2020, Springer, Dr. Bamu, Aurangabad
5. Pham NH, Le TL, Grand P, Nguyen VN (2013) Computer aided plant identification system. In: International conference on computing, management and telecommunications (ComManTel), pp 134–139. <https://doi.org/10.1109/commantel.2013.6482379>
6. Raut SP, Bhalchandra AS (2018) Plant recognition system based on leaf image. In: Second international conference on intelligent computing and control systems (ICICCS), Madurai, India, pp 1579–1581. <https://doi.org/10.1109/iccons.2018.8663028>
7. Che Hussin N, Jamil N, Nordin S, Awang K (2013) Plant species identification by using scale invariant feature transform (sift) and grid based colour moment (gbcm). In: IEEE conference on open systems (ICOS), pp 226–230. 10.1109/ICOS.2013.6735079
8. Wu SG, Bao FS, Xu EY, Wang Y, Chang Y, Xiang Q (2007) A leaf recognition algorithm for plant classification using probabilistic neural network. In: IEEE international symposium on signal processing and information technology, Giza, pp 11–16. <https://doi.org/10.1109/isspit.2007.4458016>
9. Beghin T, Cope JS, Remagnino P, Barman S (2010) Shape and texture based plant leaf classification. In: Blanc-Talon J, Bone D, Philips W, Popescu D, Scheunders P (eds) Advanced concepts for intelligent vision systems. ACIVS 2010. Lecture Notes in Computer Science, vol 6475. Springer, Berlin, Heidelberg
10. Liu J, Yang S, Cheng Y, Song Z (2018) Plant leaf classification based on deep learning. In: Chinese automation congress (CAC), Xi'an, China, pp 3165–3169. <https://doi.org/10.1109/cac.2018.8623427>
11. Ying X (2019) An overview of overfitting and its solutions. J Phys Conf Ser 1168:022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
12. Almisreb AA, Jamil N, Din NM (2018) Utilizing AlexNet deep transfer learning for ear recognition. In: Fourth international conference on information retrieval and knowledge management (CAMP), Kota Kinabalu, pp 1–5. <https://doi.org/10.1109/infrkm.2018.8464769>
13. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1–9
14. Kumar N, Belhumeur PN, Biswas A, Jacobs DW, Kress WJ, Lopez IC, Soares JVB (2012) Leafsnap: a computer vision system for automatic plant species identification. In: Proceedings of the 12th European conference on computer vision (ECCV)
15. Object detection part-1 blog. <https://www.saagie.com/blog/object-detection-part1/>. Last accessed 30 July 2020



# Twitter Sentiment Analysis Using Supervised Machine Learning



Nikhil Yadav, Omkar Kudale, Aditi Rao, Srishti Gupta,  
and Ajitkumar Shitole

**Abstract** Sentiment analysis aims to extract opinions, attitudes, as well as emotions from social media sites such as twitter. It has become a popular research area. The primary focus of the conventional way of sentiment analysis is on textual data. Twitter is the most renowned microblogging online networking site in which user posts updates related to different topics in the form of tweets. In this paper, a labeled dataset publicly available on Kaggle is used, and a comprehensive arrangement of pre-processing steps that make the tweets increasingly manageable to normal language handling strategies is structured. Since each example in the dataset is a pair of tweets and sentiment. So, supervised machine learning is used. In addition, sentiment analysis models based on naive Bayes, logistic regression, and support vector machine are proposed. The main intention is to break down sentiments all the more adequately. In twitter sentiment analysis, tweets are classified into positive sentiment and negative sentiment. This can be done using machine learning classifiers. Such classifiers will support a business, political parties, as well as analysts, etc., and so evaluate sentiments about them. By using training, data machine learning techniques correctly classify the tweets. So, this method doesn't require a database of words, and in this manner, machine learning strategies are better and faster to perform sentiment analysis.

---

N. Yadav (✉) · O. Kudale · A. Rao · S. Gupta · A. Shitole  
International Institute of Information Technology, Pune, India  
e-mail: [nikhilyadav2698@gmail.com](mailto:nikhilyadav2698@gmail.com)

O. Kudale  
e-mail: [omkark.py@gmail.com](mailto:omkark.py@gmail.com)

A. Rao  
e-mail: [adirr0910@gmail.com](mailto:adirr0910@gmail.com)

S. Gupta  
e-mail: [sg685641@gmail.com](mailto:sg685641@gmail.com)

A. Shitole  
e-mail: [ajitkumarsh1@gmail.com](mailto:ajitkumarsh1@gmail.com)

**Keywords** Supervised machine learning · Sentiment analysis · Twitter · Data mining · Product evaluation · ROC · Classification · Naive Bayes · Logistic regression · Support vector machine · Linear SVC

## 1 Introduction

The Internet has been very useful in helping today's world express their perspectives globally. This is done through blog entries, on the web discussions forums, item audit sites, and so on. Individuals worldwide depend on this client, produced content extensively. For example, if someone wants to buy some product, then they first look up its reviews and comments before finalizing it [1]. But it is not humanly possible for a particular person to sit and look at every single review available. It would simply be a waste of time. Hence, to make this process easier, it can be automated. Machine Learning (ML) plays a significantly important part here. The process of Sentiment Analysis (SA) falling under ML helps the system understand the sentiment of a particular statement made. The system is built using several ML algorithms that can understand the nature of sentiment or a set of the same. In research, methods of ML have prevailed over knowledge- and dictionary-based methods to determine the polarity [2]. Polarity here is a semantic orientation that lies between two extremities, 0 and 1 or positive and negative. The paper proposes a system wherein data from twitter will be extracted on which SA will be performed. That data is saved in data frame. Then, some cleaning and pre-processing steps are performed on it so that, accurate information is utilized to fit the ML model which helps to predict labels for unknown cleaned and pre-processed data samples. Twitter is one of the popular sources containing a relatively huge amount of data. For performing sentiment analysis, certain supervised machine learning methods (algorithms) have been utilized to accomplish precise outcomes. Some of them are multinomial naive Bayes, linear support vector classifiers, and logistic regression classifiers. One is free to compose tweets in any form, without following any rules. This is what causes twitter more well known than other blogging locales. Due to this service, individuals tend to use abbreviations, make spelling mistakes, exaggerate reviews, use emoticons, etc., [3]. These formats usually make analysis a little difficult but still, there are methods such as feature extraction and mapping emoticons to their actual meanings that can be utilized to investigate the tweets. Movie and item audit easily accessible nowadays or thoughts on religious and political issues, so they become fundamental wellsprings of client slant and sentiment. This paper majorly focuses on data that are related to product reviews for product evaluation. It is restricted mainly to the data of vendors, manufacturers, entrepreneurs, and others of a similar domain. Messages can change from general opinion to individual idea [4].

## 2 Related Work

Sentiment analysis is the careful examination of how feelings and points of view can be identified with one's feeling and mentality appears in regular language regard to an occasion. The principle motivation behind choosing twitter's profile information is that subjective data can get from this platform [5]. Ongoing occasions show that sentiment analysis has reached incredible accomplishment which can outperform the positive versus negative and manage the entire field of behavior and feelings for various networks and themes. In the field of sentiment analysis utilizing various techniques, great measure of exploration has been done for the expectation of social sentiments. Pang and Lee (2002) proposed the framework, where an assessment can be positive or negative was discovered by the proportion of positive words to total words. Later in 2008, the creator built up a methodology in which tweet results can be chosen by term in the tweet [6]. Another study on twitter sentiment analysis was done by Go et al. [7] who stated the issue as a two-class classification, meaning to characterize tweets into positive and negative classes. M. Trupthi, S.Pabboju, and G.Narasimha proposed a system that mainly makes use of Hadoop. The data is extracted using SNS services which are done using twitter's streaming API. The tweets are loaded into Hadoop and are pre-processed using map-reduce functions. They have made use of uni-word naive Bayes classification [8]. The paper [9] analyzes the utilization of SA in business applications. Besides, this paper exhibits the text analysis process in auditing the popular assessment of clients toward a specific brand and presents hidden information that can be utilized for decision making after the text analysis is performed. In paper [10], the sentiment analysis has been done in four phases. Collecting real-time tweets up to a given limit, tokenizing every tweet as part of pre-processing, comparing them with an available bag of words, and classifying the tweets as positive or negative.

The proposed system is domain specific. A user interactive GUI will be available for the users to type in the keywords related only to the commercial products. Not many existing systems have been made so specific. Also, the system aims to compare various ML algorithms and choose the one which will produce results with the highest accuracy. Making the system domain specific reduces processing time as tweets regarding specific products are only searched based on the keywords typed.

## 3 Proposed System

The system intends to carry out sentiment analysis over tweets gathered from the twitter dataset. Various algorithms have been utilized and tested against the available dataset, and the most appropriate algorithm has been chosen. Figure 1 gives the idea about how the sentiment analysis will be carried out. Once the dataset has been cleaned and divided (isolated) into preparing (training) and testing datasets, it will be pre-processed using the techniques mentioned below. Features will be extracted

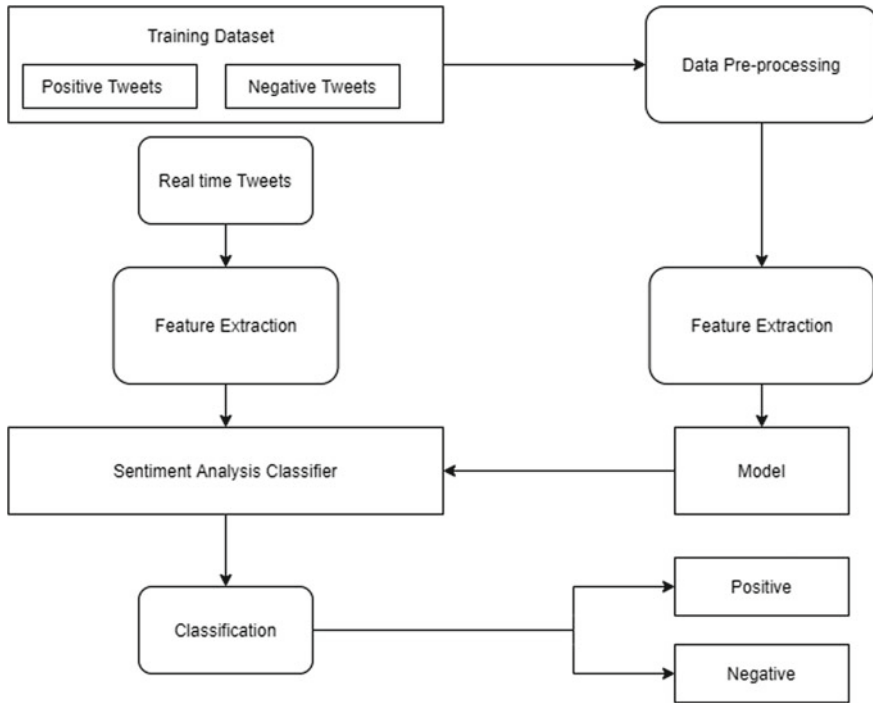


Fig. 1 Outline of proposed system

to reduce the dimension of the dataset. The next stage is to create a model that will be given to the classifier to classify the tweets into positive and negative tweets. Again real-time tweets will be given to the classifier for testing the real-time data. The proposed system does not engage in performing sentiment analysis on every tweet belonging to every other domain. The system is strictly domain restricted, where the sentiment analysis is performed to classify the tweets related to products in the market into a negative or positive category. The end-user will be provided with an interactive GUI wherein he/she can type the keywords or sentences related to a particular product. All tweets which are identified with that product will be available to the user. The user will be able to see the number of positive and negative statements made by others. This will help them in revising their production and work strategies accordingly which will be useful in improving their businesses.

Below are steps involved in handling large incoming data:

1. Data cleaning:
  - i. Use of various data tools that can help in cleaning the dataset.
  - ii. Use of several AI tools that help in identifying duplicates in large corpora of data and eliminate it.
  - iii. For correcting the corrupted data, the source of errors should be tracked and monitored constantly.

- iv. Validate once, when the existing data is cleaned.
2. Data pre-processing:
  - i. Assessing data quality.
  - ii. Identification of inconsistent values to know what the data type of the features should be.
  - iii. Aggregate the features to give better performance.

Sections 3.1 and 3.2 will give detailed ideas to handle large incoming data.

### 3.1 Data Cleaning

The data collected was not in the correct format. Information cleaning is the way toward guaranteeing that information is right, predictable, and usable. Usually, datasets need to cleanse because they consist of a lot of noisy or unwanted data called outliers too. The existence of such outliers may lead to inappropriate results. Data cleaning ensures the removal and improvisation of such data and results in a much more reliable and stable dataset.

Data cleaning can be done in given ways:

- **Monitors errors.** The entry point or source of errors should be tracked and monitored constantly. This will help in correcting the corrupted data.
- **Process standardization.** The point of entry should be standardized. By standardizing the data process, the risk of duplication reduces.
- **Accuracy validation.** Data should be validated once the existing database is cleaned. Studying and using various data tools that can help in cleaning the datasets is very important.
- **Avoid data duplication.** The identification of duplicate data is a very mandatory process. Several AI tools help in identifying duplicates in large corpora of data.

The above-mentioned steps are a few of the many ways to clean datasets. Making use of these methods will end up giving good, usable, and reliable datasets.

### 3.2 Data Pre-processing

The next step after data cleaning is data pre-processing. It is a major step in machine learning. It is the process in which data gets transformed or encoded to a state which is understandable to the machine. In simple words, the features of the dataset can be easily interpreted by the algorithms. The feature is a measurable property of an entity being observed. For example, height, age, gender can be considered as features of a person. A twitter stream will extract every related tweet from twitter, which will be in an unstructured structure. These unstructured tweets need to experience pre-handling before applying any classifier over it. The tweets will be pre-handled with

tokenization and cleaning. Initially, all the HTML contents are expelled from the tweets by giving a URL structure. Further, cleaning happens by expelling non-letters or images using python regular expressions. All the tweets must be in the same case to processes; thus, it will change over to bring the lower case, and each word is split based on space. Following this, gather all stop words and structure it as a solitary set and evacuate it. At last, return a string of important words [11]. Thus, a pre-processing step is performed for filtering out the slang words and misspellings before extracting the features[12]. Following steps can help with data pre-processing:

- **Data quality assessment.** Since the data is collected from multiple sources, it will be unrealistic to consider it to be perfect. Assessing the data quality must be the first step while pre-processing it.
- **Inconsistent values.** Data can be inconsistent at times. Like the "address" field can contain a phone number. Hence, the assessment should be done properly like to know what the data type of the features should be.
- **Feature aggregation.** As the name says, features are aggregated to give better performance. The behavior of aggregated features is much better when compared to individual data entities.
- **Feature sampling.** It is a way of selecting a subset of the original (first) dataset. The central matter of sampling is that the subset should have nearly the same properties as the original dataset.

Coming to the proposed system, the pre-processing done will be as follows:

- Converting tweets to lowercases.
- Supplant at least two dots with spaces.
- Replace extra spaces with a single one.
- Get rid of spaces and quotes at the end of the tweets.

Two types of features are extracted from the dataset, namely unigrams and bigrams. A frequency distribution is created for the extracted features. Later, the top N unigrams and bigrams are chosen to carry out the analysis. Also, tweets contain special features like URLs, user names, emoticons, etc. Retweets are also a feature of tweets. These features are not required while performing sentiment analysis. Hence, these features are replaced with common keywords or markers like "URL," "USER\_MENTION," "EMO," respectively. Again, removal of stop words and lemmatization are necessary steps to be done.

**Stop words.** Stop words will be words that don't have any criticalness in search inquiries. For example, "I like to write." After removing stop words becomes, "like write." "I" and "to" are termed as stop words.

**Stemming.** It is an element procedure of delivering morphological variations of a base word. The words like "chocolatey," "chocolates" are converted to their root word "chocolate."

**Lemmatization.** Lemmatization decreases the inflected words appropriately guaranteeing that the root word has a place with the language.

### 3.3 Classifiers to Be Used

**Naive Bayes.** The naive Bayes is a supervised machine learning algorithm that returns probability values as the output. Naive Bayes classifier is very useful in solving high-dimensional problems. It assumes the probabilities of the different events that are completely independent. Naive Bayes [13] is a straightforward model, where class  $C$  is assigned a tweet  $t$  such that:

$$C = \arg \max P(c|t) \tag{1}$$

$$P(c|t) \propto P(c) \prod^n P(f_i|c) \tag{2}$$

The probability of event A happening can be found, by giving the occurrence of event B. Naive Bayes algorithm can be used to tackle large scale classification problems.

**Logistic regression.** Logistic regression predicts a binary outcome, i.e., (Y/N) or (1/0) or (True/False). It also works as a special case of linear regression. It produces an S-shaped curve better known as a sigmoid. It takes real values between 0 and 1.

The model of logistic regression is given by:

$$\text{Output: } 0 \text{ or } 1 \tag{3}$$

$$\text{Hypothesis: } Z = WX + B \tag{4}$$

$$h_\theta(x) = \text{sigmoid}(Z) \tag{5}$$

Basically, logistic regression has a binary target variable. There can be categories of target variables that can be predicted by it. The logistic classifier uses a cross-validation estimator.

**Support vector machine.** It is a non-probabilistic model that utilizes a portrayal of text models as focuses in a multidimensional space. These examples are mapped with the goal that the instances of the diverse categories (sentiments) have a place with particular areas of that space. Later, the new messages are mapped onto that equivalent space and are predicted to have a place with a classification dependent on which category they fall into. In the SVM algorithm, the fundamental goal is to boost the edge between information points and the hyperplane. The loss function that helps with this is called a hinge loss. The equation of the hyperplane is given as:

$$w \cdot x - b = 0 \tag{6}$$

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases} \tag{7}$$

The cost is 0 if the predicted, and the actual value is of a similar sign. On the off chance that they are not, at that point, figure the loss esteem.

For performing sentiment analysis, logistic regression is considered over naive Bayes because naive Bayes assumes all the features used in model building to be conditionally independent whereas logistic regression splits feature space linearly and typically works reasonably well even when some of the variables are correlated, and on the other hand, logistic regression and SVM with a linear kernel have similar performance but depending on the features, one may be more efficient than the other.

### 3.4 Plotting Results

The plotting of the result will be done using graphs, and the comparison of algorithms is done using a ROC curve [14]. It is a plot of true positive rate and false positive rate. Figure 2 shows the ROC curve for a MultinomialNB Model. The Area Under the Receiver Operating Characteristics curve (AUROC) for the MultinomialNB Model is 0.89.

Figure 3 shows the ROC curve for a logistic regression model. The Area Under the Receiver Operating Characteristics curve (AUROC) curve for the logistic regression model is 0.90.

Figure 4 shows the ROC curve for a linear SVC model. The Area Under the Receiver Operating Characteristics curve (AUROC) for the linear SVC model is 0.83.

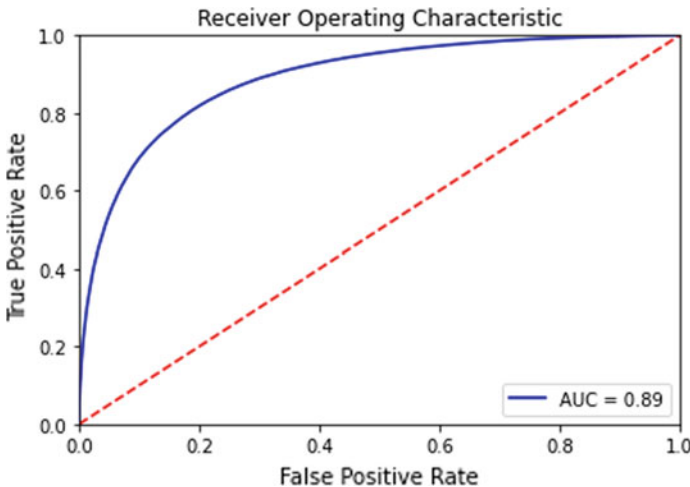


Fig. 2 ROC curve for multinomial naive Bayes classifier



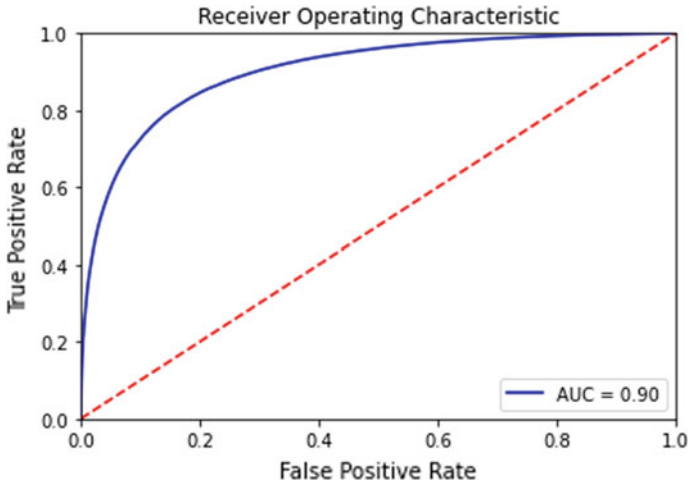


Fig. 3 ROC curve for logistic regression classifier

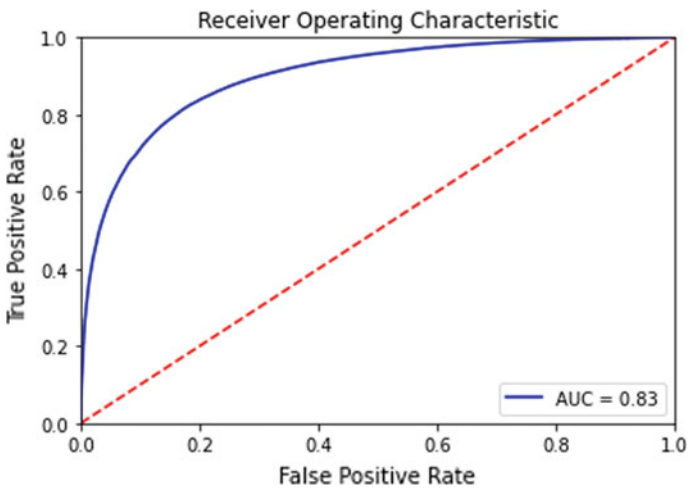


Fig. 4 ROC curve for linear SVC

## 4 Results

The dataset used for the training model is the Sentiment140 dataset. It is a balanced dataset with 1.6 million tweets among which 8 lakh tweets belong to the positive class, and the remaining 8 lakh tweets belong to negative class. The splitting is done using the `train_test_split` method with a `test_size` of 0.20. 12 lakh tweets are used for training the model, and the remaining 4 lakh tweets are used for testing the model.

**Table 1** Classification report of logistic regression

	Precision	Recall	F1 Score	Support
0 (Negative label)	0.81745	0.83662	0.82692	160,156
1 (Positive label)	0.83236	0.81280	0.82246	159,844
Accuracy			0.82472	320,000
Macro avg	0.82490	0.82471	0.82469	320,000
Weighted avg	0.82490	0.82472	0.82470	320,000

**Table 2** Classification report of multinomial naive Bayes

	Precision	Recall	F1 Score	Support
0 (Negative label)	0.78090	0.85148	0.81466	160,156
1 (Positive label)	0.83637	0.76064	0.79671	159,844
Accuracy			0.80610	320,000
Macro avg	0.80864	0.80606	0.80569	320,000
Weighted avg	0.80861	0.80910	0.80569	320,000

#### 4.1 Logistic Regression

Table 1 gives the classification report of the logistic regression model. The accuracy of the model is 82.47. It also shows the precision and recall of the model. Precision is the positive predictive value, and recall is the sensitivity of the model [15].

#### 4.2 Multinomial Naive Bayes

Table 2 gives the classification report of multinomial naive Bayes model. The accuracy of the model is 80.61.

#### 4.3 Linear Support Vector Machine

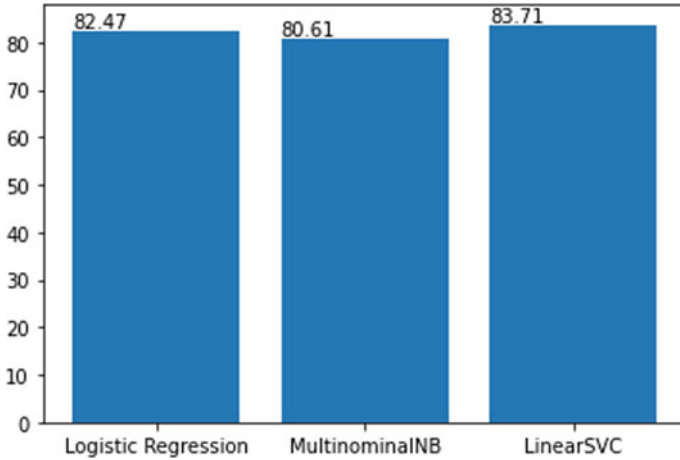
Table 3 shows the classification report of the linear SVC model. The accuracy of the model is 83.71.

ROC curves are appropriate when the observations are balanced between each class and since the dataset used for training and testing is a balanced dataset ROC curves which are considered for measuring the performance of the model.

AUROC is a superior measure of classifier performance than accuracy because it does not bias on size of test or evaluation data whereas accuracy is always biased on size of test data, and also AUROC is the best summary of the performance of a

**Table 3** Classification report of linear support vector machine

	Precision	Recall	F1 Score	Support
0 (Negative label)	0.85970	0.90315	0.88889	160,011
1 (Positive label)	0.78447	0.70515	0.74270	79,989
Accuracy			0.83716	240,000
Macro avg	0.82288	0.88415	0.81179	240,000
Weighted avg	0.83462	0.83716	0.83483	240,000



**Fig. 5** Comparison of accuracies of classifiers

classifier as it incorporates different aspects of the performance into a single number. Figure 5 shows a comparison between the three algorithms used for sentiment analysis; comparatively, linear SVC gives the highest accuracy of 83.71 but its AUROC is less than logistic regression having an accuracy of 82.47, and hence, logistic regression is considered for classification purpose.

## 5 Conclusion and Future Work

The work in this paper is done to classify a relatively huge corpus of twitter data into two groups of sentiments, positive and negative, respectively. Higher accuracy is achieved by using sentiment features instead of conventional text classification. This feature can be used by various establishments, business organizations, entrepreneurship, etc., to evaluate their products and get a deeper insight into what people say about their products and services. Future work includes working not only in the English language but in other regional languages too. Also, it will include analysis

of complex emotions like sarcasm and generate a hybrid classifier to get the best accuracy.

## References

1. Neethu MS, Rajasree R (2013) Sentiment analysis in twitter using machine learning techniques. In: 2013 Fourth international conference on computing, communications and networking technologies (ICCCNT), Tiruchengode, pp 1–5
2. Kumar A, Sebastian TM, Sentiment analysis on twitter. Department of Computer Engineering, Delhi Technological University Delhi, India
3. Joshi R, Tekchandani R (2016) Comparative analysis of Twitter data using supervised classifiers. In: 2016 International conference on inventive computation technologies (ICICT). ISBN: 978-1-5090-1285-5
4. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R, Sentiment analysis of twitter. Passonneau Department of Computer Science Columbia University New York, NY 10027 USA
5. Hasan A, Moin S, Karim A, Shamshirband S, Machine learning-based sentiment analysis for twitter accounts. Department of Computer Science, Air University, Multan Campus, Multan 60000
6. Pang B, Lee L (2008) Opinion mining and sentiment analysis
7. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1, no 12
8. Trupthi M, Pabboju S, Narasimha G (2017) Sentiment analysis on twitter using streaming API. In: 2017 IEEE 7th international advance computing conference (IACC), Hyderabad, pp 915–919
9. Halibas AS, Shaffi AS, Mohamed MAKV (2018) Application of text classification and clustering of twitter data for business analytics. In: 2018 Majan international conference (MIC), ISBN: 978-1-5386-3761-6
10. Neethu MS, Rajasree R (2013) Sentiment analysis in twitter using machine learning techniques. In: 2013 Fourth international conference on computing, communications and networking technologies (ICCCNT), Tiruchengode, pp 1–5
11. Shamantha RB, Shetty SM, Rai P (2019) Sentiment analysis using machine learning classifiers: evaluation of performance. In: 2019 IEEE 4th international conference on computer and communication systems (ICCCS), Singapore, pp 21–25
12. Tyagi P, Tripathi RC (2019) A review towards the sentiment analysis techniques for the analysis of twitter data. In: Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE)
13. Yadav N, Kudale O, Gupta S, Rao A, Shitole A (2020) Twitter sentiment analysis using machine learning for product evaluation. In: 5th International conference on inventive computation technologies (ICICT-2020)
14. Shitole A, Devare M (2019) TPR, PPV and ROC based performance measurement and optimization of human face recognition of IoT enabled physical location monitoring. *Int J Recent Technol Eng* 8(2):3582–3590. ISSN: 2277-3878
15. Shitole A, Devare M (2018) Optimization of person prediction using sensor data analysis of IoT enabled physical location monitoring. *J Adv Res Dyn Control Syst* 10(9):2800–2812. ISSN: 1943-023X

# Classification of Skin Disease Using Traditional Machine Learning and Deep Learning Approach: A Review



Honey Janoria, Jasmine Minj, and Pooja Patre

**Abstract** It is generally known that the skin infections are found in all the living organisms. Skin ailment is a specific sort of ailment introduced by either microorganisms or a disease. Out of the three essential kinds of skin malignant growth, namely Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), and Melanoma, the Melanoma is observed as one of the most hazardous in which endurance rate is extremely low. The early location of Melanoma can conceivably improve the endurance rates. Innovations in the skin disease recognition are extensively partitioned into four essential segments, viz., picture preprocessing that incorporates hair evacuation, de-clamor, honing, resize of the given skin picture followed by division. In this paper, a survey is carried out on the best in class in a PC helped analysis framework and further observes the ongoing practices in various strides of these frameworks. Measurements and results from the most significant and ongoing executions are broke down and announced. This research work has analyzed the presentation of late work that remains dependent on various boundaries like precision, dataset, computational time, shading space, AI procedure, and so on further the investigation carried out in this paper will help the scientists and researchers of the significant field.

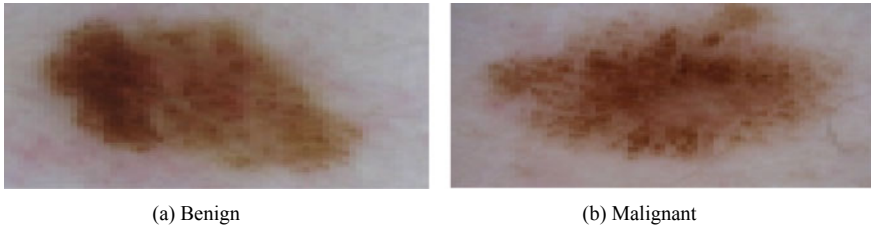
**Keywords** Skin disease · CNN model · Segmentation

---

H. Janoria (✉) · J. Minj · P. Patre  
Department of Computer Science, Vishwavidyalaya Engineering College Lakhanpur,  
Chhattisgarh Swami Vivekananda Technical University, Chhattisgarh, India  
e-mail: [honeyjanoria1992@gmail.com](mailto:honeyjanoria1992@gmail.com)

J. Minj  
e-mail: [jasmine5feb@gmail.com](mailto:jasmine5feb@gmail.com)

P. Patre  
e-mail: [poojapatre.89@gmail.com](mailto:poojapatre.89@gmail.com)



**Fig. 1** Major types of skin cancer

## 1 Introduction

One of the most common types of malignant growth is skin cancer. It is found in about 50% of the all-out proportion of malignancy patients. Skin cancer is typically found and classified in different kinds, for example, Melanoma, Basal, Squamous, Benign, and so forth. [1, 2] From all the malignancies found and recognized as of not long ago, the sort melanoma is probably the most unnerving kind of disease, as this disease reports around 7300 deaths for every year in the USA [3, 4]. There are two kinds of pictures accessible for skin malignant growth locations. The skin picture is caught by a specifically devoted framework in the neurotic community with an emphasis on the locale of enthusiasm with high zoom (e.g., 20 $\times$ ), which needs a skilled dermatologist to finish up the picture as positive or negative [5]. This kind of picture can be feed to an electronic semi-computerized framework for grouping. Be that as it may, in this innovation, the casualty in every case needs to stroll into the obsessive community and needs to take consultancy of the talented dermatologist [6]. Then again, if there is PC programming that can consequently recognize skin malignant growth from an advanced picture caught by any computerized picture that catches framework with a little spotlight on the area of intrigue, the casualty can play out the test at any place even at home [7]. Skin disease is classified into two categories benign and malignant as shown in Fig. 1 [8, 9]. Benign is the starting stage of cancer whereas malignant is considered a higher level of cancer.

## 2 Literature Survey

Numerous authors utilize diverse machine learning procedures for the characterization of skin infections. Some of the most widely used machine learning techniques like support vector machine, artificial neural network, decision tree are used by many researchers earlier, but nowadays, deep learning-based approach [2] becomes more popular for classification of such disease as it gives a better classification, accuracy, and very suitable for image input. This extensive literature work has reviewed the traditional methods used for classification of skin diseases and then deep learning-based approaches are used for deploying such classification. Furthermore, some of

the authors use transfer learning and ensemble learning to perform the desired operations. Approaches like support vector machine [10] and artificial neural network [10, 11] were very common. Whereas the deep learning approach, convolutional neural network-based model [1, 5, 8, 9], AlexNet model [3, 12], ResNet 50/ResNet 101 [4, 13, 14], and dense CNN model [7, 12, 13, 15] are commonly used by many of the researchers; some researchers have also proposed the use of transfer learning, where they extracted the feature from deep learning model and apply that feature on some other model. Bashar A. presented a survey on various neural networks and deep learning techniques with its applications in different areas like speech recognition and image classification [16]. Raj S. J. discussed applications of machine learning methods with its limitation and concluded that the use of deep learning and optimization techniques may increase the accuracy of classification [17]. All the approaches are tabulated below in Table 1.

## **2.1 Available Datasets**

There are many online available datasets like ISIC, PH2, EDRA datasets, which are used by many authors in their work, in Table 2 some of the most widely used datasets are tabulated.

## **3 Methods Used for Classification**

### **3.1 Traditional Machine Learning Approaches**

A portion of the traditional machine learning methods which are utilized for grouping of skin disease are support vector machine [26], K-nearest neighbor, and the decision tree. The traditional approach comprises basic five steps: data acquisition, preprocessing, segmentation, feature extraction, and then classification, where data acquisition can be done by taking any one of the global datasets available online, then preprocessing steps contain the noise removal and anomaly removal present in the input image; for this purpose, they applied mini noise removal and blur removal filter on the input image in skin image also there is a process of small hair on skin segmentation and region of interest extraction will be a very important step to remove such unwanted regions from the image; for this purpose, many researchers use the morphological and fuzzy-based system. For feature extraction, many of the others use GLCM feature [11], some of the authors uses hybrid features like feature texture feature and morphological feature of input image; finally, all these features are divided into training and testing datasets to put into a machine learning-based classifier for the classification purpose SVM and decision tree and many other traditional classifiers have been applied by many researchers.

**Table 1** Comparison of existing methods

Author/journal detail	Models applied	Dataset used	Results (highest accuracy or AUC)	Findings
Haenssle et al. 2018 [1]	CNN (Google inception)	Self-created dataset	0.82(specificity) 0.86 (AUC)	Problem of imbalance of validation dataset
Walker et al. 2019 [2]	Deep learning	ISIC 2017 dataset	AUC = 0.99	Mostly trained with melanoma type cancer only
Brinker et al. 2018 [3]	CNN (AlexNet)	NA	94% (Accuracy)	Small size of the dataset
Maron et al. 2019 [4]	CNN, ResNet50 Model	ISIC 2018	Sensitivity = 74.4% Specificity = 91.3%	Tested on just 100 images from an external dataset
Schandl et al. [5]	CNN	Local dataset of 13,724 images (7895 normal and 5829 diseased)	0.742 (AUC) 0.51 2 (specificity) 0.80.5(sensitivity)	–
Xue et al [6]	ResNet-101 classifier, DNN, OUSM	ISIC 2017 challenge dataset, ISIC archive	Accuracy = 84.5%	Accuracy may be increased using cascade models
He et al [7]	Dense deconvolutional network	ISBI2016, ISBI 2017	Accuracy ISBI 2016 = 96.0% ISBI 2017 = 93.9%	–
Burdick et al. [8]	CNN, transfer learning	The ISIC 2016	Accuracy = 75% AUC = 0.693	–
Brinker [9]	CNN, deep learning	ISIC image archive	Sensitivity = 0.741	Focuses on melanoma detection only
Zhang et al. [18]	GoogleNet, InceptionV3 models	Local dataset	Accuracy = 87.25	Very specific deep learning model used. Not included the result on public datasets
Hekler et al. [19]	ResNet50 model	Local dataset of 595 images	Accuracy = 82%	Very small dataset used. Accuracy is also less than other deep learning models

(continued)



**Table 1** (continued)

Author/journal detail	Models applied	Dataset used	Results (highest accuracy or AUC)	Findings
Tschandl et al. [20]	CNN, UNet16, LinkNet34	HAM10000 dataset	JI = 76.3%	Mobile net model proposed
Mahbod et al. [21]	CNN, ensemble-based CNN, inter and intra-architecture network fusion	ISIC 2016 and ISIC 2017	ISIC 2016 = 0.87 (AUC) ISIC 2017 = 0.95 (AUC)	Pre-trained models are used
Marchetti et al. [22]	Ensemble learning an SVM	ICVMC dataset of 379 images	82% (Sensitivity) 76% (Specificity) 0.86 (AUC)	Very small dataset used
Chaturvedi et al. [23]	CNN, MobileNet	HAM10000 Dataset	Precision=0.89 Recall =0.83 F1-Score=0.83	Chances of improvement in accuracy
Hosny et al. [12]	DCNN, Alex-net, transfer learning	MED-NODE, Derm (IS & Quest) and ISIC dataset2017	Accuracy MED-NODE = 96.86% Derm (IS & Quest) = 97.70% ISIC 2017 = 95.91%	–
Bisla et al. [15]	U-Net, DCGANs, ResNet-50	ISIC 2017, 2018, PH2 dataset, Edinburgh dataset	AUC = 0.95 (melanoma)	Data purification techniques applied to increase accuracy
Zhang et al. [24]	Attention residual learning convolutional neural network (ARL-CNN) model	ISIC-2017 skin lesion dataset.	Accuracy = 87.5% (melanoma) and 95.8% (Seborrheic Keratosis)	Not applied on ISIC 2018 dataset
Shi et al. [25]	CNN, ResNet-101	ISIC 2017 skin lesion classification challenge dataset	Accuracy =90.8%	Model may be applied on latest dataset
Khan et al. [13]	DCNN, kurtosis-controlled principle component (KcPCA), RESNET-50 and RESNET-10	ISIC, ISBI 2017, and ISBI 2016 dataset	Accuracy PH2= 97.9, ISBI2016 = 99.1, ISIC (MSK-1 and MSK-2)= 98.4 ISIC UDA= 93.8	–

(continued)

**Table 1** (continued)

Author/journal detail	Models applied	Dataset used	Results (highest accuracy or AUC)	Findings
Han et al. [14]	CNN, ResNet-152 model	Asan dataset Edinburgh datasets	Asan test dataset 0.91 (AUC) Edinburgh dataset 0.89 (AUC)	Tested on limited datasets
Abuzaghle et al. [10]	SVM	PH2	Average accuracy = 90.4	Deep learning models are not tested.
Taufiq [11]	SVM, ANN	NA	Sensitivity: 80% Specificity: 75%	Less accuracy than the deep learning models
Wadhawan et al. [26]	SVM	ISIC 2017	Sensitivity: 80.76% Specificity: 85.57%	Only SVM classifier applied
Do et al. [27]	SVM	1300 images	Accuracy: 92.09%	Only SVM classifier applied, deep learning architecture may apply

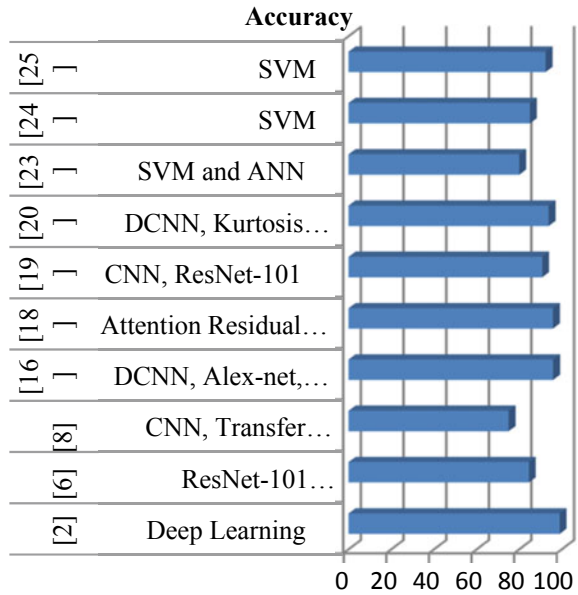
**Table 2** Description of online available global datasets

Dataset	Number of images	Number of classes
ISIC 2016	900	2
ISIC 2016	2000	3
PH2	200	3
EDRA	200	2

### 3.2 Deep Learning-Based Approaches

The deep learning approach is concerned very useful approach for the classification and detection of disease from images by using convolutional neural network-based model [1] and there were some of the authors who also use convolutional neural network model and transfer deep learning method [12] does not require any feature extraction because here features are already extracted by deep learning models. This feature contains local, global, and middle-level features so it will be very helpful for classification because it contains all the possible features. This model uses a pre-trained model for training is not required in deep learning approach user need to just provide the input image for the testing purpose; first, the model extended the local and Middle-level features from the input image and then education based on a pre-trained model what they already have in memory [2].

**Fig. 2** Comparison of accuracy for traditional machine learning and deep learning approach on ISIC datasets



Furthermore, some hybrid deep learning model [13] and transfer learning-based model [12] are added with some more flavor in the classification and identification tasks. This method uses more than one deep learning models [11] in two or more stages. So the classification accuracy of such a model is expected to be very high. On the other hand, the classification time will also be high for such a model. While using the transfer learning model, at the first stage, it generally use deep learning models, and at the second stage, other machine learning approaches are used for classification purposes; in such case, the time can be saved as compared to previous approaches. Another way to save classification time and increase accuracy was the use of ensemble learning where different iterations of training are performed to increase the accuracy.

The performance comparison of the traditional machine learning approach and the latest deep learning approaches is shown in Fig. 2. Here, the average performance of a deep learning-based approach is remarkably higher than the traditional approaches.

## 4 Conclusion

After reviewing lots of paper, it is easily concluded that the traditional machine learning-based approach has many steps like preprocessing, segmentation feature extraction, and classification process for classification of such disease and also from the literature survey, it is observed that the accuracy of traditional machine learning approach was near about 80–90%, whereas the use of deep learning-based approach

not only saves a lot of time but also it gives accuracy in the range of 90–99%. So, with this survey, it is finally concluded that if the available dataset is large in amount, it avoids the use of traditional machine learning-based approach because it will be time consuming for the large datasets. Whereas the incorporation of deep learning method will also be very useful.

## References

1. Haenssle HA et al (2018) Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 29(8):1836–1842
2. Walker BN et al (2019) Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: laboratory and prospective observational studies. *EBioMedicine* 40:176–183
3. Brinker TJ et al (2018) Skin cancer classification using convolutional neural networks: a systematic review. *J Med Internet Res* 20(10):e11936
4. Maron RC et al (2019) Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer* 119:57–65
5. Tschandl P et al (2019) Expert-level diagnosis of non-pigmented skin cancer by combined convolutional neural networks. *JAMA Dermatol* 155(1):58–65
6. Xue C et al (2019) Robust learning at noisy labeled medical images: applied to skin lesion classification. arXiv preprint [arXiv:1901.07759](https://arxiv.org/abs/1901.07759)
7. He X et al (2018) Dense deconvolution net: multipath fusion and dense deconvolution for high-resolution skin lesion segmentation. *Technol Health Care* 26(S1):307–316
8. Burdick J et al (2018) Rethinking skin lesion segmentation in a convolutional classifier. *J Digital Imaging* 31(4):435–440
9. Brinker TJ et al (2019) Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer* 113:47–54
10. Abuzaghleh O, Faezipour M, Barkana BD (2015) A comparison of feature sets for an automated skin lesion analysis system for melanoma early detection and prevention. In: 2015 IEEE long Island systems, applications, and technology conference
11. Taufiq MA, Hameed N, Anjum A, Hameed F (2016) m-skin doctor: a mobile-enabled system for early melanoma skin cancer detection using support vector machine. In: EAI international conference on mobile medical multimedia technologies, applications, and services. Hungary
12. Hosny KM, Kassem MA, Foad MM (2019) Classification of skin lesions using transfer learning and augmentation with Alex-net. *PLoS ONE* 14(5):e0217293
13. Khan MA et al (2019) Multi-model deep neural network-based features extraction and optimal selection approach for skin lesion classification. In: 2019 international conference on computer and information sciences (ICCIS). IEEE
14. Han SS et al (2018) Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol* 138(7):1529–1538
15. Bisla D et al (2019) Towards automated melanoma detection with deep learning: data purification and augmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops
16. Bashar Abul (2019) Survey on evolving deep learning neural network architectures. *J Artif Intell* 12:73–82
17. Raj JS (2019) A comprehensive survey on the computational intelligence techniques and its applications. *J ISMAC* 1(3):147–159
18. Zhang X, Wang S, Lui J, Tao C (2018) Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. *BMC Med Inf Decis Making* 18(2):59

19. Hekler A et al (2019) Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur J Cancer* 115(2019):79–83
20. Philipp T, Sinz C, Kittler H (2019) Domain-specific classification pre-trained fully convolutional network encoders for skin lesion segmentation. *Comput Biol Med* 104:111–116
21. Mahbod A et al (2019) Skin lesion classification using hybrid deep neural networks. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech, and signal processing (ICASSP). IEEE
22. Marchetti MA et al (2018) Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *J Am Acad Dermatol* 78(2): 270–277
23. Chaturvedi SS, Kajol G, Prasad P (2019) Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet. arXiv preprint [arXiv:1907.03220](https://arxiv.org/abs/1907.03220)
24. Zhang J et al (2019) Attention residual learning for skin lesion classification. *IEEE Trans Med Imaging*
25. Shi X et al (2019) An active learning approach for reducing annotation cost in skin lesion analysis. arXiv preprint [arXiv:1909.02344](https://arxiv.org/abs/1909.02344)
26. Wadhawan T, Situ N, Rui H, Yuan X, Zouridakis G, SkinScan (2011) A portable library for melanoma detection on handheld devices. In: 2011 IEEE international symposium on biomedical imaging: from nano to macro, pp 133–136
27. Do TT, Zhou Y, Zheng H, Cheung NM, Koh D (2014) Early Melanoma diagnosis with mobile imaging. In: 36th IEEE annual, international conference of the engineering in medicine and biology society (EMBC), pp 6752–6757

# Spam SMS Filtering Using Support Vector Machines



P. Prasanna Bharathi, G. Pavani, K. Krishna Varshitha,  
and Vaddi Radhesyam

**Abstract** In recent years, SMS spam messages are increasing exponentially due to the increase in mobile phone users. Also, there is a yearly increment in the volume of mobile phone spam. Filtering the spam message has become a key aspect. On the other side, machine learning has become an attractive research area and shown the capacity in data analysis. So, in this paper, two popular algorithms named Naive Bayes and support vector machine are applied to SMS data. The SMS dataset is considered from Kaggle resource. The detailed result analysis is presented. Accuracy of 96.19% and 98.79% is noticed for the chosen algorithms, respectively, for spam SMS detection.

**Keywords** SMS · Spam · Ham · Machine learning · Classification · Naïve Bayes · SVM

## 1 Introduction

In recent years, due to more increasing number of mobile phone users, SMS is considered to be useful and trusted service. But there is an exponential increase in SMS spams. It is becoming increasingly difficult to handle the problem of maintaining spam SMS. The various types of mobile messaging attacks seen in networks today are described below:

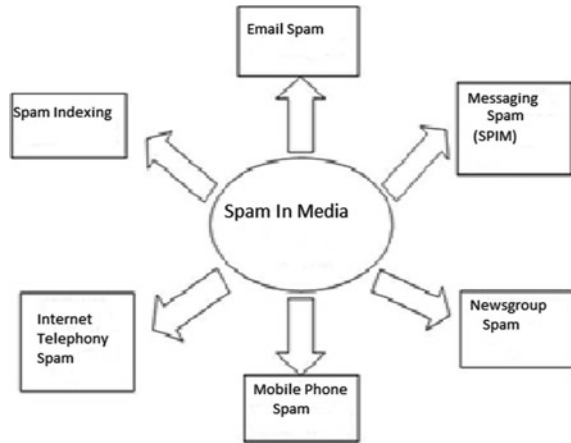
**SMS Spam:** The text messages are delivered to mobile phones in the form of SMS. Generally, this type of attack happens as a part of business promoting measures.

---

P. Prasanna Bharathi · G. Pavani · K. Krishna Varshitha  
IV/IV B.Tech Information Technology, V R Siddhartha Engineering College, Vijayawada, India

V. Radhesyam (✉)  
Assistant Professor, Information Technology, V R Siddhartha Engineering College, Vijayawada,  
India  
e-mail: [syam.radhe@gmail.com](mailto:syam.radhe@gmail.com)

**Fig. 1** Types of spam



**Premium-rate fraud:** Sometimes, for selected customers, uninvited messages will be sent to call premium-rate numbers or register for the subscription services. The following is the example of premium-rate message.

“CONGRATULATIONS! YOUR CELL NO. HAS WON 500.000 lb IN THE ONGOING SONY ERICSSON MOBILE PROMO. FOR CLAIM CALL + 447045754969”.

On the other hand, malware attacks also happen. Three of the most common forms of malware include:

**Virus:** It is computers program that which imitates itself. It can only infect other computer programs.

**Worm:** It is a malicious computer program. It imitates itself and makes disturbance over the computer network.

**Trojan:** A computer program or malicious code that damages the data. Different types of spams are shown in Fig. 1.

**E-mail Spam:** The most widespread and familiar form of spam is e-mail spam. The goal of this is the users through direct mails. A lot of memory and bandwidth loss happens due to e-mail spam.

**Instant Messaging Spam:** Yahoo! Messenger, Windows Live Messenger chat rooms are general examples for IM systems. These become the main channels for attackers or spammers.

**Newsgroup Spam:** Newsgroup spam refers to the posting of some random advertisement to the newsgroups. Spammers target those users who read news from these newsgroups. The advertisements are thus posted to many newsgroups at a time. A barrage of advertising and other irrelevant posts overwhelm the users and they robbed of the utility of the newsgroups through Newsgroup Spam.

**Mobile Phone Spam:** Mobile phone spam occurs as a part of business development activities.

**Internet Telephony Spam:** The pre-recorded, needless, unwanted bulk of phone calls that are robotically administered.

**Spamdexing:** Spamdexing consists of two words “Spam” and “Indexing”. “Spam” means flooding the Internet with many copies of a single message and Indexing means the systemic arrangement of data. So search engine spamming can be done by this practice of spam indexing.

## 2 Related Work

Different Bayesian-related classifiers are tested by Gomez Hidalgo et al. [5]. Two popular SMS spam datasets, namely the Spanish and English test were introduced in this work. Moderate results are reported after the test with a large number of messages.

Spam content in the summary of e-mail and comments on blogs is tested by Cormack et al. [2]. This work is done through the use of content-based spam filtering. Role of features as words of messages is explored in this work.

Nuruzzaman et al. [3] focused on how to reduce time and space complexities for machine learning algorithms which are used to classify spam SMS messages. In particular, text classification approaches are tested. Good accuracy is reported.

Bloom filters [4] are introduced by Coskun and Giura et al. SMS spam messages are identified using the online detection technique in a single network system.

Delany et al. [6] adopted a k-way spectral clustering approach to detect spam. 1353 known spam messages without duplicates are compiled and used as dataset for this work.

According to the analysis work done by Almeida et al. [1], SVM is a better approach for SMS text classification.

## 3 Proposed Method

The overall workflow of the proposed method is shown in Fig. 2. Dataset collection phase includes the collection of spam and ham messages. In the current work, the Kaggle dataset in comma separated values (CSV) format is used.

At the second level of the experiment, preprocessing is done for a better quality input either by removal of unwanted data. Then, the pre-processed data is transformed into a machine-readable form or non-contextual form by converting to vector or by doing discretization. Training and testing are performed to design the model. This model is validated on spam messages. To assess the performance of the algorithm, confusion matrix has been constructed. Tables 1 and 3 represent the confusion matrices for Naive Bayes and support vector machine algorithms, respectively.

### Support Vector Machines

Support vector machines (SVM) are also called support vector networks (SVN) are supervised machine learning methods [7], generally used for classification and



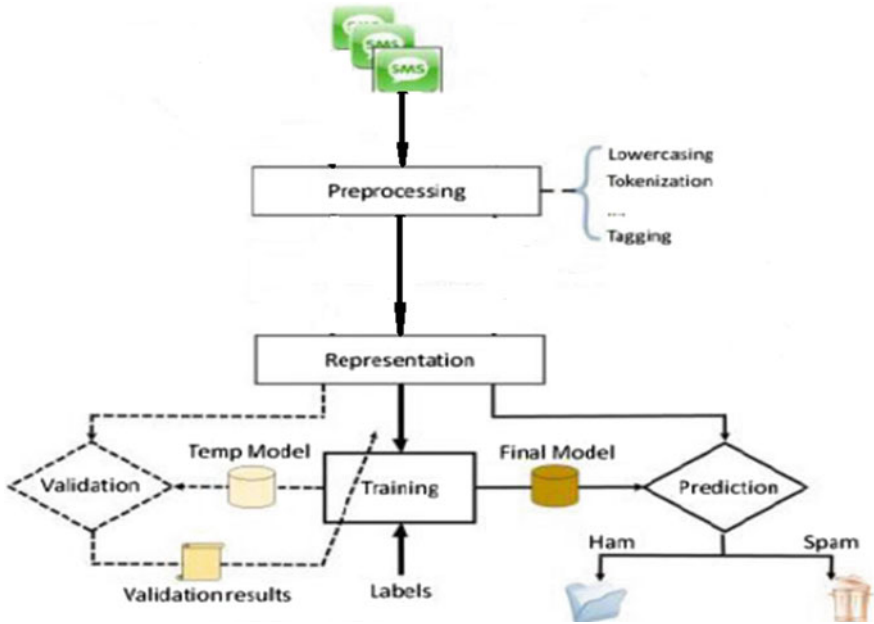


Fig. 2 Design methodology

Table 1 Confusion matrix

Messages = 5574	Predicted: ham	Predicted: spam
Actual: ham	True positive = 4820	False negative = 7
Actual: spam	False positive = 738	True negative = 9

regression purposes. SVM is a non-probabilistic binary linear classifier that assigns training data into one category or more. It also can be used for nonlinear classification problems using kernel trick. The hyperplane is used by the SVM algorithm to separate training data points into different groups efficiently, as shown in Fig. 3.

Support vector machines are of two types (1) linear SVMs and (2) nonlinear SVMs. Here, the linear SVMs are being used.

The above Fig. 4 shows the linear SVM that means it separates the spam and ham samples linearly based on a linear function.

$$A = wx + b$$

where  $w$  is the weight assigned to a training instance  $x$  and  $b$  are some constant.

If  $wx + b = 0$ , then it is a margin or a boundary line.

If  $wx + b > 0$ , then it is a spam message.

If  $wx + b < 0$ , then it is a ham message.

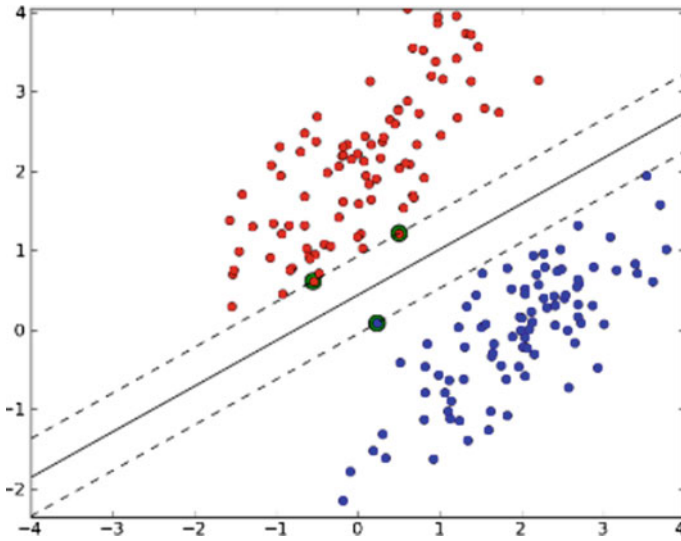


Fig. 3 Support vector machines

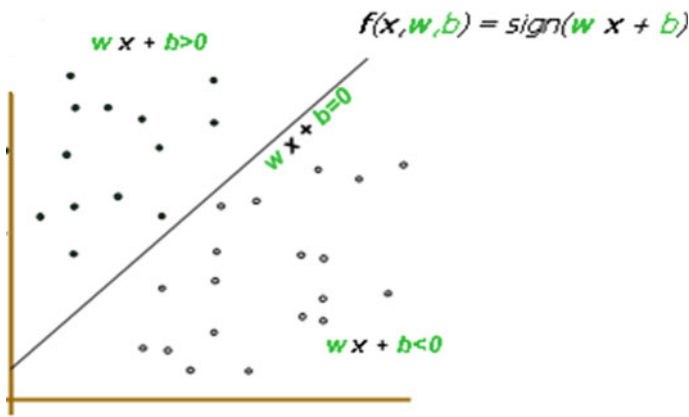


Fig. 4 Linear SVM

Figures 3 and 4 show the misclassified linear SVM that means ham contains spam messages. So, to use some important techniques for solving this problem is needed.

In Fig. 5, a maximum margin classifier is being used for correctly classifying the data means the width of the margin is being increased.

### Naïve Bayes

Literature studies reveal that classification based on probabilistic decisions [8, 9] will give promising results, especially in text classification. Naive Bayes is one such algorithm based on Maximum A posteriori decision rule in Bayesian learning and

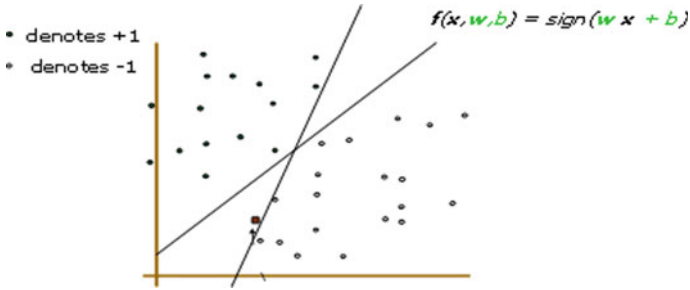


Fig. 5 Misclassified linear SVM

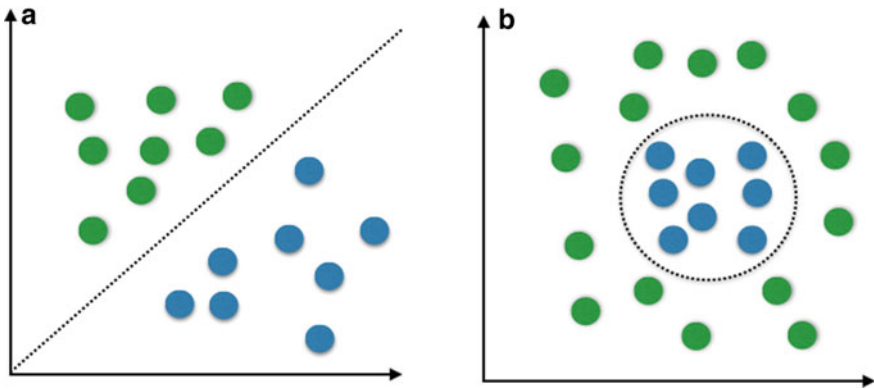


Fig. 6 Naïve Bayes classifier

can be extended for the problems like spam sms detection. Figure 6 shows a simple example.

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability  
↓
↓  
Posterior Probability
Predictor Prior Probability

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

## 4 Result Analysis

This section is to analyze classification results on SMS spam collection data for both Naïve Bayes and SVM algorithms.

### Dataset Collection

Dataset is taken from the kaggle. Dataset consists of a collection of 5574 SMS and 2 attributes [10, 11]. The first attribute is the class attribute whereas the second attribute is text attribute, i.e., SMS. The class attribute has two possible values, namely spam and ham. Among 5574 SMS, 747 SMS are of type spam and 4827 SMS are of ham type [12].

### Naïve Bayes

Naïve Bayes algorithm [13] has applied n the dataset for filtering of spam messages. The table shows the confusion matrix of the Naïve Bayes algorithm. Totally, 4827 messages are ham but the Naïve Bayes algorithm correctly classified 4820 messages only remaining 7 messages it classified as spam but it not correct similarly for spam messages it correctly classified only 738 messages remaining 9 messages are incorrectly classified as ham. So that is the reason the accuracy of the Naïve Bayes algorithm is 96.19% (Table 2).

Table 1 shows that the Naïve Bayes algorithm is applied to the dataset contains 2 attributes namely label, message and the labels present in the dataset are spam, ham. Time taken for executing the Naïve Bayes algorithm is 5 s and the accuracy for the filtering of the spam messages is 96.19%.

SVM algorithm is applied n the dataset for filtering of spam messages. The below table shows the confusion matrix of the SVM algorithm. Totally, 4827 messages are ham but the SVM algorithm correctly classified 4825 messages only remaining, 2 messages classified as spam but it is not correct. Similarly, for spam messages, it correctly classified only 743 messages and remaining 4 messages are incorrectly classified as ham. So that is the reason the accuracy of the SVM algorithm is 98.77%

Table 3 shows the SVM algorithm application on the dataset contains 2 attributes, namely label and message. and the labels present in the dataset are spam, ham. Time taken for executing the SVM algorithm is 10 s and the accuracy for the filtering of the spam messages is 98.77% (Tables 4 and 5).

**Table 2** Observations in Naïve Bayes

Field	Naïve Bayes algorithm
Time taken	5 s
Labels	2
Features	2
Accuracy	96.19%

**Table 3** Confusion matrix

Messages = 5574	Predicted: ham	Predicted: spam
Actual: ham	True positive = 4825	False negative = 2
Actual: spam	False positive = 743	True negative = 4

**Table 4** Observations in Naïve Bayes

Field	SVM algorithm
Time taken	10 s
Labels	2
Features	2
Accuracy	98.77%

**Table 5** Predictions on real time messages

S. No.	Message	Actual class	Predicted class	Result
1	PRIVATE! your 2004 account statement for 07742676969 shows 786 unredeemed Bonus points. To claim call 087912345 Identifier code: 45239	Spam	spam	Pass
2	SMS ac Sptv: The new jersey Devils and the Detroit Redwings play Ice Hockey. Correct or Incorrect? End? Reply END SPTV	Spam	Ham	Fail
3	I call u later, don't have the network. If urgnt, sms me	Ham	Ham	Pass
4	Customer service announcement. You have a New year delivery waiting for u. Kindly call 07046774435 now to arrange delivery [14]	Spam	Spam	Pass

## 5 Conclusion and Future Work

The recent technology solutions solved the delay in communication systems. But, at the same time, facing a problem of spam messages. The current work tried to classify SMS messages into spam and ham. Two standard machine learning algorithms like Naïve Bayes and support vector machine are implemented in this regard. Experiments are carried out on standard sms dataset from Kaggle. Accuracy of 96.19% is obtained with Naïve Bayes algorithm. For SVM, it is 98.77%. Future work includes the adoption of deep learning algorithms for the same. Also, the suitable algorithms need to be tested on large datasets from various spam sources like WhatsApp, snap chat, etc.

## References

1. Liu W, Wang T (2010) Index-based online text classification for SMS Spam filtering. *J Comput* 5(6):844–851. <https://doi.org/10.4304/jcp.5.6.844-851>
2. <https://www.python.org/>. Accessed 23 Dec 2018
3. Nuruzzaman TM, Lee C, Abdullah MFA, Choi D (2012) Simple SMS spam filtering on independent mobile phone. *J Secur Commun Netw* 5(10):1209–1220. Wiley, USA
4. Mahmoud TM, Mahfouz AM (2012) SMS spam filtering technique based on artificial immune system. *IJCSI Int J Comput Sci* 9(2(1)):589–597. ISSN (Online) 1694-0814. [www.IJCSI.org](http://www.IJCSI.org)
5. Cormack GV, Gómez JM, Sáenz EP (2007) Spam filtering for short messages. In: *CIKM '07 proceedings of the sixteenth ACM conference on information and knowledge management*, pp 313–320. <https://doi.org/10.1145/1321440.1321486>
6. Delany SJ, Buckley M, Greene D (2012) SMS spam filtering: methods and data. *Expert Syst Appl* 39(10):9899–9908. ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2012.02.053>
7. Ahmed I, Ali R, Guan D, Lee Y, Lee S, Chung T (2015) Semi-supervised learning using frequent itemset and ensemble learning for SMS classification. *Expert Syst Appl* 42(3):1065–1073
8. Alper K, Serkan G, Semih E, Efnan G (2013) The impact of feature extraction and selection on SMS spam filtering. *Elektronikair Elektrotechnika* 19(5)
9. Chan PPK, Yang C, Yeung DS, Ng WW (2015) Spam filtering for short messages in adversarial environment. *Neurocomputing* 155:167–176. ISSN 0925-2312. <https://doi.org/10.1016/j.neucom.2014.12.034>
10. Mathew K, Issac B (2011) Intelligent spam classification for mobile text message. In: *Proceedings of 2011 international conference on computer science and network technology*, Harbin, pp 101–105. <https://doi.org/10.1109/iccnsnt.2011.6181918>
11. Wu C-H, Tzeng G-H, Goo Y-J, Fang W-C (2007) A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Syst Appl* 32(2):397–408
12. Kawade DR, Oza KS (2015) SMS spam classification using WEKA. *Int J Electron Commun Comput Technol (IJECCCT)* 5(Issue ICICC):43–47. [www.ijecct.org](http://www.ijecct.org). ISSN 2249-7838
13. Yu B, Xu Z (2008) A comparative study for content-based dynamic spam classification using four machine learning algorithms. *Knowl Based Syst*. <https://doi.org/10.1016/j.knosys.2008.01.001>
14. Yadav K, Kumaraguru P, Goyal A, Gupta A, Naik V (2011) SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering. In: *Proceedings of 12th workshop mobile comput. Syst. Appl.* 1–6

# Design of Open and Short-Circuit Stubs for Filter Applications



S. Aiswarya and Sreedevi K. Menon

**Abstract** A resonator is designed to operate as a filter resonating at 2.485 GHz using standard techniques. The design is initiated from the standard LC circuit. Using the insertion loss method, band pass or band stop second order Butterworth filters, with inductive and capacitive elements, are realized. On applying filter transformations to the  $\lambda/4$  low pass filter, the standard  $\lambda/4$  open and short circuit filter design is obtained. The open/short circuit realization of the filter using Kuroda's identity was modified as closed loop resonators, with open and closed stubs, to achieve band stop and band pass operations. The optimization of the obtained circuit is carried out by adding stubs on to the section. A proof of concept of the operation of a resonator as a BPF for short circuit  $\lambda/2$  lines and as BSF for open circuit  $\lambda/2$  lines is carried out in this work. The miniaturization of the filter is done by adding open circuit stubs to BSF and short circuit stubs to BPF maintaining the electrical length of the design fixed. The further miniaturization of the design is carried out using standard techniques. The softwares used for designing and validation of the filters are ANSYS Designer and HFSS.

**Keywords** Filter · Resonator · Lumped · BSF · BPF

---

S. Aiswarya (✉)

Center for Wireless Networks and Applications (WNA), Amrita Vishwa Vidyapeetham,  
Amritapuri, India

e-mail: [aiswaryas@am.amrita.edu](mailto:aiswaryas@am.amrita.edu)

S. K. Menon

Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham,  
Amritapuri, India

e-mail: [sreedevikmenon@am.amrita.edu](mailto:sreedevikmenon@am.amrita.edu)

## 1 Introduction

A resonator is an inevitable part of every communication system. The resonators are designed to provide a maximum output at particular frequencies called resonating frequency. These resonators can be used for different applications based on their performance. The application of resonator in reducing the electromagnetic radiation hazards is discussed in [1]. The details of the design and optimization of planar microstrip based resonators is discussed in [2]. For some designs perturbations are made in the ground plane in order to enhance the better performance [3]. The shifting of the resonator frequency of a resonator can be made possible by using tuning circuits [4]. These resonators find application as filters [5], antennas [6, 7] couplers [8], RFID tags [9–11], diplexers [12] etc. in the communication based systems. Filter is one of the important parts of all the Radio frequency (RF) and microwave based designs. These filters may be active or passive. On the basis of operation the filters can be band pass or band stop. The details of operation of a resonator as Band Pass Filter and Band Stop Filter is discussed in [13]. Depending upon the application the filters may be band pass [14, 15] or band stop [16, 17].

Starting from lumped circuit design and moving on to the planar designs [18] is discussed in this paper. Using the standard methods, the filter designs are miniaturized [19]. A summary of the design and miniaturization of band stop filter (BSF) and band pass filter (BPF) starting from the lumped circuit is discussed in this paper. The planar description and further analysis are described in [20].

## 2 Proposed Designs and Results

An equivalent circuit was derived initially, for the designs of a band pass filter (BPF) and a band stop filter (BSF), via the method of insertion loss (IL) at 2.48 GHz, for the desired bandwidth, cut-off frequency and depth, based on the filter response. The Insertion Loss (IL) of the filter is given by the following equation:

$$IL = 10 \log \frac{P_{in}}{P_{out}} = -10 \log(1 - |\Gamma|^2) - 20 \log |S_{21}| \quad (21)$$

where,  $P_{in}$  is the input power at a source and  $P_{out}$  is the output power delivered to a load; ' $\Gamma$ ' is the reflection coefficient.

The depth of  $S_{21}$  determines filter performance.  $S_{21} = 0$  dB implies the filter is a band pass filter,  $S_{21} = -3$  dB specifies the filter bandwidth and  $S_{21} < -3$  dB delineates the attenuation in the stop band. The design of a BPF and a BSF operating at 2.48 GHz, with a bandwidth of 1.5 GHz and attenuation of  $-30$  dB, using LC circuits, and open and short-circuited stubs, is discussed as follows.



### 2.1 Lumped Element Design for BSF

A second-order BSF can be designed, for operating at 2.48 GHz having a bandwidth of 1.5 GHz and attenuation  $-30$  dB, using an LC circuit in series and parallel combinations. The schematic circuit of the second-order BPF is given in Fig. 1. The design equations for the series and parallel inductors, and capacitors, are noted as follows:

$$C_{1p} = \frac{1}{2\pi(f_2 - f_1)g_{1k}Z_L} \tag{2}$$

$$L_{1p} = \frac{1}{\omega^2 C_{1s}} \tag{3}$$

$$L_{2s} = \frac{1}{2\pi(f_2 - f_1)g_{2k}} \tag{4}$$

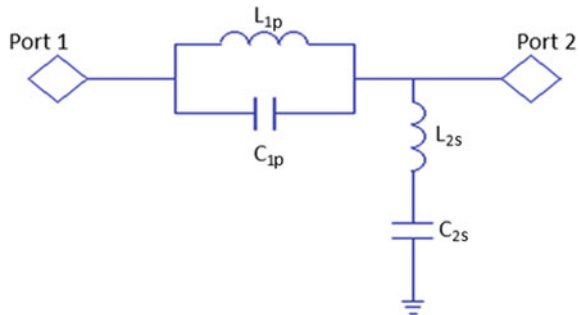
$$C_{2s} = \frac{1}{\omega^2 L_{2s}} \tag{5}$$

where  $L_{1p}$  and  $L_{2s}$  are the parallel and series inductors,  $C_{1p}$  and  $C_{2s}$  are the parallel and series capacitors and  $Z_L = 50 \Omega$ , is the load impedance  $g_{1k}=g_{2k}=1.414$ , are the coefficients of maximally flat band pass filter.

$f_1$  and  $f_2$  are the lower and upper, cut-off frequencies;  $(f_2 - f_1)$  specifies the bandwidth and  $\omega = 2\pi f$  is the angular frequency.

On solving these equations, the values of series and parallel inductors and capacitors are  $L_{1p} = 2.7$  nH,  $C_{1p} = 1.5$  pF,  $L_{2s} = 3.75$  nH and  $C_{2s} = 1.1$  pF. The circuit is build using the designed values. The theoretically designed circuit is validated using simulation software ANSYS Designer and the frequency response is shown in Fig. 2.

**Fig. 1** Schematic LC circuit of second-order BSF



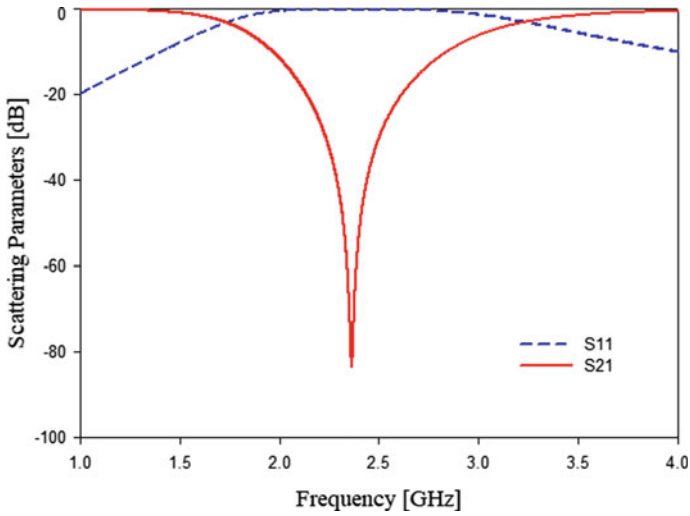


Fig. 2 Frequency response of BSF

### 2.2 Lumped Element Design for BPF

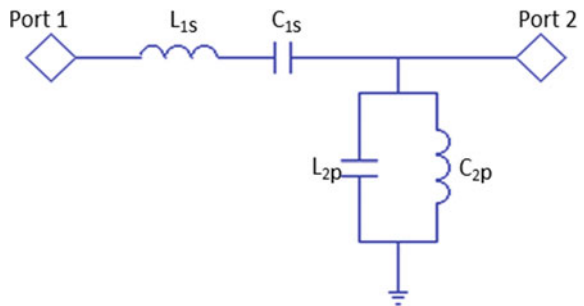
A second-order BPF, using LC components, can be realized as depicted in Fig. 3, with the following parameters:

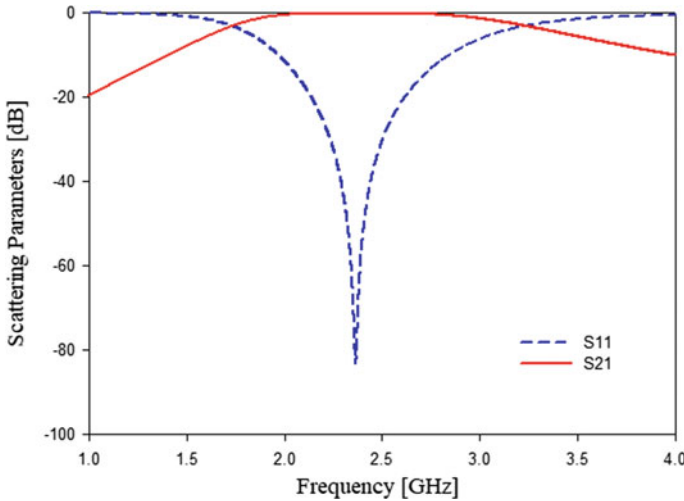
$$L_{1s} = \frac{g_{1k}Z_L}{2(f_2 - f_1)} \tag{6}$$

$$C_{1s} = \frac{1}{\omega^2 L_{1s}} \tag{7}$$

$$C_{2p} = \frac{g_{2k}}{2\pi(f_2 - f_1)Z_L} \tag{8}$$

Fig. 3 LC circuit of BPF





**Fig. 4** Frequency response BSF

$$L_{2p} = \frac{1}{\omega^2 C_{2p}} \tag{9}$$

where  $L_{1s}$  and  $L_{2p}$  are the series and parallel inductors,  $C_{1s}$  and  $C_{2p}$  are the parallel and series capacitors and  $Z_L = 50 \Omega$ , is the load impedance  $g_{1k}=g_{2k}=1.414$ , are the coefficients of maximally flat band pass filter.

$f_1$  and  $f_2$  are the lower and upper, cut-off frequencies;  $(f_2 - f_1)$  specifies the bandwidth and  $\omega = 2\pi f$  is the angular frequency.

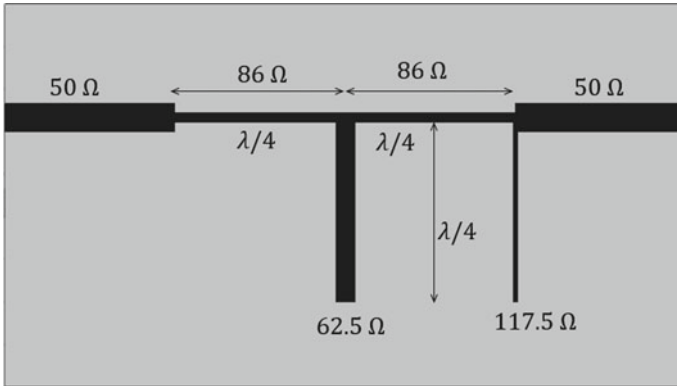
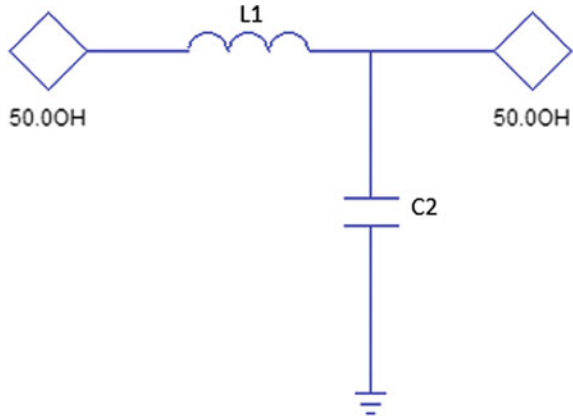
On solving these equations, the values of series and parallel inductors and capacitors are  $L_{1p} = 8.0$  nH,  $C_{1p} = 0.51$  pF,  $L_{2s} = 3.75$  nH and  $C_{2s} = 3.23$  Pf. The theoretically designed circuit is validated using simulation software ANSYS Designer and the frequency response is shown in Fig. 4.

### 2.3 BSF Design Using Open and Short-Circuit Stubs

The followed steps illustrate the design a BSF at 2.48 GHz having bandwidth 1.5 GHz, using open and short-circuited stubs, via insertion loss method:

Realize LPF of the same order with  $f_c = 2.48$  GHz. An equivalent circuit of LPF is realized, as shown in Fig. 5. For a second-order filter, the coefficients are  $g_1 = 1.414$ ,  $g_2 = 1.414$  and  $g_3 = 1$ . For realization, inductors are replaced by short-circuited stubs having length  $\lambda/4$  and capacitors by open-circuited stub having length  $\lambda/4$ . Convert LPF to BPF using Eqs. (10), (11) and (12).

**Fig. 5** LC circuit of second-order LPF



**Fig. 6** BSF design using open and short circuit stub

$$bf = \cot\left(\frac{\pi \omega_L}{2 \omega_0}\right) \tag{10}$$

where,  $\omega_L$  is the lower cut-off frequency and  $\omega_0$  is the center frequency.

$$Z_1 = \frac{1}{Y_1} = \frac{1}{bf \cdot g_1} \tag{11}$$

$$Z_2 = bf \cdot g_2 \tag{12}$$

To the circuit obtained, on applying Richards’ transformations and Kuroda’s Identities, the simplified stub based design is got. The final optimized BSF design, using open and short circuit stubs on designing in simulation software ANSYS HFSS, is shown in Fig. 6. The design is done on FR4 Epoxy having dielectric permittivity 4.4 and loss tangent 0.02. The simulated response of the design is shown in Fig. 7.

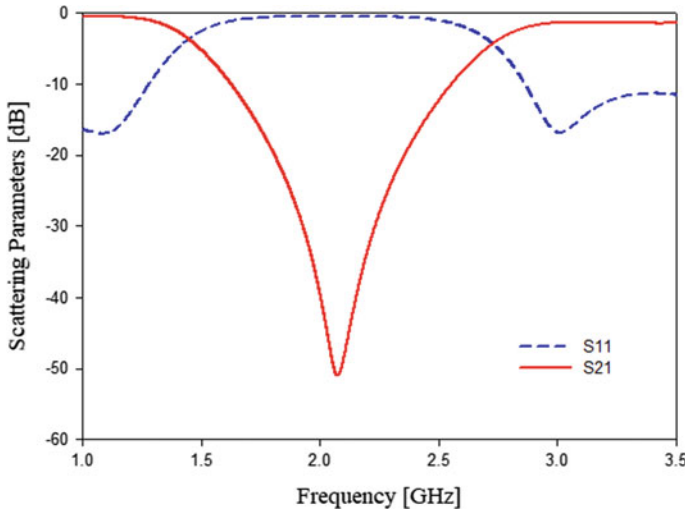


Fig. 7 Frequency response of a BSF using open and short-circuit stubs

### 2.4 BPF Design Using Open and Short-Circuit Stubs

The design of a second-order BPF with center frequency,  $f_c$  2.48 GHz and having bandwidth 1.5 GHz, follows the same procedure as BSF design, carried out with circuit realization, using Eqs. (13) and (14).

$$Z_{01} = \frac{\pi Z_0 \Delta}{4g_1} \tag{13}$$

$$Z_{02} = \frac{\pi Z_0 \Delta}{4g_2} \tag{14}$$

where,  $\Delta$  is the fractional bandwidth  $\Delta = \frac{BW}{f_c}$ , BW is the bandwidth of the filter and  $f_c$  is the center frequency.

The optimized design of a BPF, using open and short-circuit stubs designed in simulation software ANSYS HFSS, is shown in Fig. 8. The design is done on FR4 Epoxy having dielectric permittivity 4.4 and loss tangent 0.02. The frequency response of the designed BPF is shown in Fig. 9.

## 3 Conclusion

The paper describes the arrival open circuit short circuit designs from the standard lumped-element circuit. Initially the filter parameters like the order of the filter,

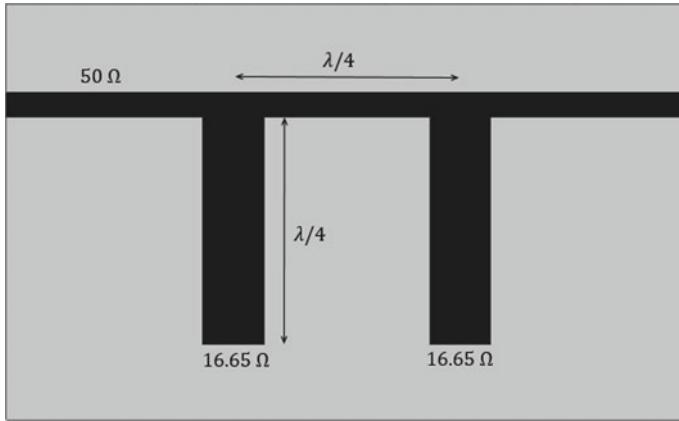


Fig. 8 BPF design using open and short circuit stub

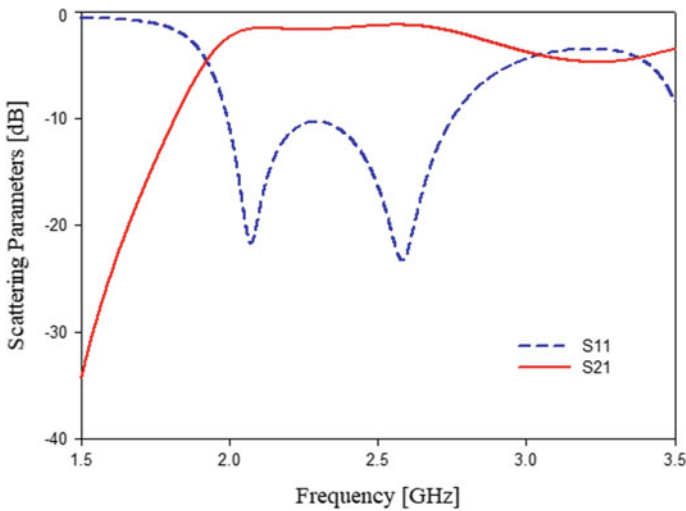


Fig. 9 Frequency response of a BSF using open and short-circuit stubs

operating frequency, band width,  $S_{11}$ ,  $S_{21}$  etc. are fixed. Then by applying filter transformations and Kuroda's identities, the open and short circuit stub based design can be obtained. Again on applying these identities, the filter structures can be again miniaturized. The miniaturized structures are designed and validated using simulation software's ANSYS Designer and HFSS. The same technique can be used in the design of other communication-based structures. These designed filter structures have a wide range of applications in communication systems.

**Acknowledgements** We would like to express our sincere gratitude to our beloved Chancellor and world-renowned humanitarian leader, Dr. Mata Amritanandamayi Devi, popularly known as Amma, for the wonderful atmosphere provided at our campus to promote and encourage research.

## References

1. Meenu L, Aiswarya S, Menon SK (2017) Reducing electromagnetic radiation hazards using resonators. In: 2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM). IEEE; Chen S et al. (2017) An alternate circuit for narrow-bandpass elliptic microstrip filter design. IEEE Microwave Wirel Compon Lett 27.7:624–626
2. Zhang D et al (2016) High selective metamaterial resonator based on complementary split ring resonator. In: 2016 IEEE conference on electromagnetic field computation (CEFC). IEEE
3. Virdee BS (2009) DGP stripline resonator for wideband filter application. In: 2009 Asia Pacific microwave conference. IEEE
4. Potelon B et al (2017) Tunable bandstop resonator based on dual behavior resonator principle. In: 2017 IEEE Africon. IEEE
5. Meenu L et al (2019) Stub loaded semi-circular resonator for filter applications. In: 2019 Sixteenth international conference on wireless and optical communication networks (WOCN). IEEE
6. Meenu L, Aiswarya S, Menon SK (2017) Asymmetric coplanar waveguide fed monopole antenna with perturbed ground plane. In: 2017 Progress in electromagnetics research symposium-spring (PIERS). IEEE
7. Meenu L, Aiswarya S, Menon SK (2017) Compact monopole antenna with metamaterial ground plane. In: 2017 Progress in electromagnetics research symposium-fall (PIERS-FALL). IEEE
8. Oliver BM (1954) Directional electromagnetic couplers. Proc IRE 42(11):1686–1692
9. Aiswarya S, Ranjith M, Menon SK (2017) Passive RFID tag with multiple resonators for object tracking. In: 2017 Progress in electromagnetics research symposium-fall (PIERS-FALL). IEEE
10. Aiswarya S et al (2017) Microstrip multiple resonator assisted passive RFID tag for object tracking. In: 2017 2nd International conference on communication and electronics systems (ICCES). IEEE
11. Karthik A, Aiswarya S, Menon SK (2018) Compact low cost passive rfid tag for object tracking. In: 2018 International conference on advances in computing, communications and informatics (ICACCI). IEEE
12. Ranjith M, Aiswarya S, Menon SK (2017) High isolation diplexer for rf circuits using loop resonators. In: 2017 International conference on advances in computing, communications and informatics (ICACCI). IEEE
13. Vinod A et al (2020) Design and analysis of planar resonators for wireless communication. In: 2020 5th international conference on communication and electronics systems (ICCES). IEEE
14. Chen S et al (2017) An alternate circuit for narrow-bandpass elliptic microstrip filter design. IEEE Microwave Wirel Compon Lett 27.7:624–626
15. Babu NC, Menon SK (2016) Design of a open loop resonator based filter antenna for WiFi applications. In: 2016 3rd International conference on signal processing and integrated networks (SPIN). IEEE
16. Alyahya M et al (2017) Compact microstrip low-pass filter with ultra-wide rejection-band. In: 2017 International conference on wireless technologies, embedded and intelligent systems (WITS). IEEE
17. Mol AS, Aiswarya S, Menon SK (2017) Study of an electromagnetically coupled resonator energized using U-shaped microstrip feed. In: 2017 International conference on advances in computing, communications and informatics (ICACCI). IEEE

18. Pozar DM (2009) Microwave engineering. Wiley
19. Sushmeetha RR, Natarajamani S (2018) Dual-band bandpass filter using complementary split ring resonators for x-band applications. In: 2018 International conference on wireless communications, signal processing and networking (WiSPNET). IEEE
20. Aiswarya S et al (2019) Analysis and design of stub loaded closed loop microstrip line filter for Wi-Fi applications. In: 2019 Sixteenth international conference on wireless and optical communication networks (WOCN). IEEE



# Wireless Sensor Networks and Its Application for Agriculture



**Kirankumar Y. Bendigeri, Jayashree D. Mallapur,  
and Santosh B. Kumbalavati**

**Abstract** The wireless sensor network (WSN) is the most proven technology in today's world. WSN has gained its applications that are different from other networks and even the usage of WSN has also inferred with other networks like VANETS that uses different types of sensors in its network. Arduino or raspberry pi uses different sensors to monitor and data can be obtained from any remote location at a very lower cost. Thus, miniaturization is possible using WSN. In this paper, WSN is being discussed in terms of evolution, scenario, hardware design, application, research issues, design constraints, research problem to provide an insight into WSN in different disciplines. Finally, the applications of WSN in agriculture with routing protocols are also being discussed. Thus, remote monitoring of agriculture using a sensor is demonstrated through this paper.

**Keywords** Wireless sensor network (WSN) · Agriculture · Sensor node

## 1 Introduction

The adoption of wireless sensor network (WSN) has been heard over a decade, and its applications are now witnessed a global presence. Characterized by low computational capability and minimal resources, a sensor node is programmed to extract environmental data and forward to the sink. This data forwarding mechanism is affected by many factors, e.g., routing technique applied, deployment strategy

---

K. Y. Bendigeri (✉) · J. D. Mallapur

Department of Electronics and Communication, Basaveshwar Engineering College (A), Bagalkot  
587103, Karnataka, India

e-mail: [kiranbendigeri@gmail.com](mailto:kiranbendigeri@gmail.com)

J. D. Mallapur

e-mail: [bdmallapur@yahoo.co.in](mailto:bdmallapur@yahoo.co.in)

S. B. Kumbalavati

Department of Electronics and Instrumentation, Basaveshwar Engineering College (A), Bagalkot  
587103, Karnataka, India

e-mail: [sbkumbalavati@gmail.com](mailto:sbkumbalavati@gmail.com)

adopted, security protocols implemented, etc. Out of all these factors, energy is one of the prominent factors that affect the data forwarding operation of a node. This proposed work shows that the real-time sensors can also be used using routing protocols as most of the authors consider single-hop communication between sensor node and gateway. Routing protocols used in sensors makes it possible to have 2 or more sensors with a single gateway so that communication of larger distance can be done, thus using minimal sensors large coverage of agriculture area is possible. This paper introduces the fundamental concept utilized in the proposed study followed by the research problem adopted, research goals, and methodologies with one of the applications of WSN.

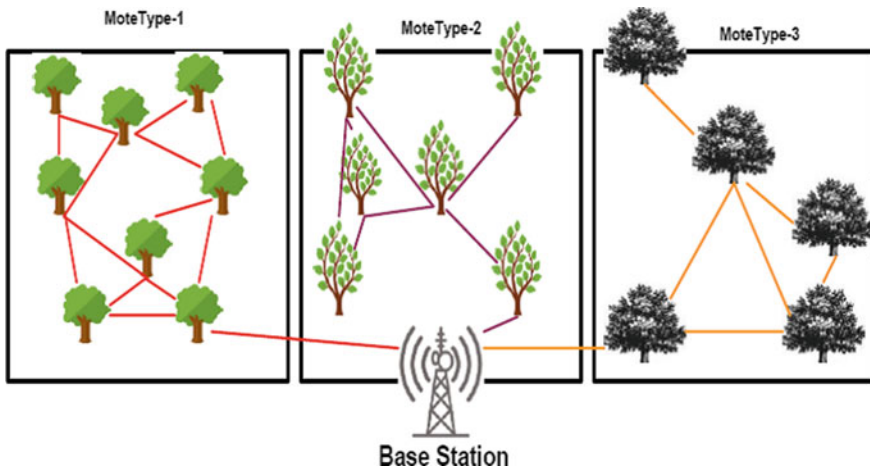
### ***1.1 Wireless Sensor Network Background***

WSN is a network that is built up of multiple sensors (popularly called ‘Motes’), which consists of sensors, computing also and communication subunits. Ad hoc technologies and computation are initiating the network collaboration among these nodes in distributed manner. The first ever known WSN [1] application was implemented in 1950s by United States Military to detect and track the opponent USSR submarines, by the name sound surveillance system (SOSUS). This application is still serving US Military for peaceful functions of monitoring underwater wildlife and volcanic activities. Research activities on WSN started in academia at the late 1970s. Carnegie Mellon University and Massachusetts Institute of Technology joined their hands with US Military and started the distributed sensor network (DSN) program under the banner of United States Defense Research Project Agency (DARPA) [2]. They began exploring the challenges in implementing distributed/wireless sensor networks. The scope of WSN is for many defense and civil society applications where the various environmental parameters act as core data or information to make application workable. The collected data routes in a hop wise manner towards the primary data sink. The data from the sink traverse through the gateway to the monitoring stations [3].

### ***1.2 Wireless Sensor Network Scenario***

Fig. 1 below depicts a typical WSN (IEEE 802.15.4.) supervising the area of interest will initiate by deploying the application-specific sensor. Clusters are created to minimize the overload on the entire network; wherein cluster-head governs each cluster, also takes the responsibility of collecting, aggregating, and forwarding the data to the base station [4].

Although in the beginning, WSN were employed for a subtle deployment and that limited to the defense-related applications, today the invent of internet of things (IoT) has made WSN as a core component of it [4]. Applications which are built from



**Fig. 1** A scenario of typical WSN

WSN and to name the few includes, smart cities, smart homes, healthcare, predictive maintenance, energy-saving smart grids, high confidence transport and asset tracking and intelligent buildings, etc. [5, 6]. A WSN will comprise several groups of ‘nodes,’ ranging from a few hundred nodes to several thousands of nodes, where each node will connect to another node or multiple neighboring nodes. A wireless sensor node is a typical equipment consisting a radio transceiver with inbuilt or external antenna, an application-specific sensor (transducer), a microcontroller, an interfacing electronic circuit housing Digital-Analog converter, a memory unit and an energy source in the form of battery. Occasionally, sensor nodes may also be facilitated with a solar cell and a charging circuit to extend the energy supply. A sensor node can have a size of a brick to a magnitude of a visible dust particle depending on the application. A sensor node, the familiar name is ‘mote,’ any functioning mote with a microscopic size is still under research. The cost of a mote may vary from a few dollars to several hundred dollars; it is application-specific, i.e., depending on the sensor used and the complexity of the mote. Earlier WSN technology [7] was narrow due to factors like insufficient bandwidth, limitations in the integrated circuit design, the size of the network nodes, the cost of the node, inefficient power management techniques [8].

The improvements in WSN technology allows robust monitoring, better power management techniques, control over remote monitoring, and even successful implementation in human inaccessible areas. Features like bidirectional communication, data transfer with acknowledgment, and also output conditions are configurable. The research in WSN and RF technology has extended the applications of WSN to agribusiness, water management, and underwater acoustic and deep-space networks. But, mostly WSN is preferred in human inaccessible and nodes are prone to damages due to natural calamities, battery life, stepping of human or animals, etc. A lot of research challenges are open for developing a rugged node, better node connectivity

as WSN is an infrastructure-less and self-configurable network, lifetime optimization as the battery is not substitutable in the unreachable areas [9].

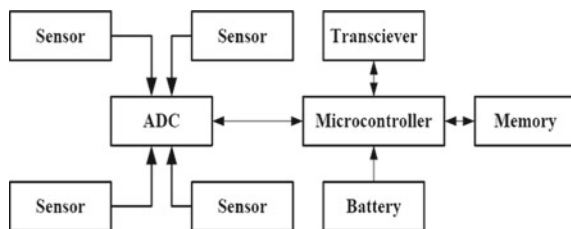
It is imperative that if any sensor node is not completely operational, necessary action should be taken to overcome any possible situation that may lead to degraded performance of the WSN upon identification of malfunction. Such defective nodes also result in an adverse influence on the WSN connectivity. The affected node not only stops the communication from itself, but also acts as an impediment of progressive routing from the other nodes as well. Hence, various existing topologies in WSN must be studied carefully and evolves up with better topological-based research to understand the coverage issues. Studies towards such direction can affirm a series of nodes to form a network which has less chance of getting broken. At the same time, equal stress should be given to the energy efficiencies, without which the sensor nodes are nonfunctional.

### 1.3 Hardware Design of Mote

A WSN requires all its constituent nodes to be productive and readily accessible, as it is most important during the construction. As per the demand of the application, the nodes participating in the construction of WSN should satisfy the prerequisites of the target application. They must be miniature in size, energy-efficient, modest, they must be equipped with an appropriate sensor, appropriate transceiver, sufficient memory hardware, and computing facilities. Extensive research work is in progress focusing miniaturization of the node size, optimization of energy consumption, economical design, based on the generalized node to application-specific requirements in the node.

Figure 2 exhibits the building blocks of a typical wireless sensor node are the microcontroller, transceiver, memory, sensor, an electronic integrated circuit, and a power source [10]. The controller schedules tasks like executing algorithms efficient energy utilization, controlling and processing the data collected from the sensor, memory management and controls the functionalities of all other components. Usually, a controller may be microcontroller/microprocessor. The transceiver is a combination of transmitter and receiver; transceiver operates in four different states like transmit, receive, idle, and sleep. Wireless sensor nodes are usually equipped

Fig. 2 Architecture of typical sensor node



with RF antennas, as radio frequency is the best suitable communication medium for WSN. Also, sensor nodes can use optical communication or an infrared medium. Infrared does not require an antenna, but they are limited to broadcasting only. Wireless sensor nodes equip with two different types of memories, one is program memory for programming the microcontroller, and the other is user memory for storing the data.

WSN are usually implemented in human unreachable regions, replacing the battery is ruled out. So, while developing an energy source for a wireless sensor node requires provisioning of adequate power available to energize the system for a long duration. A sensor node may get equipped with either rechargeable batteries through solar cells or a non-rechargeable battery [11]. Sensors are nothing but transducers having the capability of converting one form of energy to another form, for example, a thermal sensor senses atmospheric temperature and converts heat to electrical energy. Similarly, humidity, vibration, chemical, radioactive, sound, sonar, light, etc. Sensors are used based on the application requirements. Integrated electronic circuitry is also included to interface sensor, transceiver, the power source to the microcontroller, to provide an interface between microcontroller and analog to digital converter, an essential component to process the sensed data for processing and the output data from the microcontroller to antenna.

## 2 Literature Review

In [1] the author discusses various aspects of WSN like programming, security, localization, time synchronization, power management, network layer, physical layer, operating systems, node architecture. WSN is again well briefed in [2] on different issues like the anatomy of a sensor node, radio communication, link management, multi-hop, clustering. WSN is discussed in [3] with its application, MAC protocol, energy-centric simulation, privacy, and security [7]. Here, author concludes that clustering is a method to increase WSN lifetime and introduces the working of WSN. Here, author [4] discusses the advancement of sensor technology, smart instrumentation, WSN miniaturization [5]. Discusses design and modeling, network management, data management, security, and WSN applications. Here, [6] data collection, routing and transport protocols, energy-saving approaches, data storage and monitoring, physical layer and interfacing, and WSN application are discussed. This paper [8] discusses an overview of design fabrication and test-based MEMS inertial sensors and also shows how an application can be transformed into a practical design. This paper [9] discusses single-path routing protocols like LEACH, PEGASIS. Discussion is also extended for multipath routing protocols. Here [10] WSN applications have been discussed.

### 3 Wireless Sensor Network Application

At present, the sensor network has multiple commercial applications that demand both time-bound and mission-bound response. Some of the widely-known application:

- *Security Applications:* Sensor nodes are applicable in capturing data in stealth mode in various adverse conditions. Such applications are useful in monitoring line-of-control to check intrusive movements. Consequently, observation of unmanned regions, assets, perimeter, borders, and plateaus can be productive with the help of sensor networks.
- *Environmental Monitoring:* A sensor node contribute to a significant role in environmental monitoring, some key research areas under environmental monitoring are watershed, scientific investigation, pollution monitoring, weather, natural calamities, chemical disaster, mining safety, etc.
- *Industrial Development:* Industrial process automation, process control, production optimization, to manage the impact of environmental influences on production, remote monitoring of production process, etc.
- *Agriculture:* The modern approach of farming enables to improve the yield of the crop and minimize the cost incurred using sensors. It helps to measure various atmospheric and soil parameters. It also supports for intruder prevention, pest detection, disease prevention, protection against fire accidents, irrigation automation, etc.
- *Natural Disasters:* Collaborative decision techniques have enhanced wireless sensor networks to warn against natural disasters like earthquake, volcanic eruption, landslide, forest fire, twisters, tsunami, etc.
- *Automotive:* Wireless sensor motive industry has a broad range of application in the automotive sector, e.g., resisting vehicle collision, smart parking, identification, monitoring natural calamities, etc.
- *Health:* Body area networks have enabled WSN to monitor various parameters of the human body. Wireless autonomous sensors implanted in different parts of the body or wearable sensors used to measure the critical parameters of the patient and forward the clinically-important data to the physician based on which the doctor remotely monitors patient with the help of data received.
- *Monitoring of Structures:* Continuous monitoring of structures like bridges, buildings, etc., for their health status. Continuous monitoring is critical for structural integrity, environmental factors, wear and tear due to load, temperature humidity, corrosion, etc.
- *Under Water Sensor Networks:* Sensor nodes are useful in monitoring event under the water using acoustics. It is also important to note that wireless sensor networks can be utilized only at thermally solid depth; else acoustic communication is not much favorable due to reflection and refraction.
- *Future Markets:* Today WSN is becoming popular in retail outlets and public places for automatic door opening, motion detection, fire alarm, CCD devices, etc.

### 3.1 *Research Issues in Wireless Sensor Networks*

Some of the key research areas are:

- **Transducer design:** Development of bio-degradable, environmental friendly sensors, also adhering to the constraints compact in size, low power consumption, and economical designs.
- **Electronic system design:** Miniaturization of the integrated circuit with modern techniques is the real challenge today. Digital data extraction from integrated sensors through appropriate electronic circuitry, data aggregation, and transmission through electronics and removal of noise using sensor arrays are possible in today's technology.
- **Node design:** Design and development of robust and rugged nodes, yet consumes very less power, high processing capability, and excellent network communication.
- **System Design:** This problem pertains in using the existing nodes while applying a scheduling scheme.

### 3.2 *Design Constraints for Wireless Sensor Networks*

The design constraints of the sensor network are as follows, [12].

- **Fault Tolerance:** WSN deployed in unmanned areas, may fail due to blockages, power failure, physical damage, environmental issues or intruder, etc. However, such issues should not be affecting the usual networking performance of the sensor node.
- **Scalability:** Scalability is another critical design constraint over sensor network which states that some sensors used over the area affect the performance. Environmental monitoring may need several thousand nodes; a megastructure may require millions of nodes. But, a simple application like surveillance may require several hundred to less than a hundred nodes. The new scheme developed should be able to work with several nodes to millions of nodes.
- **Production Cost:** The deployed sensor nodes are not easily replaceable, and also, they are prone to damages due to various reasons. The number of nodes deployed may vary from several hundred to several million nodes application specific; the production cost is paramount and should be very low.
- **Hardware:** A sensor node is always application-specific, apart from the essential components. The design of the sensor node should accommodate the entire required element, yet the size should be minuscule sometimes to the size of the dust.
- **Network Topology:** Wireless sensor networks are deployed based on the purpose. In the deployment stage, it can be placed uniformly, thrown randomly, dropped by plane, etc. In the post-deployment phase, the topology of the network may change due to node placement and its position, the distance between two consecutive

nodes for connectivity, obstacles, damages, etc. In some cases, additional nodes may be redeployed to re-establish the coverage and connectivity.

- *Environment*: Sensor nodes are also deployed in chemically contaminated areas, war field, deep ocean, valleys, mountains, megastructures, etc.
- *Transmission Media*: Sensor nodes today adopt different transmission media like infrared, bluetooth, optical, radio waves, etc. Enabling the network for universal operation requires, proper transmission media should be made available. The traditional transmission medium is radio frequency 2.4 GHz. But, smart dust adopts optical medium for communication.
- *Power Utilization*: The sensor being a microelectronic device can be equipped with the limited battery source, typically <0.5Ah, 1.2 V. The node has to perform three essential operations like sensing, communication, and data processing. Therefore, a sensor node lifetime depends on the battery life of the sensor.
- *Heterogeneity*: Modern application demands programming to support communication between different hardware technologies, as the applications areas are becoming more and more challenging. Sensor development is also very fast, so the program should support the broad range of sensor node technologies.
- *Coverage*: A connected sensor defines how effective the coverage range is and how stable the route establishment has been for a given area. When nodes distributed sparsely, only a particular area of interest may get covered. Redundant nodes may occur in some physical locations where accuracy and redundancy are required.
- *Connectivity*: The physical location of the individual sensor and the distance between two consecutive sensors, whether they are in communication range or not, defines the connectivity of the network. Connectivity is irregular due to the partitioning of the network, even if some part of the network gets partitioned, transmission can be done by using mobile nodes.
- *Lifetime*: The battery of a sensor node has a definite lifetime which consistently drains during the communication. Development of extended life battery with proper power scheduling algorithm is very essential.

### ***3.3 Research Problem for Wireless Sensor Network***

Multiple dependable factors contribute to ensuring connectivity among the nodes in WSN, where the primary factor is energy and secondary factor is node deployment strategy. However, the presented study inclines towards investigations in the direction of energy efficiency mainly along with the fault-tolerant operation of the nodes. Such identification will help in setting up a proper wireless sensor network for various applications.

The occurrences of node failures are usually caused owing to security breaches in the wireless sensor network. Due to the various types of attacks, there is a possibility that a compromised node can act maliciously and can tend to either drop a packet or corrupt the routes. There is also a possibility of uncertain hardware circuitry problems



owing to certain external reason causing node failures. Imperfect usage of routing protocol may also lead to dissemination of error-prone data.

The present study focuses on the critical cause of node placement as well as fault-tolerant issues on the energy factor. In the long run, when the sensor nodes start depleting the energy, the declined performance of that node affects the entire network performance negatively. Such a situation very often gives rise to network partitioning problems.

The availability of sufficient residual energy significantly supports the potential radio transmission in wireless sensor network [13, 14]. When a node depletes energy, other nodes connected to it have to dissipate extra energy to make the communication operational, thereby causing faster degradation of available power. Another significant cause of energy dissipation is external environmental factors, e.g., rain, the surface condition of the earth, interference, noise, etc. Such environmental factors strongly influence the performance of the sensor nodes for capturing the real-time event of the surroundings and uses extra battery life to process the raw information. The presence of various types of natural objects like trees, hills, etc., also affects the transmission quality resulting in degradation of communication performance. The prime research problems of the proposed study are as follows:

- **Node Placement:** Node placement is one of the significant problems that initiate from the beginning of the cluster setup stage. Majority of the existing trends of research towards WSN uses random deployment of the sensor for the large scale area and uniform (or grid) deployment for the smaller-scale area. This is mainly done for the convenience of deployment viewpoint. However, random deployment does not ensure that all the nodes are in sensing range of each other while grid deployment does not ensure effective optimization of energy and resource.
- **Energy Efficiency:** From more than a decade, energy has always been the primary problem. Even the existences of hierarchical protocols are in constant phases of enhancement, which means lack of any protocol to ensure maximized energy conservation. Moreover, with lower computational capability, the sensor node will need to deplete more energy to perform transmission. In the concept of data aggregation, a cluster-head consumes maximum energy which is still not found to be lowered. Hence, energy is one of the ongoing issues which need a smart alternative solution without degrading communication.
- **Fault Tolerance:** Sensor network is always characterized by a large number of interconnected nodes, where the presence of a direct link is extremely less. Majority of the nodes performs communication with the help of the relay node. Therefore, any form of fault in a sender can generate a faulty sensory reading.
- **Agricultural Application:** Sensors used over the crop field normally extracts multiple forms of environmental data for ensuring better information gain for crop cultivation. These sensors are normally affected by harsh climatic condition, and hence, they will not operate as anticipated over an as large scale. At the same time, physical damage in any sensor will not draw any attention due to large scale operation. Hence, at any cost, the node connectivity management gets adversely affected which indirectly affects the crop cultivation.

**From the survey of WSN, it is found that WSN has narrow work in a few areas which can be taken up for research**

1. Node placement and higher fault tolerance that ensures energy efficiency, reliability, and
2. Cost-effectiveness with better applicability over agricultural-based application.

## 4 Proposed Work

In proposed work, our discussion will be towards the usage of WSN for agriculture. Real-time sensor nodes along with humidity, temperature, moisture sensor (HTP) are considered. Routing protocol level-based (LBR) and MAC-based (MBR) are considered. Figure 3 shows communication between four sensor nodes and a radio module with a gateway with neighbors involved in communication. Such instance would cause a higher range of communication, as with four sensors will be able to communicate for about a km long. Thus, a unique approach is being found to replace the typical way of agriculture by introducing technology and calling it as e-agriculture.

Figure 4 shows the working with sensors and radio modules, which considers the parameter Humidity, temperature, pressure (HTP) that is related to the crop. As shown in Fig. 4, all the three parameters at times can be measured from a crop of the greenhouse. The algorithm is such that any change with temperature, pressure, or humidity is automatically detected onto the screen. The readings are brought down to the computer interface and displayed on screen which is a desktop or laptop, which in future can be extended for remote users with the farmer from anywhere using mobile. The HTP sensor is again been used with both LBR and MBR routing protocols, where LBR is performing well with HTP. Reason for better performance is as the data is through neighbors, part of data is lost with MBR protocol.



**Fig. 3** Scenario of two sensors setup for measuring various parameters of crop

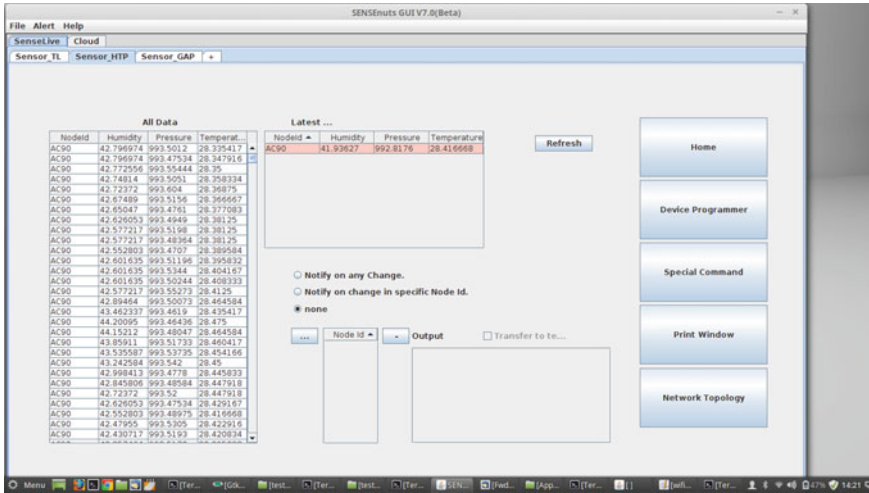


Fig. 4 LBR-based routing protocol to measure temperature, humidity, pressure

### 5 Simulation

The sensor used for sugarcane monitoring, monitor data for HTP and able to successfully receive data at the gateway which is away from the sensor nodes. Figure 5 shows a measurement of light intensity with the crop using both LBR and MBR routing protocols; the methodology is similar to that done with the temperature measurement. Better readings can achieve with LBR routing that has been again compared with a simple hop to hop communication using a radio module sensor with the radio module gateway, of which readings are same as that of LBR routing which is used with fewer nodes for a shorter range of communication.

Figure 6 shows a graph of humidity measurement of the crop in the greenhouse, using LBR and MBR routing protocols and Fig. 7 shows a measurement of pressure using same routing protocols, and dynamo is measuring pressure and humidity related with the crop. As these are the parameters related to water content with crop and the water presence may be due to many reasons like rainwater or natural water from the stored tank or vapor content during the morning hours, because of these reasons variation in the readings may be obtained. Hence, the measurement with the greenhouse effect is preferred. However, the LBR protocol can achieve better with both the measurements.

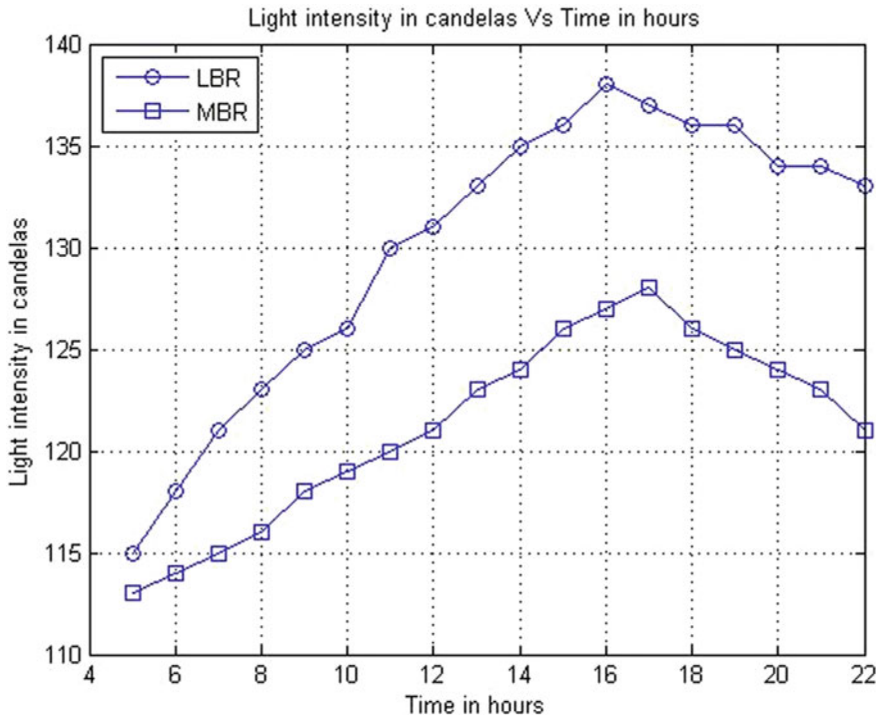


Fig. 5 Showing light intensity measurement for LBR and MBR routing protocols

## 6 Conclusion

In this paper, the architecture, application, design constraints, research problem, etc., of WSN are discussed so that researcher can get to know all the basic information of WSN. Further, one application of WSN in agriculture is considered. Most of the time sensors used as single-hop communication to the gateway are replaced by routing technique to increase the communication range of WSN and to save energy of nodes, demonstrated through the proposed experiment.

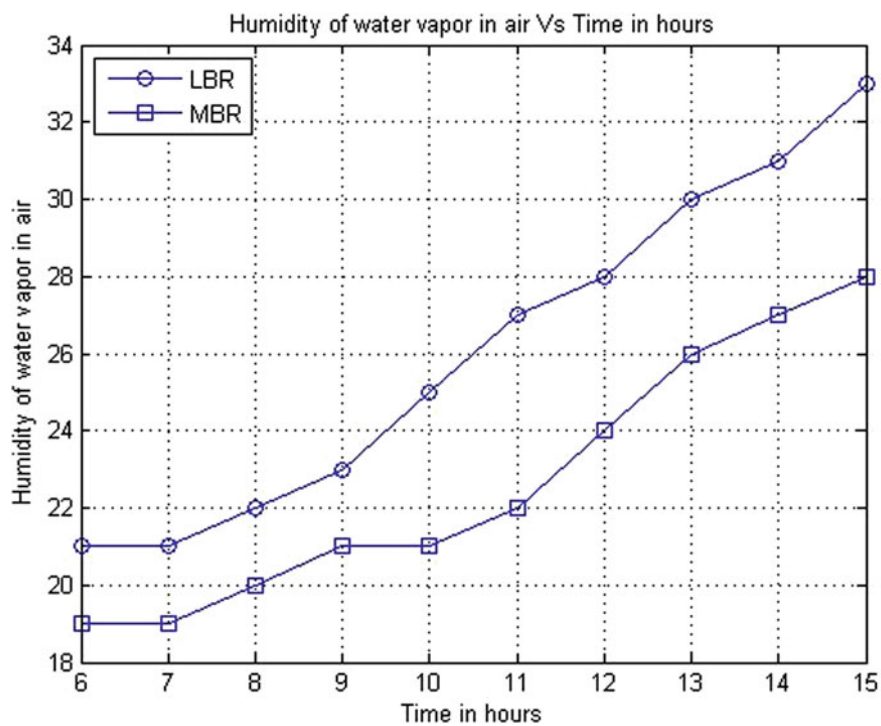


Fig. 6 Showing humidity measurement for LBR and MBR routing protocols

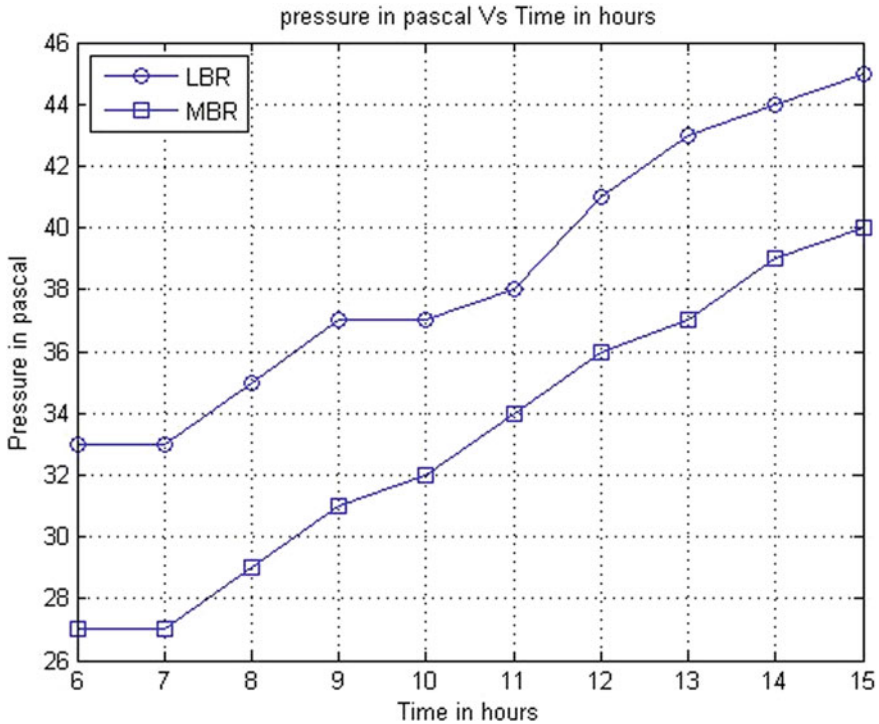


Fig. 7 Showing a pressure measurement for LBR and MBR routing protocols

## References

1. Dargie (2016) Christian Poellabauer, Fundamentals of Wireless Sensor Networks: Theory and Practice, John Wiley & Sons
2. Forster A (2016) Introduction to wireless sensor networks. Wiley
3. Khan S, Pathan ASK, Alrajeh NA (2016) Wireless sensor networks: current status and future trends. CRC Press
4. Mukhopadhyay SC (2014) Internet of things: challenges and opportunities. Springer Science & Business Media
5. Li Y, Thai MT (2008) Wireless sensor networks and applications. Springer Science & Business Media
6. El Emery IMM, Ramakrishnan S (2013) Wireless sensor networks: from theory to applications. CRC Press
7. Aungnoo MA, Laloo N, Sunhaloo MS (2014) Wireless sensor networks: clustering concepts. Omniscryptum GMBH & Company KG
8. Kempe V (2011) Inertial mems: principles and practice. Cambridge University Press
9. Akyildiz IF, Vuran MC (2010) Wireless sensor networks. Wiley
10. Kumar V (2013) Fundamentals of pervasive information management systems. Wiley
11. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) A survey on sensor networks. IEEE Commun Mag 102–114
12. Romer K, Mattern F (2004) The design space of wireless sensor networks. IEEE Wirel Commun 54–61

13. Kumar AD, Smys S (2019) An energy efficient and secure data forwarding scheme for wireless body sensor network. *Int J Netw Virtual Organ* 21(2):163–186
14. Mugunthan SR (2019) Security and privacy preserving of sensor data localization based on internet of things. *J ISMAC* 1(02):81–92
15. Tsai CW, Hong TP, Shiu GN (2016) Metaheuristics for the lifetime of WSN: a review. *IEEE Sens J* 16(9):2812–2831
16. Smys S, Raj JS (2014) Analysis of wireless mesh structures using localized backbone constructions. *J Adv Comput Netw* 2(4)

# Human Disease Diagnosis Using Machine Learning



Sarika Jain, Raushan Kumar Sharma, Vaibhav Aggarwal,  
and Chandan Kumar

**Abstract** In the disease diagnosing process, identification of patterns is so crucial for the prediction of the disease accurately. A powerful disease prediction system is needed to detect the disease accurately by analyzing various parameters/symptoms and to make our systems learn from the past diagnosed disease and capable of adapting to new methods. To deliver the medical association for the analysis of syndrome among patients, a graphical user alliance will be refined. Doctors and medical practitioners can readily utilize the GUI as a screening tool. In this paper, various techniques available for disease prediction such as k-nearest neighbor (KNN), Naïve Bayes, support vector machine (SVM), logistic regression, and decision tree have been explained. A general review is done on the prevailing and anticipated models for human disease prediction and a reasonable study on these techniques based on quantitative dimensions such as detection rate (D-Rate), false alarm rate (FA-Rate) is carried out. Strategies have been proposed for diagnosing liver, heart, and diabetes illness in patients by utilizing machine learning procedures. The three machine learning procedures that were utilized incorporate SVM, logistic regression, and kNN. The framework was executed utilizing every model and their exhibition was assessed. Execution assessment depended on certain exhibition measurements. Finally, a model has been developed in which the person can insert his symptoms and the disease identification is done that can give the result based upon the inputs.

**Keywords** Naïve bayes · K-nearest neighbors (KNN) · Support vector machine (SVM) · Logistic regression and decision tree

---

S. Jain (✉) · R. K. Sharma · V. Aggarwal · C. Kumar  
National Institute of Technology, Kurukshetra, Haryana, India  
e-mail: [jasarika@nitkkr.ac.in](mailto:jasarika@nitkkr.ac.in)

R. K. Sharma  
e-mail: [raushan4myself@gmail.com](mailto:raushan4myself@gmail.com)

V. Aggarwal  
e-mail: [vvaibhavaggarwal9@gmail.com](mailto:vvaibhavaggarwal9@gmail.com)

C. Kumar  
e-mail: [chandankumar0165@nitkkr.ac.in](mailto:chandankumar0165@nitkkr.ac.in)



# 1 Introduction

Machine learning enables the systems to learn without being adjusted remotely. Classification algorithms are utilized to choose from an arrangement of choices that best fits a lot of perceptions. The recent machine learning planned machine learning frameworks have been balanced; however, more ought to be done on their acknowledgment rate for better precision in diagnosing the infection. However, most of the diagnosis systems that are already developed works on a particular domain that is not much pragmatic in a real life scenario. Sometimes, it happens that a patient may be suffering from multiple diseases; so in this case, the system gave the poor results.

## Various Techniques of Disease Detection:

The disease detection system can be built through numerous techniques like fuzzy logic, neural network, agent-based and machine learning. The whole disease has different symptoms known, so by using machine learning techniques, like k-nearest neighbor (KNN), Naïve Bayes, support vector machine (SVM), decision tree, and logistic regression, the disease can be diagnosed. The different machine learning models include:

- **NaïveBayes:** Naïve Bayes is a modest but amazingly prevailing prognostic demonstrating algorithm. Naïve Bayes theorem delivers a method in which the possibility of a hypothesis given our earlier facts can be computed.

*Bayes' Theorem is stated as:*

$$P(m|n) = (P(n|m) * P(m))/P(n)$$

Here,  $P(m | n)$  is the possibility of hypothesis and the facts. This is termed as the posterior probability.  $P(n | m)$  is the possibility of facts to the hypothesis was right.  $P(m)$  is the possibility of the hypothesis being right (regardless of the data).  $P(n)$  is the possibility of the facts (irrespective of the hypothesis). Keywords: m for hypothesis and n for facts after scheming the posterior probability for diverse hypotheses, the hypothesis with the maximum possibility is chosen. This is the extreme possible hypothesis and can be termed as a maximum posterior (MAP) hypothesis.

This can be termed as:

$$\text{MAP}(m) = \max(P(m|n)).$$

Bayes' Theorem is categorized under the classification algorithm. The system is easier to recognize when labeled with binary or categorical input values. It is termed as Bayes' Theorem because it computes the possibilities for an individual hypothesis. They all are easy to make our calculation manageable. Instead of trying to compute the values of individual characteristic value  $P(n1, n2, n3 | m)$ , it is assumed that they are provisionally autonomous based on the targeted value and are calculated as  $P(n1$

$|m) * P(n2 | m)$  soon. Learn the behavior of the disease and can, therefore, predict the disease to which class it belongs.

- **Support Vector Machine:** SVM aims to discover the most suitable hyperplane that separates records into one of kind classes. The scikit-research Python bundle is used to implement SVM. The pre-processed records are divided into test statistics and education units representing 25% and 75% of the whole data set, respectively. Vector support technologies are based on the idea of choice plans that describe choice parameters. A choice plan is single that splits a group of items that have diverse class members. A vector support machine acquires these data points and generates the hyperplane (which in two dimensions is simply a line) that separates the labels at best. The line is the decision limit for SVM, which maximizes the margins of both labels.
- **Decision Tree:** A decision tree is the popular algorithm of machine learning, used mainly for classification, but also for regression problems. Each branch of the tree represents a possible decision reaction or occurrence. It is also known as tree-diagram. The whole idea is to divide the given dataset into fragments during training, based on the value of a certain feature. The fragments are made progressively until all the target variables fall under one category.
- **K-Nearest Neighbor:** K-nearest neighbor is one of the simplest supervised machine learning algorithms, which works based on similarity with the existing cases. All the history cases are stored and the new data or the new case is classified based on a similarity measure. It can be used in classification as well as regression problems. It memorizes the training data. No learning is performed, all the processing happens at the time when the prediction is requested. KNN does not have a discriminative function (analysis is used to determine which variables discriminate between two or more naturally occurring groups) from a training set.

To diagnose any disease, identification is a vital task. Sometimes, various signs and symptoms can be not identified and, therefore, diagnosis is the utmost difficult task. In this work, a comparison of numerous techniques of machine learning for the identification of liver, heart, and diabetes diseases has been provided. Later, a model is developed in which the person can insert his symptoms and the disease identification is done.

Section 2 presents a literature survey that has been done on disease prediction systems. Section 3 describes the simulation of different machine learning models on the sample datasets of three diseases. Section 4 describes the proposed framework along with results and discussion. Section 5 concludes the work describing the future scope of the system.

## 2 Literature Survey

Domingos and Pazzani [1] considered two algorithms of decision tree, namely C4.5 and CART and the dataset of BUPA disease disorder was also considered for the experiments. Schiff et al. [2] also made a critical look on liver disease diagnosis by evaluating classification algorithms such as Naïve Bayes classifier, C4.5, kNN and more. Ephzibah [3] built a model to diagnose diabetes. The model proposed by the author connects fuzzy logic and genetic algorithm and used to select the finest subcategory of characteristics and also to improve the accuracy of the classification. For the experiment, the author picked the data set from the UC Irvine machine learning repository that has 8 features and 769 rows. In this paper, implementation is done through MATLAB. By means of the genetic algorithm, among all, only the three best characteristics are designated. These selected characteristics are used by the classifier known as the fuzzy logic and give 87% accuracy. The cost of around 50% is less than the original cost.

Vembandasamy et al. [4] frolicked a piece to investigate coronary infection. The precision accessible with the aid of Naive Bayes is 86.419%. The researchers Parthiban and Srivatsa [5] have worked upon locating the coronary contamination in the sufferers of diabetes. Tan et al. [6] have given a crossover strategy by utilizing genetic algorithm and support vector machine. Sarwar and Sharma [7] endorsed running in Bayes Theorem to forecast type 2 diabetes. There are three types of diabetes. There are three kinds of the diabetes type-1, type-2 and the sort-3 is gestational diabetes. Growth of Insulin resistance ends in Type-2 diabetes. There is a document of 415 patients inside the facts set and this is to be taken for a diverse purpose; the information is collected by different regions. Model is evolved by means of the MATLAB with SQL server ninety-five % of the best forecast is completed by Naive Bayes.

Rajeswari et al. [8] analyze liver disease by using mining algorithms of K star, NB, and forest tree. The data set of patients is considered from the ICU which includes 345 cases and seven characteristics. By using WEKA tool 10 cross-validations can be tested. Naive Bayes provides a precision of 96.52% in 0 s. The accuracy of 97.10% is obtained using the FT tree in 00.200 s. The K-star algorithm classifies 83.47% of instances with precision in 0 s. Based on the results, the FT tree offers the highest classification accuracy in the hepatopathy data set compared to other data mining algorithms.

In the paper [9], Aruna et al. compared different classifiers viz. RBF neural networks, NB, the SVM-RBF kernel, the decision trees (J48) and the simple CART over the breast cancer data sets. Sadhana and Sankareswari [10], in their work, compared upon the precision of two classifiers SVM and decision tree for WPBC based on cross-validation. Sivakami and Saraswathi [11] have proposed the prediction of breast cancer using the DT-SVM hybrid model. Other classification algorithms such as SMO, IBL, and NB have also been applied. Among all the other algorithms, the DT-SVM gave the best performance and accurate result.

### 3 Simulation of Existing Model

The SVM, Naïve Bayes, and KNN models over the heart, liver, and diabetes diseases datasets taken from the UCI machine learning repository have been simulated and calculated the Recall, Accuracy, F1\_measure, Precision, and ROC\_Auc Score. These quantitative measurements are used to evaluate and compare the efficiency of different models.

#### 3.1 Simulation Parameters

A True Positive (TP): the symptoms that cause disease is treated as the symptoms that play a role in disease by the system.

True Negative (TN): the symptoms which do not play a role to cause disease is treated as same.

False Positive (FP): the symptoms which not play a role in disease is treated as the symptoms which act in disease prediction.

False Negative (FN): the symptoms which cause disease but were wrongly classified as symptoms which do not cause disease by the system.

1. **Accuracy:** It is the fraction of symptoms that had been efficiently labeled. It is one of the maximum effective and typically used assessment notations.

$$\text{Accuracy} = (TP + TN)/(TN + FN + FP + TP)$$

2. **Recall:** It is the fraction of symptoms (the symptoms that have maximum chances of being wrong) correctly classified by the system.

$$\text{Recall} = TP/(TP + FN)$$

3. **Precision:** It is the number of symptoms either cause disease or not that was correctly classified. It is also known as the detection rate.

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

4. **F1\_measure:** F1\_measure is the harmonic mean among precision and take into account. High precision but decrease take into account, gives you a really accurate; however, it then misses a big variety of times which can be difficult to categorize.

$$\text{F1\_Score} = (2 * \text{precision} * \text{recall})/(\text{precision} + \text{recall})$$

5. **ROC score:** ROC score is the receiver operating characteristic and the graph is plotted against True Positive Rate (TPR) and False Positive Rate (FPR) for different threshold values. ROC\_AUC is just the area under the curve.

### 3.2 Processing

Data preprocessing is an important step in solving every machine learning problem. A large portion of the data should be cleaned, i.e., preprocessed before any machine learning calculation can be done on it. The sections which contain invalid qualities are supplanted with mean estimations of the segment. As unmistakably noticeable from the dataset, every feature except for sex is whole numbers. The feature sex is changed from string to numeric binary values (0 and 1) in this data pre-processing step. Our model is trained on the Naïve Bayes, KNN, and SVM algorithms. These three models have been chosen as they are the best among all, especially for the medical field. Based on the trained model, our system is tested by taking the input and the model predict the result according to the inputted value whether the person is suffering from the disease or not and gives the accuracy ratio that how much sure that the person is suffering in a particular disease.

**Test size: 25%, Training size: 75%** (Table 1)

From the above results, it is clear that the SVM performs the best against all the observed parameters (recall, accuracy, precision, F1\_Measure, and ROC\_AUC) and for all the three diseases datasets. It is also notable that SVM is not costly to train and, therefore, choose the SVM to train our model.

**Table 1** Results for the three diseases

Disease	Algorithm	Recall (%)	Accuracy (%)	F1-Score (%)	Precision (%)	ROC_AUC (%)
Heart	Naïve Bayes	28.57	78.94	42.85	85.71	63.37
	SVM	85.71	94.73	90.00	94.73	91.94
	KNN	71.42	78.94	65.21	60.00	76.62
Liver	Naïve Bayes	100.00	70.54	82.73	70.54	50.00
	SVM	100.00	70.54	82.73	70.54	50.00
	KNN	77.66	60.95	73.73	70.17	49.30
Diabetes	Naïve Bayes	36.12	65.54	42.73	61.54	50.00
	SVM	71.72	78.64	69.73	71.54	73.03
	KNN	69.66	71.95	54.73	64.17	49.30

## 4 Operational Analysis

A prototype of human disease diagnosis medical expert system has been developed to support the diagnosis of common diseases. This will help the physician in diagnosing a few ailments. The fundamental goal of this framework is to deliver pertinent information and data for discussions, and with the outcomes got at this stage, produce potential judgments. The dataset of the liver patients and the heart patients is merged and make a new dataset (mixed dataset) of the patients. Afterwards, the preprocessing in the dataset is done for better computation and protect our system by predicting the wrong result and replacing the missing value with encoding categorical variables, scaling, mean value, etc. The framework undergoes the following steps:

*Step 1. Import the necessary libraries (pandas, numpy, MultinomialNB, KNeighborsClassifier, SVC, make\_pipeline, confusion\_matrix)*

*Step 2. Fetch the dataset*

*Step 3. Apply pre-processing steps on the dataset*

*Step 4. Split the dataset into training and testing*

*Step 5. Build the model by applying different algorithms*

*Step 6. Test the model*

*Step 7. If the result (score) is efficient then retain this model and go to step 8; otherwise, goto step 3 to try some other model.*

*Step 8. Take the input from the user and predict the result*

## 5 Conclusion and Future Scope

This document exhibits a savvy malady analysis framework for the determination of ailments utilizing the past calculation. In this venture, strategies have been proposed for diagnosing liver, heart, and diabetes illness in patients utilizing machine learning procedures. The three machine learning procedures that were utilized incorporate SVM, logistic regression, and KNN. The framework was executed utilizing one of the models at a time and their exhibition was assessed. Execution assessment depended on certain exhibition measurements. A few properties of this model stay to be explored. It ought to be tried on a few datasets. Tragically, it is difficult to gather the enormous measure of the information and the information is exclusive and hard to get.

The machine learning is a kind of savage power component which endeavors to discover the connection between the numerical properties of contributions with coordinating yields dependent on the past information. At the end of the day, there is no reasonable calculation that can be so useful for utilizing in malady forecast as there is increasingly named information. So starting at now, there exists a few constraints notwithstanding for machine learning calculations. As future work, pre-defined-controlled vocabularies in the form of ontologies could be utilized for this purpose.

## References

1. Domingos P, Pazzani M (1997) On the optimality of the simple Bayesian classifier under zero-one loss. *Mach Learn* 29(2–3):103–130
2. Schiff ER, Maddrey WC, Sorrell MF (2011) Schiff's diseases of the liver. Wiley
3. Ephzibah EP (2011) Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis. arXiv preprint [arXiv:1103.0087](https://arxiv.org/abs/1103.0087)
4. Vembandasamy K, Sasipriya R, Deepa E (2015) Heart diseases detection using Naive Bayes algorithm. *Int J Innov Sci Eng Technol* 2(9):441–444
5. Parthiban G, Srivatsa SK (2012) Applying machine learning methods in diagnosing heart disease for diabetic patients. *Int J Appl Inf Syst (IJAIS)* 3(7):25–30
6. Tan KC, Teoh EJ, Yu Q, Goh KC (2009) A hybrid evolutionary algorithm for attribute selection in data mining. *Expert Syst Appl* 36(4):8616–8630
7. Sarwar A, Sharma V (2012) Intelligent Naïve Bayes approach to diagnose diabetes Type-2. *Int J Comput Appl Challenges Netw Intell Comput Technol* 3:14–16
8. Rajeswari P, Reena GS (2010) Analysis of Liver Disorder Using Data Mining Algorithm. *Glob J Comput Sci Technol* 10:48–52
9. Aruna S, Rajagopalan SP, Nandakishore LV (2011) Knowledge based analysis of various statistical tools in detecting breast cancer. *Comput Sci Inf Technol* 2(2011):37–45
10. Sadhana M, Sankareswari A, MCA M (2014) A proportional learning of classifiers using breast cancer datasets. *Int J Comput Sci Mobile Comput* 3(11):223–232
11. Sivakami K, Saraswathi N (2015) Mining big data: breast cancer prediction using DT-SVM hybrid model. *Int J Sci Eng Appl Sci (IJSEAS)* 1(5):418–429

# Soccer Result Prediction Using Deep Learning and Neural Networks



Sarika Jain, Ekansh Tiwari, and Prasanjit Sardar

**Abstract** In the present world, the prediction of the results of football matches is being done by both machines and football experts. Football as a game produces a huge amount of statistical data about the players of the team; the matches played between the teams and the environment in which the match is being played. This statistical data can be exploited using various machine learning techniques to predict various information related to a particular football match, namely the result of a particular game, injury of a player, performance of a player in a particular match, and spotting new talents in the game, etc. In this work, previous works are reviewed on the prediction of the outcome of a football match, evaluate the merits and demerits of different approaches and then attempt to design a prediction system powered by Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTMs).

**Keywords** Neural network · Machine learning · Deep learning · Prediction · RNN · LSTM

## 1 Introduction

Football being one of the most popular sports around the globe has a huge amount of fans following the day to day events in the game. The investment involved in football matches is increasing in billions by the end of every season and every team wants to get the best result out of the investments they have made. Multiple big teams and sports organizations depend on data available from previous meetings of the teams as well as the present condition of the team to make proper adjustments to their squads to win the competitive matches with the best results. The people associated with the football teams as well as the followers of the teams often come across instances, where it is hard to guess how their team will perform in a particular game; this is

---

S. Jain (✉) · E. Tiwari · P. Sardar  
National Institute of Technology, Kurukshetra, Haryana, India  
e-mail: [jasarika@nitkkr.ac.in](mailto:jasarika@nitkkr.ac.in)



where game result prediction systems come into play. Soccer score prediction can be a helpful means to assess the readiness of a team before going into a game so that necessary changes can be made to the team to avoid a defeat. The prediction can also help the management of different football associations in getting their teams ready for the upcoming matches to get the best result possible out of the fixture. Accurate football match outcome prediction is valuable for small and big sports telecommunication companies as it can help them increase their revenue and make the game a lot more interesting for the viewers. They can also be used for finding new talents in the sport and making better investments in the teams' future and better returns for the investments.

There have been several attempts at trying to predict the outcome of a football match but none of them has been able to match the accuracy of the human prediction. Humans are still by far superior in predicting the outcome of the match as they take into account the technicality as well as the emotional factors that affect the outcome of the game; this helps them in predicting well but it has its drawbacks as well. Humans let their emotions overpower them in the prediction which more or less leads to a wrong prediction of the outcome. Various factors affect the outcome of a football match such as the number of scored goals in previous matches by a team, the winning streak that the team is coming to play within the present game, the environment in which the team is playing the current fixture, and many more other factors associated with the current form of the players in the team as well as the current form of the team itself. Not only the outcome of a game but also various other parameters can also be predicted like the performance of a player, injury of a player, and evaluating a game strategy.

This work aims to review previous efforts on the prediction of results of a football match, to evaluate the merits and demerits of different approaches and then attempt to design a prediction system powered by Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTMs) to predict the result of a football match. Increasing the accuracy of the prediction is aimed at by taking into consideration all the events that take place during a football match and their effects on the outcome of the match.

## **2 Background and Related Work**

This section familiarizes the reader with the theory that works behind the proper functioning of the result prediction done by the RNN using LSTMs. The section starts with machine learning, deep learning, and neural networks and concludes with RNN and LSTMs. Most of the techniques aims at reducing the randomness with the game.

## 2.1 Machine Learning

Machine learning is a subfield of artificial intelligence whose popularity is growing in the field of computers at a very high rate. It is the field of study that helps a system learn on its own without any explicit intervention. Majorly machine learning is of three types supervised, unsupervised, and reinforcement learning. In *supervised learning*, the system searches for patterns and analyses the data to train itself for different scenarios according to the data provided. Once the system is well trained and starts giving the desired level of accuracy in the output in the training procedure; it is deployed on real-life data. In *unsupervised learning*, the algorithm is not provided with any structure of the data, and it is expected to find a pattern in the given data by itself. *Reinforcement learning* is the ability of the system to determine what is the best outcome taking into consideration the environment.

## 2.2 Neural Networks

The neural network works exactly like the working of the human brain. The human brain consists of nerve cells that are connected by means of axons. An artificial neural network can be considered as a pool of processing units that are connected with weighted links that process the data among themselves. It consists of nodes which are connected by links that send the data processed by one node to the other node. Each node executes a simple task to help the system learn something from the input fed to the system. The core of the neural systems, i.e. the neurons (nodes) is just simple activation functions that take multiple input and process it to give a particular output.

## 2.3 Deep Learning

Deep learning is a branch of learning that makes use of deep neural networks. They can be used to implement great real-world problems which involve complex hierarchy in themselves. While deep learning takes longer time to train the system, it gives a more accurate result after the processing and training of the data is done. One benefit of deep learning is that it can work on a very large dataset with different layers and hence give a better result as compared to other learning mechanisms like SVM and random forest.

## **2.4 Recurrent Neural Networks (RNNs)**

Neural networks have a disadvantage associated with them; that is, the input and output of the network is of fixed length. Recurrent neural networks try to overcome this drawback of ANNs. They allow loops between different nodes to allow the system to remember specific outputs related to some particular features. These loops act like memorizers which remember the data for some time in the system.

## **2.5 Long Short Term Memory (LSTMs)**

LSTMs are units of recurrent neural networks that are designed to remember the values for a long or a short period of time. If the LSTM feels that a feature is important in deciding the output, then it carries the feature in the memory for a longer period of time. The recurrent units in LSTM do not use activation function at every node so the data does not get squashed by the system by getting processed repeatedly.

## **2.6 Comparative Study**

Zdravevski and Kulakov employed classification techniques that can be found in WEKA to predict the winner of a game. Parameters that were used were, namely the no. of players injured in the team, winning streak of the team before playing the game, the fatigue of the team before playing the game, win percentage of the teams, and offensive and defensive ratings of the team [1]. Lam in their research has used a Bayesian linear regression model, Gaussian process regression model, and Sparse spectrum Gaussian process regression model to calculate the winning probabilities of a match. They use 11 attributes in total to infer the team's strength and the player's ability [2]. Igiri in their research used Gaussian combinational kernel type and generated a total of 79 support vectors at a count of 100,000 iterations [3]. Rotshtein et al. in their research propose a model which predicts the result of an upcoming match using the method of identifying non-linear dependencies in the knowledge base. Five levels were defined to construct a prediction model based on fuzzy logic, namely loss with a high score, loss with a low score, game with a draw, win with a low score, and win with a high score. To apply the fuzzy knowledge base a fuzzy knowledge approximator was used [4]. Pettersson and Nyquist used LSTMs in recurrent neural networks to predict the outcome of an ongoing football match. They considered all the players playing in the team to be at the same advantage and of the same capabilities [5]. In their system, Cui et al. [6] used 25 different variables related to a football game as terminal sets, and the fitness measure is set as the total number of the wrong prediction of the games.

**Datasets:** While a great deal of data is available related to football games across the globe, the data is not always useful or in accordance with the requirements of the researchers. Many times the researchers have to gather the data manually or through various sports companies or site to work upon. Zdravevski and Kulakov [1] used the data of two back to back seasons of National Basketball Association League from the basketball reference website. Max Lam [2] used the data of the 2014/15 season of NBA from the database available on the NBA site [7]. Igiri [3] in their research used the data available for all the seasons of the English Premier League [8] as their dataset to work upon. Rotshtein et al. [4] used the data of the championship of Finland. The dataset in Pettersson and Nyquist [5] included matches from leagues and tournaments across the globe from 54 different countries that come under the Union of European Football Associations (UEFA) [9] as well as the countries of American and Asian origin. Cui et al. [6] used the official site of the English Premier League [8]. Table 1 summarizes the just discussed works, i.e., references [1–6].

**Table 1** Comparing approaches

Author	Accuracy	Remarks
Zdravevski and Kulakov [1]	The accuracy of reference classifier was comparatively low when compared to any other classifier (around 5-10% low)	The data was collected from the official NBA source, and hence, the data was more accurate. Low prediction accuracy
Lam [2]	TLGProb had an accuracy of 85.28% in NBA 2014/2015 season	TLGProb performs well when compared to the existing NBA predictive models. A small dataset was used
Igiri [3]	Prediction accuracy was 53.3%	Can handle nominal data. Low prediction accuracy. Does not support large datasets
Rotshtein et al. [4]	85% accuracy in genetic tuning and 84% accuracy in neural tuning	High accuracy. Due to insufficient learning samples, the tuning of this kind of system is a very difficult task
Pettersson and Nyquist [5]	The accuracy was 96.63% for many to one approach and 88.68% for the many-to-many approach	High accuracy. The dataset comprised of only 2–3 seasons of the leagues and tournaments
Cui et al. [6]	Testing accuracy was 56%, and the overall accuracy of the system was near to 70%	Low accuracy. The accuracy rate of each function is not the same

## 2.7 Discussion

Taking into account the popularity of the game in the present world, many organizations will be willing to invest a huge amount of money in the prediction systems for the better performance of their teams. As interesting as it may seem, prediction of the results of a football game is a very hard task and involves a large amount of uncertainty. However, it can be said that the result of football is not a completely random event, and hence, a few hidden patterns in the game can be utilized to predict the outcome. Based on the studies of numerous researchers that is being reviewed in our study as well as of those done in the previous years, one can say that with a sufficient amount of data an accurate prediction system can be built using various machine learning algorithms. While each algorithm has its advantages and disadvantages, a hybrid system that consists of more than one algorithm can be made that can increase the efficiency of the system as a whole. There also is a need for a comprehensive dataset through which better results can be obtained. Experts can work more toward gathering data related to different leagues and championships across the globe which may help in better understanding of the prediction system. Moreover, the different features of a footballer, as well as that of the team, can also be taken into consideration while predicting as this may produce a better result as compared to when all the players in a game are treated to be having an equal effect on the game.

RNNs gave a better result when compared to the other techniques for predicting the outcome of a football match. The more information the system was fed the more accurate the result of the RNN system showed. Hence, RNN as the technique to be used in our prediction system is chosen.

## 3 Operational Analysis

Previous works on predicting the results of football matches with machine learning techniques have mainly focused on the data available about the team. The focus has only been limited to a small number of leagues as well. The approach of this work is to deduce better features from the results of the previous matches that the team has played and taken into consideration the current form of the team. A long short term memory system is exploited for this work.

### 3.1 Dataset

The dataset for the project has been taken from “<http://football-data.co.uk/data.php>.” The dataset contains the data of the seasons from 2010–11 to 2017–18 of the English Premier League. The advantage associated with this dataset is that the number of

matches played by each team, and the total number of matches in the tournament is fixed. This was beneficial in removing unnecessary information from the system. In each dataset, there are around 380 records each which are in chronological order with around 60 attributes. From the given dataset, manual feature generation was done to get new attributes. For example, the values from the "Full Time Away Goals" (FTAG) and "Full Time Home Goals" (FTHG), two new attributes "Home Team Goals Scored" (HTGS), and "Away Team Goals Scored" (ATGS) were calculated by summing over the attributes through all the rows. All the contenders in every match had their form calculated by finding the winning streaks of the teams and recording if any team had won five games in a row or three games in a row. Finally, one hot encoding was done for classification purpose. The various rows that were considered in the final dataset are given below:

<i>HTGS—Goal scored by the home team</i>	<i>HTGC—Goal conceded by the home team</i>
<i>ATGS—Goal scored by the away team</i>	<i>ATGC—Goal conceded by the away team</i>
<i>HTGD—Goal difference of the home team (HTGS—HTGC)</i>	<i>HMI(—4) —Results of the last four games won by the home team.</i>
<i>ATGD—Goal difference of the away team(ATGS—ATGC)</i>	<i>AMI(—4)— Results of the last four games won by the away team</i>
<i>HiLossStreak3—loss streak of the home team with three games (Hat trick)</i>	<i>HiWinStreak3 —win streak of the home team with three games (Hat trick)</i>
<i>AtLossStreak3—loss streak of away team with three games (Hat trick)</i>	<i>AtWinStreak3—win streak of the away team with three games (Hat trick)</i>
<i>HiWinStreak5—win streak of home team with five games</i>	<i>HiLossStreak5—loss streak of the home team with five games</i>
<i>AtWinStreak5—win streak of away team with five games</i>	<i>AtLossStreak5—loss streak of away team with five games</i>
<i>HTP—Points gained by the home team</i>	<i>HTGDGoal difference of the home team</i>
<i>ATP—Points gained by the away team</i>	<i>ATGD—goal difference of the away team</i>
<i>DiffPts-(HTP—ATP)</i>	<i>DiffFormPts— HTFormPts—ATFormPts</i>

### 3.2 Methodology

There are a few parameters in every neural network that won't change throughout the working of the model even when the calculation of the accuracy of the model is done. Such parameters are known as hyper-parameters. Once the model has made use of a certain set of hyper-parameters these parameters can be played around with to get a better understanding and accuracy of the model. RNNs can not only remember the sequence in one instance, but also the sequence of the various instances coming in because the nodes storing the weights, and the activation functions are the same. With respect to football match outcome prediction, the dataset is a relational database.

Also, the order of the various columns is unimportant as it doesn't affect the outcome of the match.

LSTM cells are a modification of RNNs. LSTM cells implement RNNs, and with that, they perform an additional calculation with previous output from the RNN cell and the new input to generate the new output. LSTM cells are mainly composed of three gates:

- Forget gate: This gate is known as the sigmoid function. It is responsible for forgetting information that is not required anymore by the system when moving to the next sequence.
- Input gate: This gate works as an input provider to the network. It uses tanh function as well as the sigmoid function.
- Output gate: This gate is responsible for giving the output that is calculated by the processing that is done at the previous two gates.

The given model is implemented with the help of TensorFlow library in python. All the data in this library is processed in the form of an array. A code written in Tensorflow works in two steps. Initially, a computational graph is formed, where all the connections among various layers of the neural network and the calculations of the weights are done but none of these variables has values. They are initialized with Tensorflow objects without any numerical values. The second step is the execution, where a session function is run, and all the variables that have to be calculated are passed as parameters. This generates numerical values that are assigned to the variables.

### ***3.3 Implementation***

The code for the implementation of LSTM and RNN for the prediction can be seen below. The hyper-parameters include:

- (1) n classes—total number of output classes.
- (2) batch size—number of rows fed in one iteration to the model.
- (3) hm epochs—number of epochs the model runs for.
- (4) chunk size—number of attributes in each chunk.
- (5) n chunks—number of chunks in one instance of data.
- (6) rnn size—number of LSTM cells in the hidden layer of the RNN.

*Pseudocode*

***Step 1: import the necessary libraries***

```
import tensorflow as tf

from tensorflow.contrib import rnn

import pandas as pd

Step 2: import the one hot dataset using pandas

Step 3: define the hyper-parameters

hm_epochs = 10

n_classes = 2

batch_size = 1

chunk_size = 27

n_chunks = 1

rnn_size = 512
```

**Step 4:** Create the RNN model with LSTM cell.

**Step 5:** Feed the system with data and LSTM cell mechanism.

**Step 6:** Train the model using the function created in step 4.

### 3.4 Results and Discussion

The experiment has been performed on the above dataset using a basic LSTM cell from the TensorFlow library. The model was run on a various set of hyper-parameters to find out the best model. Table 2 shows the various train and test accuracies with chunk size = 3 and n chunks = 9, for batch sizes from 1 to 124. The model with chunk size as 27 was considered for further analysis as depicted in Table 3.

**Table 2** RNN with chunk size 3 and n chunks as 9 and varying batch size

Batch size	1	30	60	124
Accuracy train	0.922043	0.6973118	0.6655914	0.6704301
Accuracy test	0.79875	0.6525	0.64625	0.6525

**Table 3** RNN with chunk size 27 and varying batch size

Batch size	1	30	60	124
Accuracy train	0.9811828	0.7688172	0.74139786	0.7354839
Accuracy test	0.8075	0.69875	0.69375	0.69



**Table 4** RNN with different sizes of the hidden layer

RNN Size	32	64	256	512
Accuracy train	0.9731183	0.97043014	0.9919355	0.9892473
Accuracy test	0.8125	0.805	0.805	0.8025

Table 4 shows the accuracies of the model with varying size of the hidden layer. Here, for increasing RNN size, though the training accuracy is increasing, there is no sufficient increase in the test accuracy. This suggests that the model is overfitting the data. Hence, for the present dataset, the best model values are as follows:

<i>n classes—2</i>	<i>batch size—1</i>	<i>hm epochs—10</i>
<i>chunk size—27</i>	<i>n chunks—1</i>	<i>rnn size—64</i>

## 4 Conclusion

The popularity and international effect that football has made it an interesting problem to solve. Moreover, the number of factors that affect the outcome of a match is enormous. From the results can be said that RNNs with LSTM show a visible and obvious advantage over the original ANN and the traditional machine learning. Hence, other than the mainstream uses of LSTMs, this particular path of using it for prediction of the outcome for sporting events has also shown promising results. This model can still be improved. One way to do that is, by using a better set of attributes. They can include statistics of each player. This can also help in predicting the form of a particular player from season to season.

## References

1. Zdravevski E, Kulakov A (2009) System for prediction of the winner in a sports game. In: International conference on ICT innovations, pp 55–63. Springer, Berlin, Heidelberg
2. Lam MW (2018) One-match-ahead forecasting in two-team sports with stacked bayesian regressions. *J Artif Intell Soft Comput Res* 8(3):159–171
3. Igiri (2015) Support Vector machine—based prediction system for a football match result. *IOSR J Comput Eng (IOSR-JCE)* 17(3):21–26
4. Rotshtein AP, Posner M, Rakityanskaya AB (2005) Football predictions based on a fuzzy model with genetic and neural tuning. *Cybern Syst Anal* 41(4):619–630
5. Pettersson D, Nyquist R (2017). Football match prediction using deep learning. Doctoral dissertation, Master's thesis, Chalmers University of Technology
6. Cui T, Li J, Woodward JR, Parkes AJ (2013) An ensemble based genetic programming system to predict English football premier league games. In: 2013 IEEE conference on evolving and adaptive intelligent systems (EAIS), pp 138–143. IEEE

7. <https://in.nba.com/?gr=www>
8. <https://www.premierleague.com/>
9. <https://www.uefa.com/>

# Ranking Diabetic Mellitus Using Improved PROMETHEE Hesitant Fuzzy for Healthcare Systems



K. R. Sekar, S. Yogapriya, N. Senthil Anand, and V. Venkataraman

**Abstract** In recent years, healthcare applications have been gaining market in the research area. The objective of the work is based on mellitus diabetes and seeks to sort people in such a way that people who have affected most will be handled with high preference, moderate attack people will be the next level, and mild attack people will be given less priority. The methodology applied here is hesitant PROMOTHEE fuzzy in soft computing which helps to find the outranking and the best selection. The preference function is to be defined in PROMOTHEE to find the outranking decisions for the alternatives using criteria. This method is more powerful because preference can be given to the alternatives using their criteria. The rows are always treated as alternatives and columns are considered as criteria. Dominance and sub-dominance are the factors that play a vital role in the methodology described. The outcome of the research work focuses on giving the high peace of treatment to the most affected people and to safeguard their lives from the sudden mortality. The selection and the ranking have been given the foremost preference and priority in the research work with high precision of PROMOTHEE calculations in the healthcare application.

**Keywords** Mellitus diabetes · Hesitant PROMOTHEE fuzzy · Dominance · Sub-dominance · Preference function

## 1 Introduction

Diabetic mellitus occurs due to excessive growth hormone production by the pituitary gland due to a pituitary adenoma. It comes in different types and stages and

---

K. R. Sekar · S. Yogapriya · N. Senthil Anand  
School of Computing, SASTRA Deemed University, Thanjavur, India

V. Venkataraman (✉)  
School of Arts Science and Humanities, SASTRA Deemed University, Thanjavur, India  
e-mail: [mathvvr@maths.sastra.edu](mailto:mathvvr@maths.sastra.edu)

has the ability to metastasize and infect the excretory system. Therefore, early diagnosis and treatment should be essential to alleviate this disease. This study aims to apply FUZZY preference ranking organization method for enrichment of evaluation (PROMETHEE) in the analysis of the techniques for diabetes treatment where we were able to achieve reliable results that can aid physicians and patients to embark on a favorable treatment option. Decision-Making Trial and Evaluation Laboratory Method (DEMATEL) are separately applied to derive the risk criterion and decision-makers' notions. Diabetic mellitus is a health problem which may occur if the diabetic is uncontrolled. High blood sugar will still affect the body and offers a certain lethargic feeling during the whole day. Diabetic is always a silent killer for all of the patients affected by this disease. Physicians always monitor the patient blood sugar level before treating with good medications. The history of the patient must be studied in detail by the physicians before delivering the appropriate treatment. This is also highly appreciated and valued, and the same must be followed by the physicians inshore and offshore. With the diabetic mellitus, patients are sorted by the physicians based on three thresholds using the patient's preferences. PROMOTHEE concept helps in finding out the good ranking methodology which is useful to exercise their medications. In the case of existing models, the decision-makers find it difficult to obtain a clear view of the problem and to evaluate the results. PROMETHEE is one of the fuzzy techniques which set out guidelines for weighing the attributes according to their preferences. Typically, decision-makers can provide a rational linguistics and intuitionism for the given data set which may be the reason why PROMOTHEE ends up with a good outranking framework. The hesitant fuzzy has alternatives and preferred criteria with respective membership weightage. The objectives and the fuzzy weightage for the attributes provide the good précised answer for outranking in PROMOTHEE methodology.

## 2 Relative Work

In this paper, a best-worst PROMETHEE method is deployed to rank schools available in the program for international student assessment [PISA] to guide the decision-makers to take the right decision [1]. In the paper, a failure mode and result analysis (FMEA), preference ranking organization technique for enrichment evaluation [PROMETHEE] approach and also the technique for order preferences in ideal solution [TOPSIS] are used to derive the risk factors for decision-makers [2]. In the paper, this case study emphasizes individualized shopper fulfillment, as per personal interests by using the neutrosophic PROMETHEE methodology utilized in matched selling [3]. In this paper, a decision framework is leveraged for the site selection in offshore wind power station, wherein PROMETHEE method is used for deploying the assumption and methodological support for site selection and decision-making in the proposed coastal wind power project to proceed further and enhance the benefits of the proposed integrated coastal management [4]. In another research work, the preference ranking organization method for enrichment evaluation (PROMETHEE)

method is applied to perform standard evaluation of hospital service by considering the patient's case study in a public hospital, and further, the resultant outputs are compared with the TOPSIS services [5]. In the next research work, a Pythagorean fuzzy-based PROMETHEE method is utilized for validating its applicability and feasibility in incorporating a multi-criteria decision-making challenge associated with the bridge-superstructure construction methods [6]. Consecutively, in the next paper, the parabolic trough concentrating solar power plant (PT-CSPP) and fuzzy-driven PROMETHEE II method is used to generate electricity and tackle the sharp increase in the finite amount of fossil fuel and the general energy requirements to quantify the eligible alternatives [7]. Followed by this, the next research work is primarily focused on selecting the location for incorporating a military airport using multi-criteria decision-making methodologies by integrating AHP, ranking and selection processes that are deployed by using PROMETHEE preference ranking approach [8]. The paper aims to find the destinations and measure the address of the competitiveness of tourism destinations at the regional level by using PROMETHEE method [9]. Finally, [10] has developed a fuzzy-driven PROMETHEE approach to perform material selection and further the ranking and selection processes are deployed using PROMETHEE method for validating and comparing the corresponding criteria.

In this paper, PROMETHEE is used for selecting the seal strips used in turbines to select the most appropriate stainless steel material [11]. In the next paper, PROMETHEE is employed for enhancing mobile application recommendations by reducing the cognitive efforts of users to a minimal level [12], whereas in the next paper, a unicriterion analysis based on the PROMETHEE principles is performed to deploy multi-criteria ordered clustering [13]. Next, this research paper leverages renewable energy sources accompanied with PROMETHEE II method is utilized here for analyzing the climate change mitigation, safeguard energy security and strengthens environmental protection [14]. Here, in this paper, an evaluation is performed on the Web sites of hotels, and PROMOTHEE and GAIA methods are utilized here to deploy a visual assistance for such Web sites [15]. This paper aims to determine four significant perspectives, namely societal, economical, technological and environmental perspectives to meet the increasing societal demands and development by using the PROMETHEE method [16]. In this paper, the PROMETHEE II method and ArcGIS software are used to rank the particular sub-watershed by incorporating the AHP method [17]. In this paper, PROMETHEE method is used to sort out the schools according to their academic performance and quality [18]. In this case study, PROMETHEE-driven decision-making methods are used to select the appropriate equipment to build construction that remains as a challenging task [20].

### 3 Methods

*Step 1:* Attributes are considered here as a criterion  $j = 1, 2, \dots, m$  for the **alternatives available in the rows for the decision-making problems**

*Step 2:* **Weightage of each and every criterion can be determined through**  
 $\sum_{j=1}^k w_j^*$

*Step 3:* 0-1 mapping for the decision-making problems can be obtained through normalization

$$\bar{R}_{ij} = [\bar{X}_{ij} - \text{Minimum}(\bar{X}_{ij})][\text{Maximum}(\bar{X}_{ij}) - \text{Minimum}(\bar{X}_{ij})] \quad (1)$$

For the decision-makers  $i = 1, 2, \dots, n$  can be found through the evaluation variable  $\bar{X}_{ij}$  where  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, m$ .

*Step 4:* **Pairwise comparison of the alternatives can be formed by the notation**

$$\bar{d}_j(a, b) = \bar{g}_j(a) - \bar{g}_j(b) \quad \bar{d}_j(a, b) = \bar{g}_j(a) - \bar{g}_j(b) \quad (2)$$

where  $\bar{d}_j(a, b)$  represents evaluation difference between  $a$  and  $b$  for the criteria in each tuple.

*Step 5:* Preference function can be defined as  $\bar{P}_j(a, b) = F[\bar{d}_j(a, b)]$  where  $\bar{P}_j(a, b)$  express the difference between  $a$  and  $b$  for the evaluation of alternatives in degree ranging 0–1. If the difference between the alternatives in the negative side is similar, on the other side positive value represents no similarity in the alternatives.

*Step 6:* Multi-criteria preference can be obtained through the formulae

$$\Pi(a, b) = \sum_{j=1}^k \bar{P}_j(a, b) \bar{w}_j \quad \pi(a, b) = \sum_{j=1}^k \bar{P}_j(a, b) \bar{w}_j \quad (3)$$

where  $\bar{w}_j > 0$  are the weights related to each criterion in the column.

$\pi(a, b)$  adheres that the degree preferred by  $a$  and  $b$  in the overall criteria. The numeric value of  $\pi(a, b) = 0$  refers weak preference of  $a$  over  $b$  and 1 represents a stronger preference.

*Step 7:* Determined the preference order.

In PROMETHEE I, partial ranking can be possible because of their low precision values, and in the PROMETHEE II, the complete solution of the ranking is possible because of the high precision values.

- a. Partial ranking obtained only through PROMETHEE I.

$$\begin{aligned}
 a\overline{\overline{P}} + b &: \left\{ \overline{\overline{P}} \text{ iff } \overline{\overline{\varphi}}^+(a) > \overline{\overline{\varphi}}^+(b), \text{ for every } a, b \in A \right. \\
 I & \text{ iff } \left\{ \overline{\overline{\varphi}}^+(a) = \overline{\overline{\varphi}}^+(b), \text{ for every } a, b \in A \right. \\
 a\overline{\overline{P}} - b &: \left\{ \overline{\overline{P}} \text{ iff } \overline{\overline{\varphi}}^-(a) < \overline{\overline{\varphi}}^-(b), \text{ for every } a, b \in A \right. \\
 I & \text{ iff } \left\{ \overline{\overline{\varphi}}^-(a) > \overline{\overline{\varphi}}^-(b), \text{ for every } a, b \in A \right. \tag{4}
 \end{aligned}$$

$\overline{\overline{\varphi}}^+(a)$  denotes +ve outranking flow or outgoing flow (how ‘a’ literally predominates all the alternative tuples).

$\overline{\overline{\varphi}}^-(a)$  denotes -ve outward flow or incoming flow (how ‘a’ predominates all the alternative tuples).

$\overline{\overline{\varphi}}^+(a)$  and  $\overline{\overline{\varphi}}^-(a)$  represent the preference relation for ranking alternatives.

Where former denotes partial ranking and the later denotes complete ranking without any errors and precision difference.

$$\begin{aligned}
 a\overline{\overline{P}} + b &: \left\{ \overline{\overline{P}} \text{ iff } \overline{\overline{\varphi}}^+(a) > \overline{\overline{\varphi}}^+(b), \text{ for every } a, b \in A \right. \\
 I & \text{ iff } \left\{ \overline{\overline{\varphi}}^+(a) = \overline{\overline{\varphi}}^+(b), \text{ for every } a, b \in A \right. \\
 a\overline{\overline{P}} - b &: \left\{ \overline{\overline{P}} \text{ iff } \overline{\overline{\varphi}}^-(a) < \overline{\overline{\varphi}}^-(b), \text{ for every } a, b \in A \right. \\
 I & \text{ iff } \left\{ \overline{\overline{\varphi}}^-(a) > \overline{\overline{\varphi}}^-(b), \text{ for every } a, b \in A \right.
 \end{aligned}$$

Compute the resultant outgoing flow by the following steps.

- b. PROMETHEE II provides a complete ranking among the alternatives using their preference criterion and avoids annoying comparison.

$$\overline{\overline{\varphi}}(a) = \overline{\overline{\varphi}}^+(a) - \overline{\overline{\varphi}}^-(a) \tag{5}$$

where  $\overline{\overline{\varphi}}(a)$  represents the net outranking flow for each alternative.

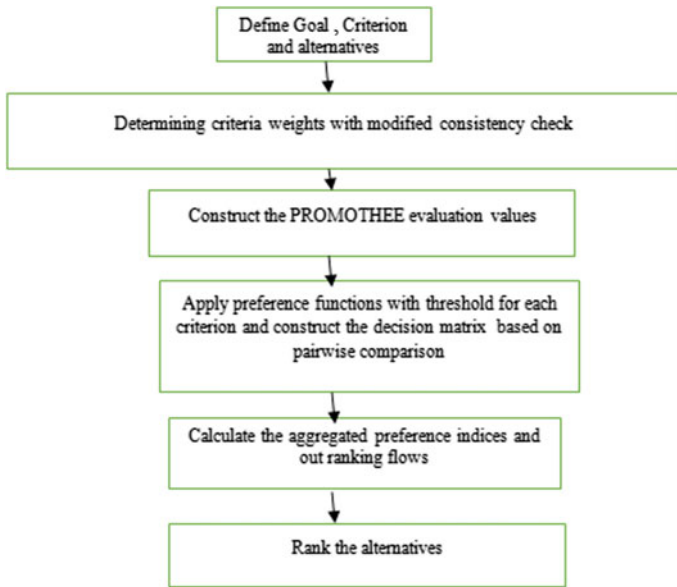
The preference relation functions are as follows:

PROMOTHEE-I is the decision-making method for all types of domains in the research origin. Preference and priorities have to be given to the alternatives according to their criteria. Here, we are considering two factors like dominance and sub-dominance which will provide high mileage and milestone in decision-making systems. Row-wise total of each alternative is referred to as dominance, and the criteria’s total is called as sub-dominance. Pairwise comparisons among

the alternatives will leave the entropy in the semantic table. PROMOTHEE-I will give an extent result rather than partial results.

PROMOTHEE-II is based on quantitative and qualitative measures among the criteria. The outranking decision-making by using this method is always balanced because of comparing positive and negative sides. PROMOTHEE-II is the complete ranking method rather than rational method in decision-making scenarios.

### 3.1 Process Flow



#### Definition 1

Type 1: Let us define the preference criterion function as

$$\bar{P}(x) = \begin{cases} 0; & x \leq 0 \\ 1; & x > 0 \end{cases} \tag{6}$$

here,  $x$  refers the variation between two tuples or alternatives.



In type 1, if  $f(a) = f(b)$ , then indifference will occur. If the decision-makers never give any importance or vital weightage given to the criteria and try to go for pairwise comparisons among the alternatives will end up in erroneous results.

**Definition 2**

Type 2: Let us define the criterion preference function in a linear form as

$$\bar{P}(x) = \begin{cases} 0; & x < 0 \\ x/m; & 0 \leq x \leq m \\ 1; & x > m \end{cases}$$

**Definition 3**

Type 4: Let us define the level criterion function as

$$\bar{P}(x) = \begin{cases} 0; & x \leq q' \\ 1/2; & q' < x \leq q' + p' \\ 1; & x > q' + p' \end{cases}$$

indifference will occur on the interval  $[-q', q']$  where  $q' = 1, 2, \dots, r (<n)$  and  $s = 1, 2, \dots, s (<n)$ . Quantitative criteria can be achieved by type 4.

**4 Results and Discussions**

See Table 1.

**4.1 Illustration Work**

Diabetic symptoms	
1	Frequent urination
2	Fatigue
3	Blurred vision
4	Slow healing
Diabetic patients	

(continued)

(continued)

Diabetic patients	
1	Patient 1
2	Patient 2
3	Patient 3
4	Patient 4
5	Patient 5

In the above-said Sects. 4 and 4.1, the research paper discussed comparison analysis about the relative articles, and it will reveal the application, methods applied and outcome of that work. The later section has got patient 1 to patient 5 and has alternatives, and diabetic symptoms are the criterion for the research scenario. Using diabetic patients and diabetic symptoms were deployed in the illustration work.

Table 2 discusses linguistics values provided by the decision-maker rather than domain expert in the particular field. Normally, linguistics refers to a low, medium, high and very high. Table 2 is the training set for the coming calculation and computation. Each and every linguistic value has got its significance for a particular cell to say the intersection of rows and columns.

In Table 3, alternatives and criteria are having intuitionistic values to linguistic values, which is available in Table 2. Intuitionistic values are 1, 5, 7 and 9, and this is referred to as scale 10. So for each and every linguistic value, one corresponding cardinal value is assigned, and it is always unique in nature.

In Table 4, the first objective has been defined with two jargons like beneficial and non-beneficial, and the same will be assigned for the criteria in terms of cost factors toward criteria, and the same will be deployed in Table 5. In Table 5, maximum and minimum values are taken from the column attributes or criteria.

Table 6 refers to the aggregation which is all calculated through pairwise comparisons among the alternatives. The need behind having pairwise comparison is only to know about the similarity between the alternatives and rather between the tuples. If the minus value comes between the pairwise comparison referred to as high similarity and positive values are considered as no similarity.

In Table 7, zeros are assigned for the pairwise comparison cells which are having a negative sign in the previous Table 6. Positive values got right from Table 6 as kept as such without any modification in this table.

In Table 8, the second objective fuzzy weightage has been given in terms of the membership function between 0.0 and 1.0 for the preference criterion. Using the first objective, the second objective has been formulated according to the cost-benefit ratio. So the criterion has got fuzzy weightage values 0.9, 0.3, 0.5 and 0.7, respectively.

In Table 9, the respective fuzzy value of the criterion has been multiplied with each and entries of that column. For example, in the first criterion frequent urination, the fuzzy membership weightage function 0.9 has been multiplied with the first entry (intersection of rows and columns) 0. Similarly, have to continue the operation of each and every respective column.

**Table 1** Comparative analysis

S. No.	Reference	Application	Methods	Aggregation	Outcome
1	#Ref 1	evaluating school performance	Best-Worst PROMETHEE method	Yes	Multi dimensional
2	#Ref 4	Industries	PROMETHEE method	Yes	To derive the risk priority of failure modes
3	#Ref 7	Marketing	neutrosophic PROMETHEE method	Yes	Accurate and more effective
4	#Ref 8	Power station	PROMETHEE method	Yes	Site selection decision-making
5	#Ref 9	Hospital	PROMETHEE method	Yes	Evaluating the standard service
6	#Ref 20	develops a Pythagorean fuzzy	PROMETHEE-method	Yes	Multiple criteria decision making
7	#Ref 13	solar power plant	PROMETHEE method	Yes	Site selection
8	#Ref 16	airport	AHP integrated PROMETHEE and VIKOR method	Yes	Location selection
9	#Ref 19	tourism	the PROMETHEE method	Yes	Destinations and measure the address of the competitiveness of tourism
10	#Ref 31	A fuzzy logic	PROMETHEE method	Yes	Material selection problems
11	#Ref 17	Turbine seal strips	PROMETHEE-GAIA method	Yes	Material selection
12	#Ref 22	mobile application	PROMETHEE	Yes	Ranking methods
13	#Ref 29	A unicriterion analysis	PROMETHEE method	Yes	Multicriteria
14	#Ref 36	Renewable energy sources	PROMETHEE II	Yes	Ranking websites to support market
15	#Ref 37	Hotel	PROMETHEE and GAIA	Yes	Multi criteria analysis
16	#Ref 33	Energy plant site	PROMETHEE method	Yes	Selection and multicriteria
17	#Ref 34	Sub-watersheds	PROMETHEE method	Yes	Multi-criteria decision analysis

(continued)

**Table 1** (continued)

S. No.	Reference	Application	Methods	Aggregation	Outcome
18	#Ref 55	Schools	PROMETHEE method	Yes	Measuring Performance Quality
19	#Ref 62	Water protection	PROMETHEE II method	Yes	evaluation of source water protection
20	#Ref 15	Construction equipment	Multi-criteria Decision Making Methods	Yes	Selection

**Table 2** DM linguistic values

Decision-making	Frequent urination	Fatigue	Blurred vision	Slow healing
P1	Low	Very high	High	Medium
P2	Very high	High	Medium	Low
P3	High	Medium	Low	Very high
P4	Medium	Low	Very high	High
P5	Very high	High	Medium	Low

**Table 3** DM intuitionistic values

Decision-making	Frequent urination	Fatigue	Blurred vision	Slow healing
P1	1	9	7	5
P2	9	7	5	1
P3	7	5	1	9
P4	5	1	9	7
P5	9	7	5	1

**Table 4** Multi-objectives

Frequent urination—Beneficial
Fatigue—Non-beneficial
Blurred vision—Beneficial
Slow healing—Beneficial

Table 10 has been normalized through their column total of each and every column. So that to remove entropy and table disorders rather make them the semantic with all equalities.

Tables 11 and 12 describe the following inference. The two factors like dominance and sub-dominance. Dominance refers to row total, and the sub-dominance is referred to as column total say  $\varphi^+$  Leaving flow and  $\varphi^-$  Entering flow. Leaving flow is the incurring loss, and the entering flow is the capital of the business.

**Table 5** Find maximum and minimum values

Decision-making	Frequent urination (B)	Fatigue (NB)	Blurred vision (B)	Slow healing (B)
P1	1	9	7	5
P2	9	7	5	1
P3	7	5	1	9
P4	5	1	9	7
P5	9	7	5	1
Maximum	9	1	9	9
Minimum	1	9	1	1

Using (3)

**Table 6** Aggregation values

Decision making	Frequent urination	Fatigue	Blurred vision	Slow healing
D(P1-P2)	-8	2	2	4
D(P1-P3)	-6	4	6	-4
D(P1-P4)	-4	8	-2	-2
D(P1-P5)	-8	2	2	4
D(P2-P1)	8	-2	-2	4
D(P2-P3)	2	2	4	-8
D(P2-P4)	4	6	-4	-6
D(P2-P5)	0	0	0	0
D(P3-P1)	6	-4	-6	4
D(P3-P2)	-2	-2	-4	8
D(P3-P4)	2	4	-8	2
D(P3-P5)	-2	-2	-4	8
D(P4-P1)	4	-8	2	2
D(P4-P2)	-4	-6	4	6
D(P4-P3)	-2	-4	8	-2
D(P4-P5)	-4	-6	4	6

Using (2)

In the above-said Table 13, dominance and sub-dominance outcomes like  $\varphi^+$  and  $\varphi^-$  are taken into account rather for calculating the value of  $\varphi$ . Using the outcome values of  $\varphi$ , the ranking has been decided right from top to bottom. Using ranking, the last criterion treatment has been credited in this research work.

**Table 7** Calculate the preference function

Decision-making	Frequent urination	Fatigue	Blurred vision	Slow healing
D(P1-P2)	0	2	2	4
D(P1-P3)	0	4	6	0
D(P1-P4)	0	8	0	0
D(P1-P5)	0	2	2	4
D(P2-P1)	8	0	0	4
D(P2-P3)	2	2	4	0
D(P2-P4)	4	6	0	0
D(P2-P5)	0	0	0	0
D(P3-P1)	6	0	0	4
D(P3-P2)	0	0	0	8
D(P3-P4)	2	4	0	2
D(P3-P5)	0	0	0	8
D(P4-P1)	4	0	2	2
D(P4-P2)	0	0	4	6
D(P4-P3)	0	0	8	0
D(P4-P5)	0	0	4	6

**Table 8** Add weightage for the criteria according to priority

WEIGHTAGE	0.9	0.3	0.5	0.7
Decision making	Frequent urination	Fatigue	Blurred vision	Slow healing
D(P1-P2)	0	2	2	4
D(P1-P3)	0	4	6	0
D(P1-P4)	0	8	0	0
D(P1-P5)	0	2	2	4
D(P2-P1)	8	0	0	4
D(P2-P3)	2	2	4	0
D(P2-P4)	4	6	0	0
D(P2-P5)	0	0	0	0
D(P3-P1)	6	0	0	4
D(P3-P2)	0	0	0	8
D(P3-P4)	2	4	0	2
D(P3-P5)	0	0	0	8
D(P4-P1)	4	0	2	2
D(P4-P2)	0	0	4	6
D(P4-P3)	0	0	8	0
D(P4-P5)	0	0	4	6

**Table 9** Find the total for each column

Weightage	0.9	0.3	0.5	0.7
Decision making	Frequent urination	Fatigue	Blurred vision	Slow healing
D(P1-P2)	0	0.6	1	2.8
D(P1-P3)	0	1.2	3	0
D(P1-P4)	0	2.4	0	0
D(P1-P5)	0	0.6	1	2.8
D(P2-P1)	7.2	0	0	2.8
D(P2-P3)	1.8	0.6	2	0
D(P2-P4)	3.6	1.8	0	0
D(P2-P5)	0	0	0	0
D(P3-P1)	5.4	0	0	2.8
D(P3-P2)	0	0	0	5.6
D(P3-P4)	1.9	1.2	0	1.4
D(P3-P5)	0	0	0	5.6
D(P4-P1)	3.6	0	1	1.4
D(P4-P2)	0	0	2	4.2
D(P4-P3)	0	0	4	0
D(P4-P5)	0	0	2	4.2
Total	23.5	8.4	16	33.6

## 5 Conclusion

In the research work, hesitant PROMOTHEE fuzzy is employed for the healthcare system to find the best selection of severity of the patients to give at most treatment and to provide the greater priority. Outranking is needed for the decision support system using the above-said method.

**Table 10** Calculate the aggregated normalized preference values

Weightage total	23.5	8.4	16	33.6
Decision making	Frequent urination	Fatigue	Blurred vision	Slow healing
D(P1-P2)	0	0.071428571	0.0625	0.083333333
D(P1-P3)	0	0.14285714	0.1875	0
D(P1-P4)	0	0.285714286	0	0
D(P1-P5)	0	0.071428571	0.0625	0.083333333
D(P2-P1)	0.3063829	0	0	0.083333333
D(P2-P3)	0.076595745	0.071428571	0.125	0
D(P2-P4)	0.1531914	0.214857143	0	0

(continued)

**Table 10** (continued)

Weightage total	23.5	8.4	16	33.6
D(P2-P5)	0	0	0	0
D(P3-P1)	0.22978723	0	0	0.083333333
D(P3-P2)	0	0	0	0.166666667
D(P3-P4)	0.0808510	0.1428571	0	0.041666666
D(P3-P5)	0	0	0	0.166666667
D(P4-P1)	0.1531914	0	0.0625	0.041666666
D(P4-P2)	0	0	0.125	0.125
D(P4-P3)	0	0	0.25	0
D(P4-P5)	0	0	0.125	0.125

Using (1)

**Table 11** Find the total for each row

Decision making	Frequent urination	Fatigue	Blurred vision	Slow healing	TOTAL
D(P1-P2)	0	0.071428571	0.0625	0.083333333	0.217261904
D(P1-P3)	0	0.14285714	0.1875	0	0.33035714
D(P1-P4)	0	0.285714286	0	0	0.285714286
D(P1-P5)	0	0.071428571	0.0625	0.083333333	0.217261901
D(P2-P1)	0.3063829	0	0	0.083333333	0.389716233
D(P2-P3)	0.076595745	0.071428571	0.125	0	0.273024316
D(P2-P4)	0.1531914	0.214857143	0	0	0.368048543
D(P2-P5)	0	0	0	0	0
D(P3-P1)	0.22978723	0	0	0.083333333	0.313120563
D(P3-P2)	0	0	0	0.166666667	0.166666667
D(P3-P4)	0.0808510	0.1428571	0	0.041666666	0.18452376
D(P3-P5)	0	0	0	0.166666666	0.166666666
D(P4-P1)	0.1531914	0	0.0625	0.041666667	0.257358067
D(P4-P2)	0	0	0.125	0.125	0.25
D(P4-P3)	0	0	0.25	0	0.25
D(P4-P5)	0	0	0.125	0.125	0.25

The following are the outcomes through the article research.

- (i) Hesitant PROMOTHEE fuzzy is the newer technique for the best ranking and selection.
- (ii) This method removes all the entropies in the semantic training set, so we cannot find any disorder.



**Table 12** Determine the leaving and the entering outranking flows

Aggregated preference function	P1	P2	P3	P4	P5	$\varphi +$ Leaving flow
P1	0	0.217261904	0.33035714	0.285714286	0.217261901	1.050595
P2	0.389716233	0	0.273024316	0.368048543	0	1.030789
P3	0.313120563	0.166666666	0	0.18452376	0.166666666	0.830978
P4	0.257358067	0.25	0.25	0	0.25	1.007358
P5	0	0	0	0	0	0
$\varphi$ —Entering flow	0.960194863	0.633928564	0.853381456	0.838286589	0.633928567	

Using (4)

**Table 13** Out ranking

$\varphi^+$	$\varphi^-$	$\varphi$	Rank	Treatment
1.0506	0.960194864	2.010794864	1	Immediate
1.03079	0.633928564	1.664718564	4	Medication
0.83098	0.853381456	1.684361456	3	immediate
1.00736	0.838286589	1.845646589	2	Immediate
0	0.633928567	0.633928567	5	Medications

Using (5)

- (iii) Alternatives are well established and given preference and priorities according to their criteria.
- (iv) Dominance and sub-dominance are calculated before the pairwise comparisons.
- (v) Pairwise comparisons are the best way for any type of best selection methodology.

Considering the above facts, this PROMOTHEE fuzzy is the best method for the best selection in healthcare systems.

## References

1. Ishizaka A, Resce G (2020) Best-worst PROMETHEE method for evaluating school performance in the OECD’s PISA project. Socio-Econ Plann Sci 15:100799
2. Zhu J, Shuai B, Li G, Chin KS, Wang R (2020) Failure mode and effect analysis using regret theory and PROMETHEE under linguistic neutrosophic context. J Loss Prev Process Ind 104048
3. Zavadskas EK, Bausys R, Kaklauskas A, Raslanas S (2019) Hedonic shopping rent valuation by one-to-one neuromarketing and neutrosophic PROMETHEE method. Appl Soft Comput 1(85):105832

4. Wu Y, Tao Y, Zhang B, Wang S, Xu C, Zhou J (2020) A decision framework of offshore wind power station site selection using a PROMETHEE method under intuitionistic fuzzy environment: a case in China. *Ocean Coastal Manag* 1(184):105016
5. Mutlu M, Kalender ZT, Sennaroglu B, Tuzkaya G Hospital service quality evaluation with IVIF-PROMETHEE and a case study
6. Chen TY (2019) A novel PROMETHEE-based method using a Pythagorean fuzzy combinative distance-based precedence approach to multiple criteria decision making. *Appl Soft Comput* 1(82):105560
7. Wu Y, Zhang B, Wu C, Zhang T, Liu F (2019) Optimal site selection for parabolic trough concentrating solar power plant using extended PROMETHEE method: a case in China. *Renew Energ* 1(143):1910–1927
8. Sennaroglu B, Celebi GV (2018) A military airport location selection by AHP integrated PROMETHEE and VIKOR methods. *Transp Res Part D Transp Environ* 1(59):160–173
9. Lopes AP, Muñoz MM, Alarcón-Urbistondo P (2018) Regional tourism competitiveness using the PROMETHEE approach. *Annals Tourism Res* 1(73):1–3
10. Gul M, Celik E, Gumus AT, Guneri AF (2018) A fuzzy logic based PROMETHEE method for material selection problems. *Beni-Suef Univ J Basic Appl Sci* 7(1):68–79
11. Zindani D, Kumar K (2018) Material selection for turbine seal strips using PROMETHEE-GAIA method. *Mater Today Proc* 5(9):17533–17539
12. Lolli F, Balugani E, Ishizaka A, Gamberini R, Butturi MA, Marinello S, Rimini B (2019) On the elicitation of criteria weights in PROMETHEE-based ranking methods for a mobile application. *Exp Syst Appl* 15(120):217–227
13. Boujelben MA (2017) A unicriterion analysis based on the PROMETHEE principles for multicriteria ordered clustering. *Omega* 1(69):126–140
14. Andreopoulou Z, Koliouka C, Galariotis E, Zopounidis C (2018) Renewable energy sources: using PROMETHEE II for ranking websites to support market opportunities. *Technol Forecasting Soc Change* 1(131):31–37
15. Ostovare M, Shahraki MR (2019) Evaluation of hotel websites using the multicriteria analysis of PROMETHEE and GAIA: evidence from the five-star hotels of Mashhad. *Tourism Manag Perspect* 1(30):107–116
16. Wu Y, Wang J, Hu Y, Ke Y, Li L (2018) An extended TODIM-PROMETHEE method for waste-to-energy plant site selection based on sustainability perspective. *Energy* 1(156):1–6
17. Vulević T, Dragović N (2017) Multi-criteria decision analysis for sub-watersheds ranking via the PROMETHEE method. *Int Soil Water Conserv Res* 5(1):50–55
18. Murat S, Kazan H, Coskun SS (2015) An application for measuring performance quality of schools by using the PROMETHEE multi-criteria decision making method. *Procedia-Soc Behav Sci* 3(195):729–738
19. Kuang H, Kilgour DM, Hipel KW (2015) Grey-based PROMETHEE II with application to evaluation of source water protection strategies. *Inform Sci* 10(294):376–389
20. Temiz I, Calis G (2017) Selection of construction equipment by using multi-criteria decision making methods. *Procedia Eng* 1(196):286–293

# Hybrid Recommendation System for Scientific Literature



Indraneel Amara, K Sai Pranav, and H. R. Mamatha

**Abstract** This paper aims to create a recommendation system for scientific articles. The system can help both researchers and students in finding suitable articles related to their fields of study in a time-efficient manner. This paper proposes a model that uses the DBLP-Citation-network V10 Dataset and a second model that uses a custom dataset created by combining an existing CORA dataset with web scraped information. The core idea of this paper is to use the latent relationships found in the citation data to find more relevant papers. An additional feature provided by this system is that it can scale to any number of new scientific articles. The proposed approach uses various pre-processing techniques and algorithms to calculate scores such as PageRank, TextRank, cosine similarity, author score, etc. Since the content-based and collaborative methods are used to calculate the scores, the proposed system is a hybrid recommendation system. The final recommendations are made by normalizing the scores and selecting the top 10 scores. The system was found to provide suitable recommendations to the user achieving accuracies of 76.4% and 81.4% for the two models, respectively.

**Keywords** TextRank · Recommendation · PageRank · NLP · Scientific articles

## 1 Introduction

Over the years, the quantity of published scientific articles has been exponentially increasing. This is because of a multiple of reasons—growing research community

---

I. Amara (✉) · K. Sai Pranav · H. R. Mamatha  
PES University, Bangalore, India  
e-mail: [sai102030amara@gmail.com](mailto:sai102030amara@gmail.com)

K. Sai Pranav  
e-mail: [saipranav152@gmail.com](mailto:saipranav152@gmail.com)

H. R. Mamatha  
e-mail: [mamathahr@pes.edu](mailto:mamathahr@pes.edu)

leading to more and more scientific articles is being published each year, the ease of publishing a paper has drastically changed, and the number of journals has also increased. According to the National Science Foundation (NSF), the number of scientific articles published in the world changed from 1,755,850 in the year 2008 to 2,555,959 in just 10 years for the topics under science and engineering only. Due to this, a researcher wastes a lot of valuable time in trying to understand the current field the researcher is working in and learn about state-of-the-art research articles. This time could otherwise be spent on understanding and improving influential works. The total amount of global research that was measured by peer-reviewed science and engineering conference articles and journal articles grew around 4% annually during the last decade. Owing to this, there is an upscaling demand for better methods to identify quality scientific articles as old ways of human consumption is lacking. Even though there are advances in search engine technology, it is not very easy and still quite hard for a researcher or student to find new technology to satisfy his/her needs. Under the current circumstances, almost all researchers browse through various works that can be found in top conferences or journals or they depend on key-based searches to find similar work. To make things easier for researchers and students, a recommendation system for scientific articles could help easily identify related or similar articles for each researcher.

## 2 Related Work

In this section, some of the research literature related to recommender systems in general, academic paper recommendation systems and evaluation of recommender systems is presented.

The first and foremost step while creating a recommendation system is to select the right dataset. There were multiple datasets to choose from [1], but the DBLP dataset was chosen since it had information on all the features that were required by the proposed system. There are various methods for research paper recommendation systems available. Some authors suggest using a bag-of-words model on the corpus followed by k-nearest neighbour (kNN) and clustering algorithms to find the recommendations [2] or use NLTK for stemming and lemmatization and then calculate TFIDF followed by creating a similarity matrix to recommend [3]. With the help of a reference-citation matrix, the references and citations by the input paper can be used to get other papers that cited and referenced the same papers respectively and then using a similarity measure recommendation can be made [4].

These approaches are primitive, and a lot of other features can be used to create a much more robust system. For example, keyword-based search, reference analysis, authors, ratings and source analysis [5] or using references and citations of citations to find scientific articles that could have potentially been cited making it a very similar and relevant recommendation [6]. This has been possible with the growing availability of citation data enabling the use of algorithms like PageRank which

provides a complete picture than using just the number of citations to build recommendation systems [7]. If recommendations are required based on past publications of the author, then the paper text and the text from its references and citations can be used to find similar and relevant papers [8]. Some authors also suggest creating an information graph by using a term extractor to achieve state-of-the-art efficiency [9]. A mixture of Paper centric features, Author centric features, Venue centric features and Citation context centric features could help understand the impact of early citations [10]. Hence, a subset of these features can be used to try to find papers that would be more impactful. Another method uses network information and content information and puts forth a method to combine the two modalities with the help of canonical correlation analysis (CCA) [11]. Numerous papers propose the use of explicit ratings as a feature for making recommendations [12], but this comes at its own cost as to obtain implicit ratings, continuous monitoring of the researcher's work is necessary, which raises privacy issues. It is also possible that users would be unwilling to spend time for explicitly rating research papers. [13, 14] suggests the use of capsule networks for text classification tasks. But the dataset used in this project also has citation information due to which graph convolutional networks will be a better fit for the problem.

Most of the articles published are not correctly evaluated. When a research paper recommender system survey was conducted, it was found that about one-fifth of the total approaches were not evaluated, and one-fourth of the total evaluated approaches did not have any baseline. It also showed more than half of these studies had too few participants, and in many cases, there was no mention of how many people participated [15]. Due to this, it is not very easy to compare the performance of different recommender systems.

Results were not mentioned in 45% of the papers that were surveyed and the ones that had any were calculated very differently. Calculate the classification accuracy and explicit rating of recommendation which is 84% for ML papers and an average rating of 3.88 out of 5, respectively [2]. Calculate precision, recall, F1 and mean average precision (MAP) for varying sizes of parameter N [4] and use mean reciprocal rank (MRR) and normalized discounted cumulative gain (nDCG) [6, 8]. [9] use F1 score for comparing the various methods they used. [9]. Use coefficient of determination ( $R^2$ ) and Pearson correlation coefficient ( $\rho$ ) and get the best results of 0.82 and 0.915, respectively [10]. Use MAP and nDCG and finds canonical correlation analysis (CCA) which gives the best results  $-0.6961$  and  $0.4510$  respectively among their other methods [11]. This shows how the comparison of these systems is not an easy task.

### 3 Methodology

In this section, how the two datasets were compiled, and the processes involved in creating the models are discussed. The first model was created using PageRank and TextRank. The reason why TextRank works well is that it includes the information

recursively drawn from the entire text and not just relies on the local context. The second model works by converting each of the articles to vectors and recommends based on the cosine similarity score combined with PageRank. In the subsequent sections, all the modules are explained separately to each of the datasets.

### ***3.1 Data Collection and Cleaning***

The DBLP-Citation-network V10 Dataset was freely available on Aminer.org [16] and consisted of information such as author, title, abstract, year of publishing and citations. The cleaned dataset finally contained over 900,000 articles. A second dataset consisting of only machine learning articles was also created. The features of this dataset were title, abstract, category, authors and citations. This was successfully created by using the CORA dataset [17]. Missing fields had to either be manually filled in or scraped from the web, and if the process failed, then the record was skipped. The titles, references and links of the various articles were in different files haphazardly. The files were parsed through, and the final dataset was brought together.

### ***3.2 Data Pre-processing***

Various pre-processing techniques were used to obtain the features to be used by the recommendation system. The abstract is split into sentences, and pre-processing is carried out. This was possible using the spacy library in python. Spacy is an advanced natural language processing library that helped in splitting the sentences and pre-processing. For every abstract, the first step was to remove all the stop words, e.g., words like me, my, ours, she, its, etc. POS tagging is labelling each word in a sentence with its appropriate parts of speech. Using POS tagging, all words other than nouns and proper nouns were discarded, and this was done because not all words in the sentence are useful. Using a window size of 4, the directed graphs were created and could get the top 20 words from each abstract. Windows are a group of words from the entire set, and each pair in a window has an undirected edge between them.

The text in the scientific articles was also quantified by applying feature extraction. Each abstract was converted to vectors so that the cosine similarity among the documents could be calculated. To do this, the first step was to tokenize each word. This is defined as a way of obtaining meaningful basic units from large samples of text. Simple tokenization can be done by splitting each sentence whenever a white space is encountered. After tokenization, important pre-processing steps like lemmatization and stemming were applied. They both share a common goal and are used to reduce inflectional forms of a word. Stemming is the process wherein the ends of the words are chopped off in the hope that the goal is achieved. Most of the time derivational affixes are also removed. Lemmatization, on the other hand, refers to the

actual use of vocabulary and morphological analysis of words and usually aims to get rid of the inflectional endings only and finally return the base form of the word which is known as the lemma. The next step was to count each word, and these counts are used to assign weights to tokens in articles. It is later also used for dimensionality reduction.

Finally, the tokens were weighed. Assigning weights to tokens can be done in a variety of ways. The method that has been adopted in this project is feature presence (FP). In FP, the weight of a token is simply given by a binary variable which is 1 if the token occurs in an article and 0 otherwise. Since the dimensionality of the obtained vector will be very high, dimensionality reduction was then done. This is done using the singular value decomposition (SVD) algorithm. SVD decomposes vectors onto their orthogonal axes. Dimensionality reduction can be achieved by simply dropping columns. But it can also be achieved by deriving new columns based on linear combinations of the original columns. It was found that the use of SVD resulted in the yield of faster model build times, faster, and a more accurate model. On applying the feature recommendation method as stated above, a list of features is obtained to numerically represent each article. Instead of checking for similarity in such a high dimension, a few features do not indicate similarity between articles (e.g., use of the words “is” or “a”), and hence, they can be removed from the feature set. This not only reduces the number of features but also improves the quality and speed of the recommendation system.

### 3.3 *Creation of Graph*

To create a graph, networkx [9], which is a python library, was used for making, manipulating and studying complex networks. Using this library, each scientific article is treated as a node and a directed graph is built. The graph is built using information from references and citations. This means each edge between the nodes represents a directed link signifying a citation or reference. Using this graph, the PageRank scores of each scientific article were calculated. To make the system scalable and also help in building graphs, graph convolution network was used. It uses a representation learning technique best suited for dynamic graphs. Graph convolution network helps us by predicting embedding of a new node, and it can do this without retraining. To do so, it learns aggregator functions using which embedding of a given node, and its features and neighbourhood can be induced. This is also known as inductive learning. Inductive learning is learning by discovery, and in this method, the model discovers rules by observing examples. Graph convolution network which assumes nodes in the same neighbourhood will have the same embedding. So, using a parameter  $K$ , which denotes the number of neighbours, how many nodes will be considered similar is decided. It was important to choose the right value for  $K$  as a large value will cause undesired sharing of information between the nodes. The parameters of the GCN model used are as follows—filter size ( $16 \times 16$ ), dropout rate 0.5, optimizer is Adam, ReLU activation in the initial layers and SoftMax activation

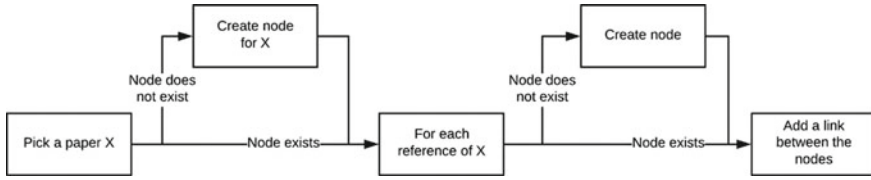
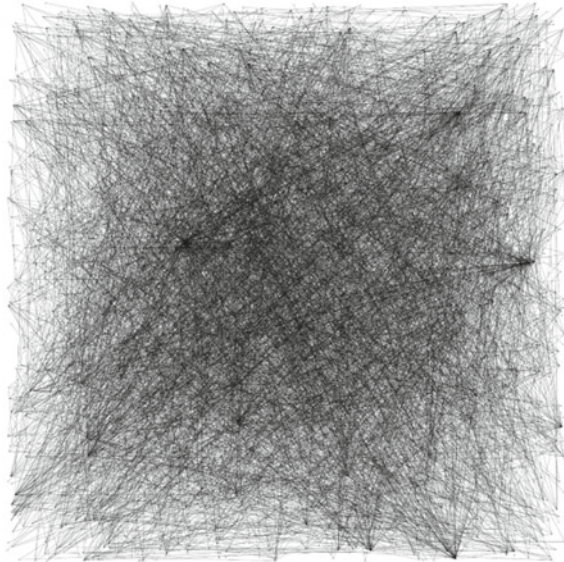


Fig. 1 Citation graph creation process

Fig. 2 Citation graph formed using ML dataset. Nodes and lines represent papers and citations, respectively



in the final layer, and the number of nodes is equal to the number of papers. Early stopping was also used to reduce over-fitting on the test dataset (Figs. 1 and 2).

### 3.4 Calculation of Scores

Two separate models were built on different datasets. The first model provides a user with two ways to get recommendations: using keywords or using the abstract of another article. Once the input is given, the TextRank algorithm is used and the top 20 keywords are identified. Next, the input keywords are compared to each of the keywords in an article and note the scores of overlapping words. This score is used to find out how relevant an article is to the input query. Using the graph created by networkx, the PageRank values were calculated. PageRank is an algorithm used by Google to measure the importance of web pages. It is calculated by analysing the number of incoming and outgoing links to all the pages. In this research, this



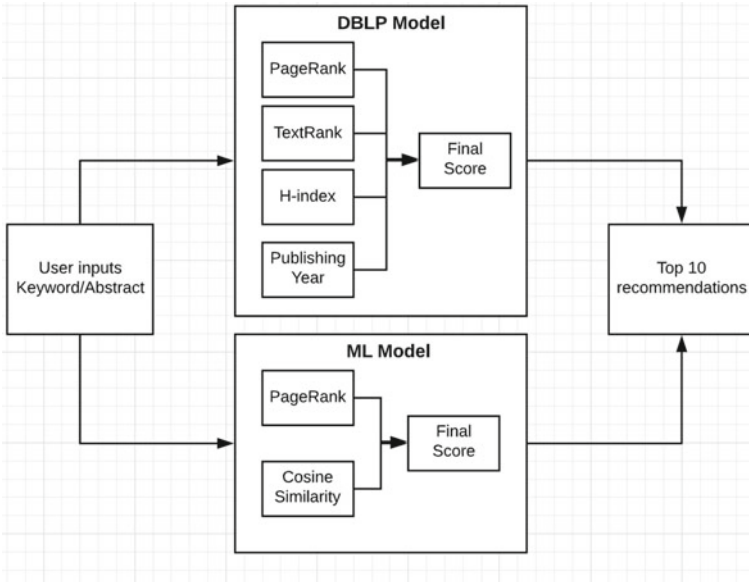


Fig. 3 Architecture of the system

was calculated using the references and the citations of each paper. Another feature used was the H-index of authors. All authors have an H-index. This is the maximum value of h such that h number of articles has at least h citations. Next, an average of the H-index of all the authors for each paper is calculated. This helps assign a score to how influential the author might be. The rationale behind using the average of H-index as a feature is that an influential author will most likely publish another influential paper and also people would be more interested in reading their articles. Finally, a preassigned score for each year is used. The more recent the paper is, the higher the value is for that paper. Doing this will enable us to give more importance to recent work. This array of scores (PageRank, TextRank, author score, year score) is then used to make recommendations. In the ML dataset, since the same data is not available, instead PageRank and cosine similarity scores were used. After vectorizing the scientific articles, cosine similarity of the input with other papers was used as the metric for calculating a score (Fig. 3).

### 3.5 Normalizing Scores and Final Recommendation

The final steps and module of the recommendation system were to normalize and make recommendations. The scores were normalized linearly between 0 and 1 using MinMax normalization to ensure equal representation of each feature. Using the final

**Table 1** Results using DBLP-Citation-network V10 Dataset

Category	Accuracy (%)
Genetic algorithms	88.2
Case based	69.3
Neural networks	65.5
Probabilistic methods	82.6

**Table 2** Results using machine learning dataset

Category	Accuracy (%)
Genetic algorithms	76.0
Case based	74.2
Neural networks	83.5
Probabilistic methods	67.7

obtained score, all the scientific articles were sorted. The recommendations are then given by choosing the top 10 scientific articles.

## 4 Results

The metric used to evaluate this system was accuracy. For each input query, the model was run on the entire corpus and the top few recommendations were taken into consideration. The accuracy was calculated by checking the abstracts of the recommended scientific articles and checking if it was coherent with keyword inputs. Below are the results for DBLP-Citation-network V10 (Table 1).

The system produced an accuracy of 76.4% for the DBLP-Citation-network V10. For the machine learning dataset, the same few categories of research were tested, and the accuracies are shown (Table 2).

Finally, a graph convolution network model is used to predict the category of a paper that a user wants to add to the corpus. This model achieved an accuracy of 81.4%. Due to this, the addition of new scientific articles becomes very easy. In this way, our dataset would become open-source with regular contributions from the research community. While these contributions have to be peer-reviewed to prevent malfeasance, the additional effort can strongly impact the performance of all the above models and hence make a robust system (Fig. 4).

## 5 Conclusion

The final result of this paper is a hybrid recommendation system that recommends suitable articles to the researchers and students. Two models were created, and they

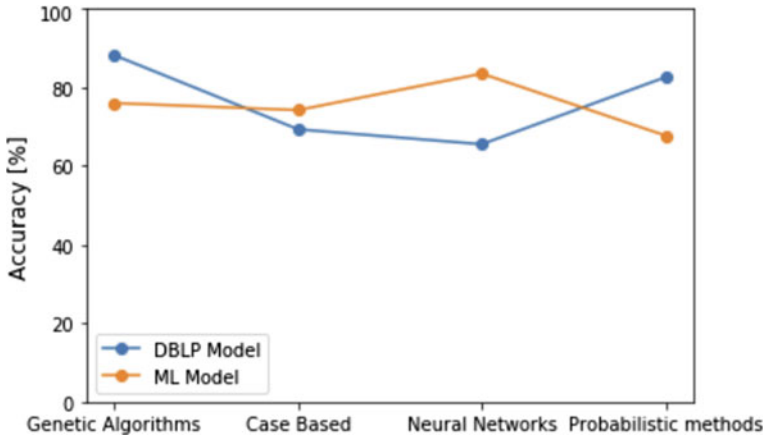


Fig. 4 Results of the two models

were each modelled on separate datasets. The techniques used in each of the models were unique since the hybrid recommender uses content-based features such as TextRank, cosine similarity and collaborative features such as PageRank, references, citations coupled with author information and year of publishing. The recommender systems formed using both models were tested and found to be producing good quality recommendations.

Most recommender systems use keyword analysis to recommend articles. The recommendation system that has been created takes into consideration various attributes like title, abstract, references, year and author importance and produces superior results. Additionally, the proposed system allows for a more narrowed search and provides queries for a specific domain, which is not a common feature of other available recommenders. Existing sites like Google Scholar, CiteSeer, ResearchGate, etc. allow querying with keywords and paper titles only. The proposed approach enables users to get recommendations using not just keywords and titles but also abstracts of papers.

To enable ubiquitous and easy usage of the proposed system, a web interface has been implemented so that people can use it on their local systems as did for implementing and testing. Students and researchers can find suitable recommendations using this web interface, hence reducing the time spent on finding papers to read.

## 6 Future Work

The recommendation system built for the first dataset uses TextRank and PageRank for recommendation and cosine similarity and PageRank for the second dataset. The system currently only uses the text present in the abstract; due to this, the accuracy of the system is less than 90%. This can be improved by considering the text present

in the entire scientific article. It would help in improving the recommendations as other features like In-Text Impact factor can also be considered while determining the top-recommended articles. One observed limitation is that the dataset corpus that is being used does not comprise of all the currently published papers. An increase in the size of the dataset can also help in improving the recommendations given by the system. When a user wants to add a new paper, he must also provide the citation details since the model will not have enough information to calculate the scores without it. With the addition of more categories and scientific articles, the PageRank algorithm becomes more effective. Various other models can be trained on the content data to obtain the most essential data. The scores obtained from these models can then be combined with the PageRank scores to recommend articles.

## References

1. Khalid H, Akmar IM (2018) Evaluation datasets for research paper recommendation systems. In: Data science research symposium 2018: data analytics and its applications, at faculty of computer science and information technology, University of Malaya, Malaysia
2. Lee J, Lee K, Kim JG (2013) Personalized academic research paper recommendation system. Georgia Inst Technol
3. Kerzendorf WE (2019) Knowledge discovery through text-based similarity searches for astronomy literature. *J Astrophysics Astron* 40(3)
4. Haruna K, Ismail MA (2018) Research paper recommender system evaluation using collaborative filtering
5. Gipp B, Beel J, Hentschel C (2009) Scienstein: a research paper recommender system. In: Proceedings of the International conference on emerging trends in computing, pp 309–315
6. Sugiyama K, Kan M-Y (2014) A comprehensive evaluation of scholarly paper recommendation using potential citation papers. *Int J Digit Libr* 16(2):91–109
7. Chen P, Xie H, Maslov S, Redner S (2007) Finding scientific gems with Google's PageRank algorithm. *J Inf* 1(1):8–15
8. Sugiyama K, Kan M-Y (2010) Scholarly paper recommendation via user's recent research interests. In: Proceedings of the 10th annual joint conference on digital libraries—JCDL '10
9. Luan Y (2019) Information extraction from scientific literature for method recommendation. University of Washington
10. Singh M, Jaiswal A, Shree P, Pal A, Mukherjee A, Goyal P (2017) Understanding the impact of early citers on long-term scientific impact. In: 2017 ACM/IEEE joint conference on digital libraries (JCDL)
11. Gupta S, Varma V (2017) Scientific article recommendation by using distributed representations of text and graph. In: Proceedings of the 26th international conference on World Wide Web companion—WWW '17 companion
12. Manouselis N, Verbert K (2013) Layered evaluation of multi-criteria collaborative filtering for scientific paper recommendation. *Procedia Comput Sci* 18:1189–1197
13. Jacob IJ (2020) Performance evaluation of caps-net based multitask learning architecture for classification. *J Artif Intell* 2(01):1–10
14. Vijayakumar T, Vinothkanna R (2020) Capsule Network on Font Style Classification. *J Artif Intell* 2(02):64–76
15. Beel J, Langer S, Genzmehr M, Gipp B, Breitingner C, Nürnberger A (2013) Research paper recommender system evaluation. In: Proceedings of the international workshop on reproducibility and replication in recommender systems evaluation—RepSys '13

16. DBLP-Citation-network V10 dataset, AMiner, Mar 2020. (Online). Available: <https://www.aminer.org/data>
17. CORA dataset, computer science department, University of Massachusetts Amherst, Mar 2020. (Online). Available: <https://people.cs.umass.edu/~mccallum/data.html>

# Classification of VEP-Based EEG Signals Using Time and Time-Frequency Domain Features



M. Bhuvaneshwari and E. Grace Mary Kanaga

**Abstract** The Evoked Potential (EP) is a term that refers to the response generated by the brain in effect to the external stimuli. The response is measured by the strength of the electric potential generated by the brain. The measured response varies depends upon the flickering speed of the stimuli through which the stimuli can be identified. In this study, the activity of the brain while perceiving the visual stimuli of varying frequency has been investigated. The study was made on the Electroencephalography (EEG) dataset designed by the authors that were acquired from the healthy seven subjects whose mean age is 22. The acquired is transformed to the time-frequency domain by applying wavelet transforms, and the statistical features were extracted from the acquired and transformed EEG signals for classification. The proposed study applied machine learning algorithms to classify the appropriate stimuli. The study has experimented with machine learning algorithms like Support Vector Machines (SVM), random forest, K-nearest neighbour, multi-layer perceptron, linear discriminant analysis with the accuracy of 93.14%, 97.85%, 71.08%, 79.65%, 81.46% and 94.09%, 99.06%, 90.02%, 85.16%, 82.91% in time and time-frequency domain, respectively.

**Keywords** Evoked potential · Electroencephalography · Time domain · Time-Frequency domain

## 1 Introduction

The human nervous system is a complex system that consists of billions and billions of neurone which communicates with each other through electrical signals. The

---

M. Bhuvaneshwari (✉) · E. Grace Mary Kanaga  
Karunya Institute of Technology and Sciences, Coimbatore, Tamilnadu, India  
e-mail: [bhvnshwari@gmail.com](mailto:bhvnshwari@gmail.com)

E. Grace Mary Kanaga  
e-mail: [grace@karunya.edu](mailto:grace@karunya.edu)

electrical impulse is encoded with some useful information which can be decoded outside the human body using a system called Brain–Computer Interface (BCI). Brain–computer interface has been emanated as a prominent communication technology between the human and the computer. BCI system benefits people with severe motor disability in rehabilitation and also to communicate with the outside world. The brain signals are acquired using electrophysiological signals such as EEG, ECoG and MEG to capture the signals from the brain. The brain signals thus acquired is processed in the computer to generate the real-time output to accomplish certain tasks. The task includes gaming, rehabilitation, communication and assistive tool for severely disabled people. This real-time output generated is not associated with the normal neural output. The neuromuscular movement plays a prominent role in building these types of application.

Based on the placement of electrodes BCI further classified as invasive (electrodes placed inside the brain region), semi-invasive (electrodes placed between the scalp and the brain), non-invasive (electrodes are placed in the scalp region). Invasive method of signal acquisition has high Information Transfer Rate (ITR) [1], and less noise hence results in high system output. Though the invasive method is more accurate and reliable because of its complexity in implanting, it has not taken for the study. In non-invasive method, electrodes are placed on the scalp, which not only records the brain signals but also the artifacts (muscular movements and voluntary actions), the later can be removed by pre-processing. Manoharan 2019 formulated pre-preprocessing techniques for text and image processing.

Based on the method of generation of electric potential BCIs is classified as active, passive and reactive. In active BCI, the electric potential is generated by voluntary actions of the brain. Passive BCI utilises the potential generated by the non-voluntary actions of the brain. Reactive BCI uses the electric potential generated in response to external stimuli which may be audio, video or somatosensory. This type of potential is called evoked potential.

Biologically the neurons are triggered when an internal stimuli is evoked; internal stimuli here is the human thoughts. Researches have proved that the potential can also be triggered by the external stimuli. This paper focuses on the prediction of stimuli based on sensory stimulation-based evoked potential which is highly useful for paralytic people to communicate their needs.

## 2 Previous Work

Most of the BCI-based communication system was based on speller [2] and oculo-graphy [3] methods. For this speller-based system, an electrocortical potential called Event Related Potential (ERP) after triggering an event is measured and analysed. For different types of stimuli, different ERPs are evoked which can be characterised by the time delay. P300 and Steady State Visual Evoked Potential (SSVEP) are the most common ERPs applied for speller-based systems. The positive deflection of the electrocortical potential derived at 300 ms after the stimulus is evoked is termed P300

[4]. Presented a visual P300-based system as a possible alternative communication for paralytic people.

SSVEP is characterised by positive and negative deflection in the EEG signal in response to a visual stimulus. The stimuli flickering at a constant speed generate a peak frequency matching the stimuli frequency is termed SSVEP [5]. Designed a high-performance SSVEP-based speller system for paralytic people. The existing works have evidenced that EP and ERP can discriminate from stimuli to stimuli and proposed work focus on flickering stimuli based system.

### **3 Methodology**

EEG is an empirical method of recording the brain signals noninvasively that generates specific signal potential based on the subject's attention. The subjects attention can be triggered by sensory stimulation methods whose response is deemed evoked potential. These potential are originated from the corresponding cortical region of the brain stalk. This study triggers the visual cortex of the brain stem to generate the evoked response. Visual-based stimuli are used in prediction of various neurodegenerative diseases such as Alzheimer, Schizophrenia.

#### **3.1 Dataset**

The EEG data is acquired from seven healthy subjects whose mean age is found to be 22.5. The seven subjects are proportioned based on gender into 6:1. The subjects were made to focus on the 5 min visual stimuli that has six images flickered between 5 and 20 Hz with a refreshing rate of 60 Hz per second. 10 s of the interval was given to each image. The placement of the electrode follows the international standard of the 10–20 system. The acquisition maintains a 256 Hz sampling frequency. For signal acquisition, g-USB amp and g-tec recorder are used.

#### **3.2 Stimuli Design**

The stimulus is designed using six flickering images of different frequency ranges. These six images are made to flicker at different frequencies say 5, 6, 9, 10, 14, 20 Hz are combined with 10 s of resting period between each stimulus. The subjects are made to focus the visual stimulus under different environmental conditions (say noisy/silent environment). The subjects view the stimulus through a LED monitor. Since the refreshing rate or flickering speed reflects in the signal strength constant distance of 75 cm is maintained for all the subjects. As the proposed system is



the naïve method in BCI-based communication system, for paralytic people no benchmark datasets are available for comparison.

### 3.3 Evoked Potential

The evoked potential is the electrical potential generated from the visual cortex in response to the visual stimuli. The strength of the non-evoked potential is stronger than evoked potential which might be of few millivolts. The generated EP is unique to objects [6] henceforth can be applied to categorise the objects [7] viewed by the subjects. Since the signals are of very low potential signal averaging is applied to boost up the signal strength. EP usually contains distinct frequency components which have constant phase and latency for a certain time period which helps in the prediction phase [8]. Used EP to recognise the alertness of the driver based on the EP generated in the brain due to the colour intensity of the stimuli.

## 4 Experimental Design

The visual stimulus included 5 min of video for every three trails. Each trial begins with 2 min of preparation time for the subject to focus on the stimulus followed by six images flickered at different frequencies. 10 s of time interval was given after every image to better differentiate the evoked and non-evoked potential (Fig. 1).

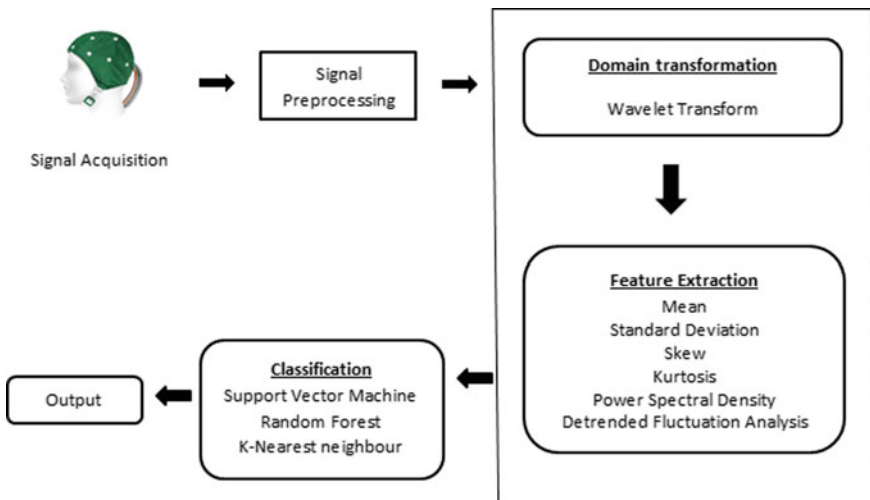


Fig. 1 Architecture diagram of the proposed system

## 4.1 Feature Extraction

Feature extraction is an important step in the classification process as it emerges out to be the essence of the whole dataset. There are many features used in the researches in recent days like time, frequency and time-frequency-related features in the field of EEG signal processing. In this study, the classification results of time and time-frequency-related features have been compared. The univariate and multivariate features like mean, standard deviation, skew, kurtosis, power spectral density (psd) and de-trended fluctuation analysis (dfa) were extracted from the original data. The extracted features were used for classification and prediction process. The archived data was split into test and training dataset in the ratio of 2:5. The data used for prediction was not used in the training phase.

### 4.1.1 Time Domain Features

The EEG signal acquired from the subject is always projected in the time domain. The features mean, standard deviation, skew, kurtosis, power spectral density and de-trended fluctuation analysis are extracted from the original data. The first four features represent the statistical features for all the domain, and the latter represents the frequency-related features. Hence, these six features are taken for the study. The descriptions about the features are elaborated in Table 1.

### 4.1.2 Time-Frequency Domain Features

The time-frequency domain features represent both temporal and spatial characteristics of the data. This can be transformed by applying wavelet transforms into the original data. The wavelet transform decomposes the signal into chunks by applying windowing technique. The basic functional unit of this transform is the wavelet. The wavelet transform can be represented as

$$F_{xy} = \int_{-\infty}^{+\infty} g(t)\varphi_{xy}(t)dt \quad (1)$$

where  $g(t)$  is the original wave, and  $\varphi$  is the arbitrary mother wavelet;  $x$  is the size of the window, and  $y$  is the translation function. After the transformation to the time-frequency domain, the features tabulated in Table 1 are extracted from the transformed data for classification.

**Table 1** Description of the hybrid features extracted in the proposed work

Features	Description	Formulae
Mean	This represents the average of the signal	$\mu = \frac{1}{m} \sum_{j=1}^m n_j$
Standard deviation	This represents the distribution of the signal	$\sigma = \sqrt{\frac{1}{m-1} \sum_{j=1}^m  n_j - \mu ^2}$
Skew	This represents the degree of asymmetry of the distribution of the signal	$S = \frac{C_3}{C_2\sqrt{C_2}}$ where $C_k = \frac{1}{m} \sum_{i=1}^m (n_j - \pi)^k$
Kurtosis	This measures the highest value of the entire dataset	$K = \frac{C_4}{C_2^2 C_2}$ where $C_k = \frac{1}{m} \sum_{i=1}^m (n_j - \pi)^k$
Power spectral density	It is defined as the power present in the signal as a function of frequency	$P_a(f) = \int_{-\infty}^{+\infty} R_a(\tau) e^{-2j\pi f\tau}$ where $j = \sqrt{-1}$
De-trended fluctuation analysis	This measures the self-affinity of the signal	$f(n) = \sqrt{\frac{1}{m} \sum_{t=1}^m (X_t - Y_t)^2}$

## 4.2 Classification

Classification is a technique to predict the class of the new observation based on the training data whose class labels are known. The classes here are the six images, each image represents a class, and since the class names are known and have many classes this classification comes under supervised multiclass classification. From the archived dataset, 70% was used as training and 30% as test data. The archived dataset is considered as raw data, and the feature extracted dataset is considered as another dataset. The performance of both the datasets is compared.

### 4.2.1 SVM

SVM, one of the most commonly used supervised machine learning algorithm for EEG classification builds a hyperplane whose equation is given by

$$W^T X + b = 0 \tag{2}$$

where  $w$  is the weight vector, and  $x$  is the input and  $b$  be the bias. Though there are facts and myths that SVM is suitable for binary classification problem, there are proven

researches [9, 10] for SVM applied in multiclass classification problem. Many hyperplanes can be drawn to distinguish different classes. From the hyperplane, optimal hyperplane for each class can be chosen based on the margin, i.e. the distance between the outer data point of a class and the hyperplane. For a non-linear dataset a function called kernel or kernel trick is applied. This method classifies the data by plotting the data points in a high dimensional space by constructing a hyperplane. As per the research work of [11], SVM performs well on sparse datasets, and for its generalisation characteristics, it is frequently used in EEG-based classification methods and also outperforms for multiclass classification. The classification accuracy results 93.14% and 99.09% with the time domain EEG data and the time-frequency domain data, respectively.

#### 4.2.2 Random Forest

Random forest is one of the tree-based learning algorithms outpaces classifier accuracy for large dataset the implementation is also extended to this algorithm. It is one of the ensemble models that construct a set of decision tree from the randomly selected subset of input data and aggregates the result of each decision tree and predicts the final decision based on the majority voting method. Let  $\{x_1, x_2, x_3, x_4\}$  be the input data and  $\{y_1, y_2, y_3, y_4\}$  be the label then the decision tree constructed by random forest be  $\{x_1, x_2, x_3\}, \{x_1, x_2, x_4\}, \{x_2, x_3, x_4\}$ . This particular classifier outperforms on raw data than on feature extracted data. This creates tree randomly taking the subset of the data. The classification accuracy results in 99.85% and 93.06% in time and time-frequency domain, respectively.

#### 4.2.3 K-Nearest Neighbour

It is a supervised machine learning algorithm that learns from the labelled input data and predicts the output when an unlabelled data is provided. This classifies the data by determining the distance between the data points and sorts them in the ascending order. From the sorted array, first k entries belong to the same label and so on until all the labels are classified. The distance proximity can be measured by distance metrics like Euclidean distance [12] which is calculated as

$$\text{Distance } d(a, b) = \sqrt{\sum_i (a_i - b_i)^2} \quad (3)$$

The classification accuracy of this algorithm varies highly for time and time-frequency domain by  $\pm 20\%$ .

#### 4.2.4 Multi-layer Perceptron

The perceptron is applied to classify the linearly separable data with some limitations for non-linearly separable data. Multi-Layer Perceptron (MLP) fixes this issue by classifying the data that are not linearly separable. MLP has input, output and multiple hidden layers hence the name MLP. At each layer dot product of the activation function such as Rectified Linear Unit (ReLU), sigmoid, tanh and the corresponding weight is pushed to the next layer.

The MLP classifies the dataset in two steps: (i) the data is fed to the input layer and the resulting flow across the hidden layers along with the activation function, (ii) the output of the hidden layer is compared with the desired output. If the derived output has deviated from the desired output, then the weights of the hidden layer are modified and backpropagated again. At the output layer, the backpropagation learning algorithm is applied to reduce the erroneous output.

#### 4.2.5 Linear Determinant Analysis

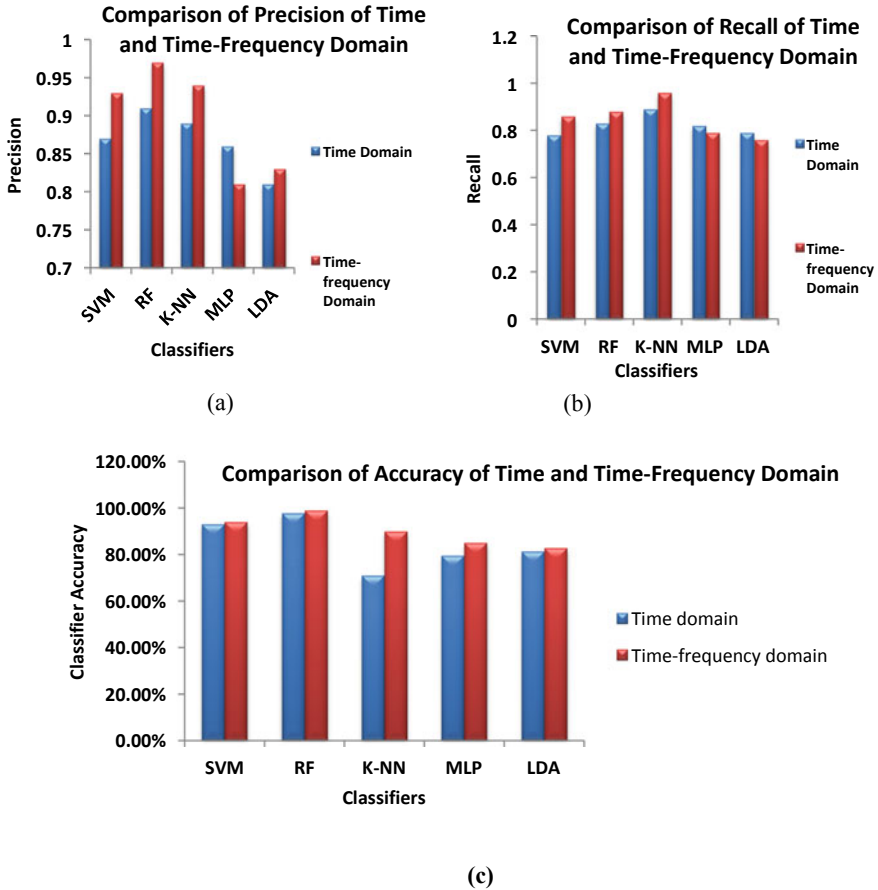
It is one of the popular dimensionality reduction algorithms which can also be used as a classification algorithm. Linear Determinant Analysis (LDA) performs binary and multiclass classification problem. The model works with an assumption that the data are (i) Gaussian and (ii) each attribute contributes the same variance to the dataset. LDA applies Bayes theorem to predict the probability of the classes. For a multiclass classification problem, the difference between the variability of the classes are defined by the sample covariance of the class which is given by

$$\sum b = \frac{1}{X} \sum_{i=1}^X (\mu_i - \mu)(\mu_i - \mu)^T \quad (4)$$

where  $\mu$  is the mean of each class. LDA classifies the provided dataset with the accuracy of 81.46% and 82.91% of time and time-frequency domain features, respectively.

## 5 Results and Discussion

VEP signals were recorded when the subject look into the stimuli of a certain frequency [13]. The potential of the EEG signal varies with the stimuli frequency which was at a specific range when analysed with different subjects. Figure 2 shows the VEP acquired from channel C4 of two different subjects while perceiving the 5 min visual stimuli. The time to evoked potential graph shows a distinct variation among the two subjects for the same stimuli. This shows that the evoked potential falls in a certain range for particular stimuli frequency. By fixing a certain range for



**Fig. 2** Comparison of (a) precision (b) recall and (c) classification accuracy with time and time-frequency domain

specific frequency, the stimuli focused by the subjects can be identified. The classification result of statistical features only data has given poor results. Hence, the hybrid features were extracted, a combination of statistical and frequency features, as tabulated in Tables 2 and 3.

**Table 2** Performance metrics of time domain features

Classifiers	Accuracy (%)	Precision	Recall
SVM	93.14	0.87	0.78
Random forest	97.85	0.91	0.83
k-NN	71.08	0.89	0.89
MLP	79.65	0.86	0.82
LDA	81.46	0.81	0.79

**Table 3** Performance metrics of time-frequency domain features

Classifiers	Accuracy (%)	Precision	Recall
SVM	94.09	0.93	0.86
Random forest	99.06	0.97	0.88
k-NN	90.02	0.94	0.96
MLP	85.16	0.81	0.79
LDA	82.91	0.83	0.76

The original EEG data acquired from the healthy subjects is the time domain data. The statistical features extracted from the original data refer to the time domain data. The transformation algorithm, wavelet transform is applied to the acquired data that represents the data both in time and frequency domain forms the time-frequency domain data. The statistical features extracted from both the data are used for the classification process, and the results of the same are shown in Tables 2 and 3.

Fig. 2 shows the classification accuracy of various classifiers in time and time-frequency domain. From the figure, it was very clear that the classification accuracy of k-NN in time-frequency domain outperforms the time domain approximately by 20%. Hence, for this type of data time-frequency domain is suitable. On comparing the results of precision and recall, time-frequency domain features is found to be better than time domain features.

## 6 Conclusion

This study infers the brain stem response of visual cortex region when subjected to visual stimuli of a certain frequency. It clarifies that the stimuli with high frequency show notable variation than the normal potential. The proposed work applied statistical features and hybrid features separately on the machine learning techniques, in which hybrid features outperforms. The study also compared the classification results of the time domain and time-frequency domain. On comparing the two domains, the time-frequency domain produces and enhances the classification accuracy 20% approximately. Further research can be explored in this area using P300, P100 peaks generated from some other region of the brain.

**Acknowledgements** The authors would like to acknowledge the DST-ICPS chapter for their financial assistance for this study.

## References

1. Lin K, Gao S, Gao X (2019) Boosting the information transfer rate of an SSVEP-BCI system using maximal-phase-locking value and minimal-distance spatial filter banks. *Tsinghua Sci*

- Technol 24(3):262–270
2. Won K, Kwon M, Jang S, Ahn M, Jun SC (2019) P300 speller performance predictor based on RSVP Multi-feature. *Front Human Neurosci* 13(July):1–14
  3. Tuisku O, Surakka V, Rantanen V, Vanhala T, Lekkala J (2013) Text entry by gazing and smiling. *Adv Human-Comput Interact*
  4. Guy V, Soriani MH, Bruno M, Papadopoulo T, Desnuelle C, Clerc M (2018) Brain computer interface with the P300 speller: usability for disabled people with amyotrophic lateral sclerosis. *Annals Phys Rehabil Med* 61(1):5–11
  5. Saravanakumar D, Ramasubba Reddy M (2019) A high performance hybrid SSVEP based BCI speller system. *Adv Eng Inf* 42(February):100994
  6. Sanchez-Hernandez SA, Contreras-Ortiz SH (2018) Analysis and classification of evoked potentials in response to familiar and unfamiliar faces. In: 2018 IEEE Andescon, Andescon 2018—conference proceedings, pp 1–3
  7. Laton J, Van Schependom J, Gielen J, Decoster J, Moons T, De Keyser J, De Hert M, Nagels G (2014) Journal of the neurological sciences single-subject classification of schizophrenia patients based on a combination of oddball and mismatch evoked potential paradigms. *J Neuro Sci* 347(1–2):262–267
  8. Bekdash M, Asirvadam VS, Kamel N, Yanti DK (2016) Estimation of driver alertness to different colors and intensities using brain visual evoked. *Potentials* 20:16–19
  9. Chen J, Wang C, Wang R (2009) Adaptive binary tree for fast SVM multiclass classification. *Neurocomputing* 72(13–15):3370–3375
  10. Thirumala K, Pal S, Jain T, Umarikar AC (2019) A classification method for multiple power quality disturbances using EWT based adaptive filtering and multiclass SVM. *Neurocomputing* 334:265–274
  11. Edla DR, Ansari F, Chaudhary N, Dodia S (2018) Science direct classification of facial expressions from EEG signals using wavelet packet transform and SVM for wheelchair control operations. *Procedia Comput Sci* 132(Iccids):1467–1476
  12. Bablani A, Edla DR, Dodia S (2018) Classification of EEG data using k-nearest neighbor approach for concealed information test. *Procedia Comput Sci* 143:242–249
  13. Stehlin SAF, Nguyen XP, Niemz MH (2018) ScienceDirect EEG with a reduced number of electrodes: where to detect and how to improve visually, auditory and somatosensory evoked potentials. *Integrative Med Res* 38(3):700–707
  14. Manoharan S (2019) A smart image processing algorithm for text recognition, information extraction and vocalization for the visually challenged. *J Innov Image Process* 1(01):31–38



# Performance Analysis of Matching Criteria in Block-Based Motion Estimation for Video Encoding



Awanish Kumar Mishra and Narendra Kohli

**Abstract** Nowadays, in the era of Covid-19, the demand for video data have grown exponentially due to online activities at every level. Online classes, conferences, meetings are everywhere all around us. All business activities are governed by online platforms. Due to the high demand for video data, it has been the main area of research to compress the video size. For all these activities, lossy compression is very useful, due to the limitation of the human visual system, as it can tolerate micro details in any image or video. Block matching algorithms are mostly used to determine motion estimation of the blocks for the video compression in available standards for the video encoding like AVC–Advanced video coding, HEVC–High-efficiency video coding, HDR video encoding compression (N-HVEC) and VVC–Versatile video coding. Motion estimation along with deep learning-based method is showing very motivating results in the field of video compression. Performances of block matching algorithms depend on the matching criterion. In this paper, various available matching criteria are reviewed based on—mean of the number of search locations for every block, average MAD per pixel and average peak signal to noise ratio (PSNR).

**Keywords** Motion estimation · Matching criterion · Search window · Block matching · MAE

## 1 Introduction

Video data have redundant data in terms of spatial redundancy and temporal redundancy. Size of video data may be reduced in both ways, i.e. by reducing spatial redundancy and by reducing temporal redundancy. In any real-time video, most of

---

A. K. Mishra (✉) · N. Kohli  
Harcourt Butler Technical University, Kanpur, India  
e-mail: [mishra.awanish@gmail.com](mailto:mishra.awanish@gmail.com)

N. Kohli  
e-mail: [nkohli@hbtu.ac.in](mailto:nkohli@hbtu.ac.in)



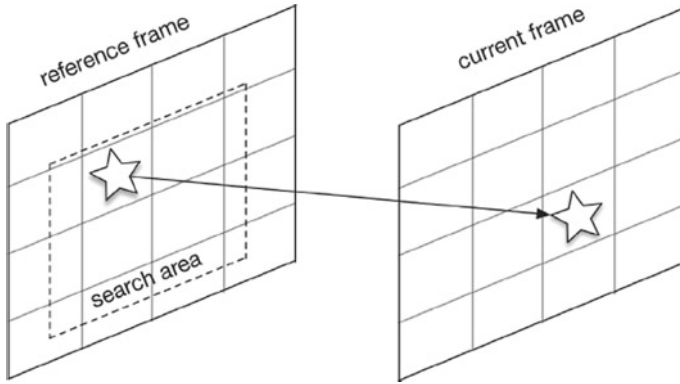
**Fig. 1** Video encoder and decoder

the data in adjacent frames are redundant, and hence, by removing temporal redundancy, the size of video data can be reduced up to great scale. There are many effective algorithms for the reduction of temporal redundancy, but among them, block matching algorithms are producing the best results in terms of predicted image quality and execution time. In block matching algorithms for the motion estimation and compensation of the blocks, one matching criterion is also used to measure the matching degree of the blocks. In this process, video data is converted into frames, and the first frame works as the reference frame for the reduction of temporal redundancy from the second frame, i.e. current frame. This motion vector is the root of the encoding and decoding of the frames in video data. Any real-time video has approximately 30 frames per second, and hence, the movement of the block in adjacent frames is assumed to be within some range  $p$ . Full search block matching algorithm produces the best results, but it is too time-consuming as it searches all locations. To reduce the time complexity, various block matching algorithms have been presented like three-step search [1], four-step search [2], diamond search [3], hexagon-based search [4], adaptive rood pattern search [5], fast motion estimation [6], zero motion prejudgment [7], adaptive diamond search algorithm [8]. These block matching algorithms are using only selected points in the search window to search the matching block. Block matching algorithms are used to match blocks of the current frame and reference frame using the matching criterion. This paper is reviewing matching criteria available to be used with block matching algorithms based on various parameters mean of the number of search locations for every block, average MAE per pixel, the average number of bits per pixel value and average peak signal to noise ratio (PSNR) (Fig. 1).

Performance of matching criteria is evaluated when used with the adaptive diamond search algorithm to check the similarity between the blocks of the current frame and reference frame. Size of the window that is used to search the similar block is 7. Size of the search window tells about the maximum deviation of the block from the current frame to the reference frame.

## 2 Block Matching Criteria

A video sequence contains many objects in it, and these objects are moving in all directions with some geometric operations like rotation, translation, scaling and zooming. To study any moving object in detail, it is essential to extract frames from the video data and then to process. To study about the contents of any frame, it is preferred to divide the frame into equal size blocks. Blocks of the current frame



**Fig. 2** Motion compensation

are mapped with blocks of the reference frame in the searched window based on some suitable matching criterion. There are various matching criteria to study the matching degree of the blocks. The most common and effective criterion to measure the similarity of the block is MAD–Mean absolute difference (Fig. 2).

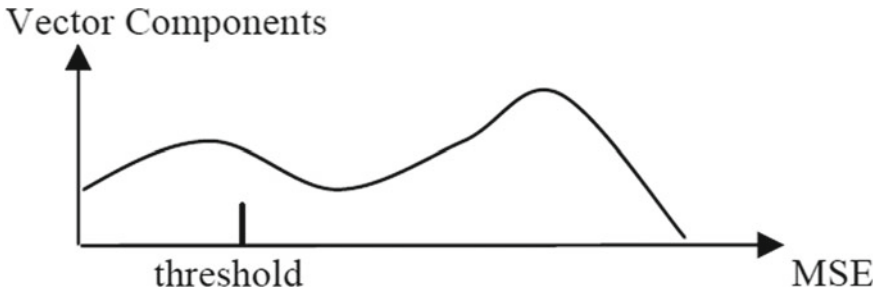
### 2.1 Mean Absolute Difference (MAD)

The matching criterion that is mostly used with all block matching algorithm and found in all the literature of block matching techniques is mean absolute difference or mean absolute error. Block with minimum MAD is chosen as the block of choice.

$$\text{Func\_MAD}(u, v) = \frac{1}{N^2} \sum_{x,y} |\text{curr}(x, y) - \text{ref}(x + u, y + v)| \quad (1)$$

Position of the block in the current frame is represented by the point  $(u, v)$  and the size of the block is  $N \times N$ . Search window is of size  $(2p + 1) \times (2p + 1)$ ; hence,  $-p \leq u$  and  $j \leq +p$ .  $\text{curr}(x, y)$  and  $\text{ref}(x, y)$  are intensity values of pixel  $(x, y)$ , respectively, in the current frame and reference frame. The motion vector for the given arrangement is  $(u, v)$ , which produces the least cost using the function  $\text{Func\_MAD}(u, v)$ . The minimum value of the  $\text{Func\_MAD}$  gives the measure of the residual error between the predicted block and the actual block. Mean absolute difference and mean square error (MSE) both are equivalent in terms of performances but MSE highlights the residual error.

$$\text{Func\_MSE}(u, v) = \frac{1}{N^2} \sum_{x,y} (\text{curr}(x, y) - \text{ref}(x + u, y + v))^2 \quad (2)$$



**Fig. 3** Vector matching criterion

## 2.2 Vector Matching Criterion (VMC)

In the mean absolute difference criterion for the block matching, the decision is taken based on the average of the absolute differences between the corresponding intensities of the pixels of the current block and reference block. MAD does not consider any high error value of any particular pixel. MAD rejects those matching blocks in which error of very few pixels dominates all the other matching pixels. Due to this reason, some times MAD selects blocks from the reference frame with least residual error but mismatching block. This problem of MAD was resolved in vector matching criterion proposed by Wang and Chen [9]. In vector matching criterion, block of size  $N \times N$  is further subdivided into  $\frac{N^2}{4}$  macroblocks each of size  $2 \times 2$ . All macroblocks are compared with the corresponding macroblock of the block of the reference frame. In VMC, some threshold value is fixed based on allowed residual content. The macroblocks with less than threshold MAD value are counted for all candidate blocks in the search window of the reference frame. Block from the reference frame with the highest number of the counted macroblock is selected as the mapped block. It is called vector matching criterion because in this every block is represented as a vector, and every macroblock is a member of that vector. VMC removes the problem of MAD, but practical implementation of MAD is simple if compared with VMC (Fig. 3).

## 2.3 Smooth Constrained MAD (SC-MAD) Criterion

In video compression to reduce temporal redundancy, residual error plays a significant role. The residual frame is obtained by subtracting the coded frame from the actual frame. This residual frame is transformed in different form using the discrete cosine transform (DCT). As per the characteristics of discrete cosine transform, the smooth residual frame takes less bits in transformation, whereas the non-smooth residual frame takes more bits in transformation. Based on this property of DCT, a new matching criterion was proposed by Jing et al. [10] called smooth constrained

mean absolute difference (SC-MAD). Smooth constrained MAD is the criterion to measure the similarity of the block for the motion estimation and compensation to minimize the bits required to code. Smooth constrained MAD also reduces distortion. In this criterion to measure the similarity of the block for motion compensation, residual frame is an essential component, along with this SC-MAD also uses minimum residual value and the maximum residual value. To apply SC-MAD, residual frame is to be transformed in DCT, and DCT is applied in the blocks of size  $8 \times 8$ . Due to the size restriction of DCT, the residual frame of size  $16 \times 16$  is further subdivided into four small size residual frame of size  $8 \times 8$ . For each small size residual frame minimum and the maximum residual value is estimated, and for each small size residual block  $\text{MaxMinError}(i)$  is calculated as

$$\text{MaxMinError}(i) = \text{MaxResidual}(i) - \text{MinResidual}(i) \quad (3)$$

$\text{MaxResidual}(i)$  is the maximum intensity of the subblock, and  $\text{MinResidual}$  is the minimum intensity. Smooth constrained mean absolute error is computed in the following way

$$\text{Func\_SC - MAD} = \text{Func\_MAD} + \alpha \sum_{i=1}^4 \text{MaxMinError}(i) \quad (4)$$

where  $\alpha$  is the weighing factor to estimate the value of  $\text{Func\_SC-MAD}$ . Block in the search window of the reference frame with the minimum value of the  $\text{Func\_SC-MAD}$  is selected as the matching block.

## 2.4 Scaled Value Criterion

To overcome the problem of MAD, VMC and SC-MAD, a new scaled value criterion [11, 12] were suggested in which the pixel values are scaled to lie between 0 and 1. This matching criterion is very effective, especially in case of rotation, zoom and affine transformation. Ref is the reference frame, and Curr is the current frame of size  $(N \times N)$ . Intensity values of the pixels in the current frame and reference frame are, respectively, denoted by  $\text{Ref} = [\text{Ref}_1, \text{Ref}_2, \dots, \text{Ref}_{N_2}]$  and  $\text{Curr} = [\text{Curr}_1, \text{Curr}_2, \dots, \text{Curr}_{N_2}]$ . New scaled intensity values of the frames are obtained by the following formula:

$$\text{Ref}_{\text{new}} = \frac{\text{Ref}_{\text{old}} - \text{Ref}_{\text{min}}}{\text{Ref}_{\text{max}} - \text{Ref}_{\text{min}} + 1} \quad (5)$$

$$\text{Curr}_{\text{new}} = \frac{\text{Curr}_{\text{old}} - \text{Curr}_{\text{min}}}{\text{Curr}_{\text{max}} - \text{Curr}_{\text{min}} + 1} \quad (6)$$

Cost function is applied over these scaled values:

$$\text{Cost (Ref, Curr)} = \frac{1}{N^2} \sum_{k=1}^{N^2} \text{func}(|\text{Ref}_{\text{new}} - \text{Curr}_{\text{new}}, \tau|) \quad (7)$$

where

$$\text{func}(\text{val}, \tau) = \begin{cases} 1 - \tau * \text{val}, & \text{if } \tau * \text{val} \leq 1 \\ 1, & \text{else} \end{cases} \quad (8)$$

A minimum value of Cost (Ref, Curr) will decide the matching block in the reference frame and hence will produce a motion vector.

## 2.5 Modified Scaled Value Criterion

In the entire available block matching criteria, the final decision is taken based on the intensity values of the pixels in the block. These criteria will never match a block of the current frame with a matching block of the reference frame even for the duplicate but differently shaded frames. To remove this problem of the matching criterion, Purwar [13] suggested a new scaled value criterion, in which the intensity of the frames is scaled based on below described rules, and then, the cost function is applied to finally select a matching block in a reference frame. The scaled value criterion is used with block matching techniques to verify that the block from the current frame is matching to the selected block from a reference frame or not based on the cost function. Ref is the reference frame, and Curr is the current frame of size  $(N \times N)$ . Intensity values of the pixels in the current frame and reference frame are, respectively, denoted by  $\text{Ref} = [\text{Ref}_1, \text{Ref}_2, \dots, \text{Ref}_{N_2}]$  and  $\text{Curr} = [\text{Curr}_1, \text{Curr}_2, \dots, \text{Curr}_{N_2}]$ .

The scaled value criterion is very useful in case of zoomed videos and videos with rotation effect. Every frame has a different range of pixel values in it. Maximum value and minimum value of pixels in the reference frame are  $\text{Ref}_{\text{max}}$  and  $\text{Ref}_{\text{min}}$ , and maximum value and minimum value of pixels in the current frame are  $\text{Curr}_{\text{max}}$  and  $\text{Curr}_{\text{min}}$ . To apply block matching with scaled value criterion, pixel values are scaled in the following way:

Step1: If  $(\text{Curr}_{\text{max}} - \text{Curr}_{\text{min}}) \leq (\text{Ref}_{\text{max}} - \text{Ref}_{\text{min}})$ , then

$$\text{Ref}_{\text{new}} = (\text{Ref}_{\text{old}} - \text{Ref}_{\text{min}}) \quad (9)$$

$$\text{Curr}_{\text{new}} = (\text{Curr}_{\text{old}} - \text{Curr}_{\text{min}}) \text{round} \left\{ \frac{(\text{Ref}_{\text{max}} - \text{Ref}_{\text{min}})}{(\text{Curr}_{\text{max}} - \text{Curr}_{\text{min}})} \right\} \quad (10)$$

Otherwise

$$\text{Curr}_{\text{new}} = (\text{Curr}_{\text{old}} - \text{Curr}_{\text{min}}) \quad (11)$$

$$\text{Ref}_{\text{new}} = (\text{Ref}_{\text{old}} - \text{Ref}_{\text{min}}) \text{round} \left\{ \frac{(\text{Curr}_{\text{max}} - \text{Curr}_{\text{min}})}{(\text{Ref}_{\text{max}} - \text{Ref}_{\text{min}})} \right\} \quad (12)$$

Using the above given four equations modified values of the intensities for the pixels of the current frame and reference are obtained. The cost function to evaluate the matched block is applied over the new values of the intensities in the following way:

Cost(Ref, Curr) is the function to evaluate the matching cost of the blocks from the current frame and reference frame

$$\text{Cost}(\text{Ref}, \text{Curr}) = \frac{1}{N^2} \sum_{k=1}^{N^2} \text{function}(|(\text{Ref}_{\text{new}}) - (\text{Curr}_{\text{new}})|, \tau \text{au}) \quad (13)$$

where function function(val,  $\tau \text{au}$ ) is defined as,

$$\text{function}(\text{val}, \tau \text{au}) = \begin{cases} \text{val} & \text{if } \text{val} \leq \tau \text{au} \\ \max(\text{Ref}_{\text{max}}, \text{Curr}_{\text{max}}) & \text{else} \end{cases} \quad (14)$$

The cost function function( $(|\text{Ref}_{\text{new}}) - (\text{Curr}_{\text{new}}|)$ ,  $\tau \text{au}$ ) evaluates the matching degree of the blocks in the current frame and reference frame using the new scaled intensities  $\text{Curr}_{\text{new}}$  and  $\text{Ref}_{\text{new}}$ , and the value threshold tau is defined in the following given way:

$$\tau \text{au} = \max((\text{Curr}_{\text{max}} - \text{Curr}_{\text{min}}), (\text{Ref}_{\text{max}} - \text{Ref}_{\text{min}})) \quad (15)$$

it is clear that only those pixels will have the role in the decision of the matching for which the value of the threshold  $\tau \text{au}$  is greater than equal to  $|\text{Ref}_{\text{new}}) - (\text{Curr}_{\text{new}}|)$ , i.e.  $\tau \text{au} \geq |\text{Ref}_{\text{new}}) - (\text{Curr}_{\text{new}}|)$ .

Finally, the block with least value of the cost function Cost(Ref, Curr) will be chosen as the matching block, and the motion vector for the chosen block will be the correct and acceptable one.

### 3 Results

Experimental results to evaluate five criteria mean absolute difference MAD, vector matching criterion VMS, smooth constrained mean absolute difference SC-MAD, scaled value criterion SVC and modified scaled value criterion MSVC using block matching algorithm adaptive pattern selection strategy for diamond search are shown in tables based on the mean of the number of search locations for every block, average MAD per pixel and average peak signal to noise ratio (PSNR). Value of search parameter is  $\pm 7$ , and the value of threshold  $\tau$  has been chosen as 10. Total seven videos have been used in this analysis, and these are Miss America, Multibar, Susie,

**Table 1** Comparison of performance in terms of an average MAD per pixel using adaptive pattern selection strategy for diamond search

Video data	Average MAD per pixel				
	MAD-based criterion	VMC-based criterion	SC-MAD criterion	Scaled VC criterion	Modified SVC criterion
Miss America	4.77	5.92	5.88	5.05	4.55
Multibar	5.37	15.25	15.6	5.39	5.55
Susie	4.76	4.42	7.22	4.38	4.22
Airbus	10.07	12.94	15.27	6.22	9.19
Cactus Comb	14.21	14.88	13.1	12.35	11.54
Flower	17.75	22.92	19.37	16.14	15.8
Carphone	12.56	17.56	14.85	11.33	9.44

**Table 2** Comparison of performance in terms of the mean of the number of search locations for every block using adaptive pattern selection strategy for diamond search

Video data	Mean of number of search locations for every block				
	MAD-based criterion	VMC-based criterion	SC-MAD criterion	Scaled VC criterion	Modified SVC criterion
Miss America	21.89	22.57	24.08	22.46	21.9
Multibar	24.23	24.54	24.56	24.37	24.1
Susie	23.34	22.83	24.35	22.98	22.75
Airbus	23.79	24.11	25.11	23.36	22.78
Cactus Comb	23.76	23.92	24.02	23.54	22.4
Flower	23.76	23.94	23.75	23.56	23.39
Carphone	21.93	22.28	22.88	21.33	21.1

Airbus, Cactus Comb, Flower and Carphone. Results are shown in the following tables and graphs (Tables 1,2 and 3; Figs. 4, 5 and 6).

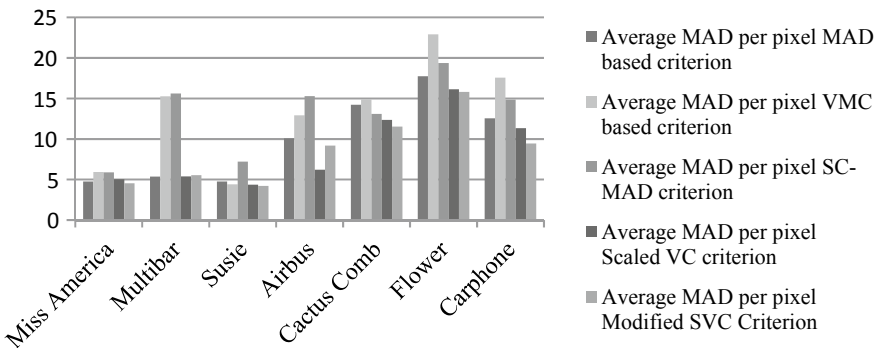
## 4 Conclusion

Various existing block matching criteria are reviewed in this paper. In the available literature, most of the block matching algorithms for the motion estimation and compensation are using mean absolute difference (MAD) as a matching criterion. In a practical scenario, other matching criteria are very effective and are used in various coding standards. One matching criterion is more suitable for one situation while in another situation some other matching criterion performs better. Mean absolute difference (MAD) is simple in understanding and application too. The problem of mean absolute difference (MAD) is resolved by the vector matching criterion (VMC).

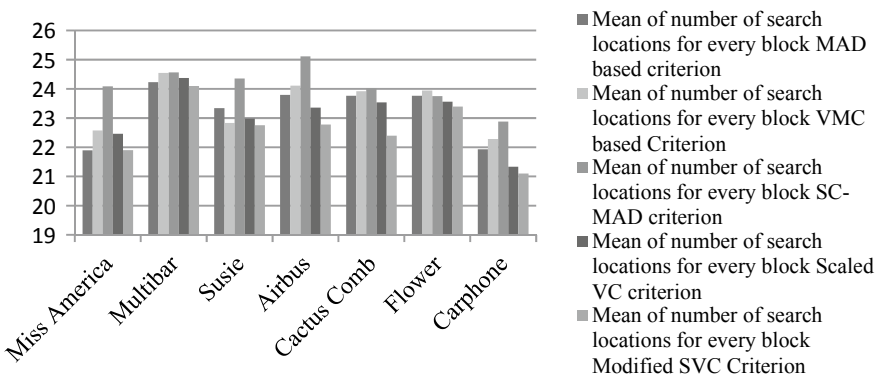


**Table 3** Comparison of performance in PSNR using adaptive pattern selection strategy for diamond search

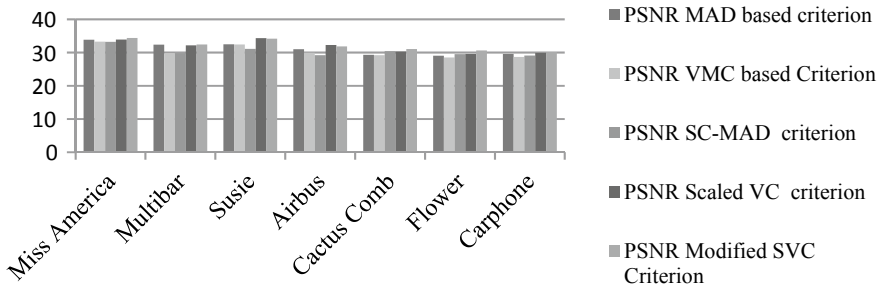
Video data	PSNR				
	MAD-based criterion	VMC-based criterion	SC-MAD criterion	Scaled VC criterion	Modified SVC criterion
Miss America	33.89	33.3	33.25	33.96	34.4
Multibar	32.38	29.77	29.85	32.22	32.45
Susie	32.5	32.44	31.1	34.35	34.2
Airbus	31	29.96	29.21	32.28	31.85
Cactus Comb	29.31	29.21	30.45	30.29	31.04
Flower	29.07	28.51	29.55	29.64	30.65
Carphone	29.61	28.74	29.14	29.93	30.18



**Fig. 4** Comparison of performance in terms of an average MAD per pixel using adaptive pattern selection strategy for diamond search



**Fig. 5** Comparison of performance in terms of mean of number of search locations for every block using adaptive pattern selection strategy for diamond search



**Fig. 6** Comparison of performance in PSNR using adaptive pattern selection strategy for diamond search

Smooth constrained mean absolute difference (SC-MAD) is better than others in tackling discrete cosine transform. To optimize the performance of the block matching algorithm, by ignoring the effect of rotation and zoom, the scaled value criterion was proposed. To further enhance the performance scaled value, criterion was modified to include affine transformation. In the coming days of tremendous growth in video data, it will be essential to further minimize the video data size using some coding. One matching criterion is not sufficient to encode video of different situations and types. All criteria are important but the performance shows that the scaled value criterion is above all.

## References

1. Li R, Zeng B, Liou ML (1994) A New three step search algorithm for block motion estimation. *IEEE Trans Circ Syst Video Technol* 4(4)
2. Po LM, Ma WC (1996) A novel four step search algorithm for fast block motion estimation. *IEEE Tran Circ Syst Video Technol* 6(3)
3. Zhu S, Ma KK (2000) A new diamond search algorithm for block matching motion estimation. *IEEE Trans Image Process* 9(2)
4. Zhu C, Lin X, Chau LP (2002) Hexagon based search pattern for fast block motion estimation. *IEEE Trans Circ Syst Video Technol* 12(5)
5. Nie Y, Ma K-K (2002) Adaptive rood pattern search for fast block-matching motion estimation. *IEEE Trans Image Process* 11(12):1442–1449
6. Ismail Y, McNeely JB, Shaaban M et al (2012) Fast motion estimation system using dynamic models for H.264/AVC video coding. *IEEE Trans Circuits Syst Video Technol* 22(1):28–42
7. Arora SM, Rajpal N, Khanna K (2016) A new approach with enhanced accuracy in zero motion prejudgment for motion estimation in real-time applications. *J Real-Time Image Proc* 1–17
8. Pan Z, Zhang R, Ku W et al (2019) Adaptive pattern selection strategy for diamond search algorithm in fast motion estimation. *Multimed Tools Appl* 78:2447–2464. <https://doi.org/10.1007/s11042-018-6353-2>
9. Wang S, Chen H (1999) An improve algorithm of motion compensation MPEG video compression. In: *Proceedings of the IEEE international vehicle electronics conference (IVEC'99)* (Cat. No.99EX257), Changchun, China, vol 1, pp 261–264. <https://doi.org/10.1109/ivec.1999.830680>

10. Jing X, Zhu C, Chau L-P (2003) Smooth constrained motion estimation for video coding. Elsevier J Signal Process 83:677–680
11. Purwar RK, Prakash N, Rajpal N (2009) A quality based motion estimation criterion for temporal coding of video. In: 2009 Seventh international conference on advances in pattern recognition, Kolkata, pp 61–64. <https://doi.org/10.1109/icapr.2009.66>
12. Purwar RK, Prakash N, Rajpal N (2008) A criterion to measure the similarity of the block for inter frame coding of video. In: 2008 International conference on audio, language and image processing, Shanghai, pp 133–137. <https://doi.org/10.1109/icalip.2008.4589962>
13. Purwar RK, Prakash N, Rajpal N (2011) A matching criterion for motion compensation in the temporal coding of video signal. SIViP 5:133–139. <https://doi.org/10.1007/s11760-009-0149-9>

# Convolutional Recurrent Neural Network Framework for Autonomous Driving Behavioral Model



V. A. Vijayakumar, J. Shanthini, and S. Karthick

**Abstract** Autonomous vehicles, without the help of a human, support challenging tasks for sensing the environment and vehicle navigation. The driving behavior is controlled automatically from the observed surroundings using many supervised learning methods that provide action output based on matching the visual inputs and training labels. Most essentially, deep learning algorithms offer improved processing of observed input data but with the increased training, the complexity in processing the real-time data eventually becomes complex. In this paper, an autonomous driving model driven by a behavioral model is designed incorporating (a) recognition, (b) planning and (c) prediction modules. Each module is designed to regulate the processing of input trajectory video data. Additionally, deep learning classifiers are included to improve the automated ability of planning and prediction modules. Initially, the recognition module is planned to limit the redundant data from the raw input data. Secondly, the planning module is designed with convolutional neural network (CNN) to classify the predictable and unpredictable objects from the surrounding trajectories occurring in the line of sight. Finally, the prediction module is designed with recurrent neural network (RNN) to predict the future driving patterns based on the present condition and past driving outputs. The simulation results show that the proposed hybrid deep learning behavioral model offers improved autonomous driving than other existing autonomous driving models. The results of different environments prove that the proposed hybrid model offers increased scalability in terms of improved recall rate of 95.15%, 96.13% and 97.72% in terrain, dense and light traffic zones, respectively, than existing methods.

---

V. A. Vijayakumar (✉) · J. Shanthini · S. Karthick  
Department of Computer Science and Engineering, SNS College of Technology, Coimbatore,  
India

e-mail: [vjkumar\\_va@hotmail.com](mailto:vjkumar_va@hotmail.com)

J. Shanthini

e-mail: [drjshanthini@gmail.com](mailto:drjshanthini@gmail.com)

S. Karthick

e-mail: [profskarthik@gmail.com](mailto:profskarthik@gmail.com)

**Keywords** Convolutional neural network · Recurrent neural network · Recognition · Planning and prediction · Autonomous driving

## 1 Introduction

Autonomous vehicles in the recent past have improved the potential of efficiency and safety. This, in turn, reduces the road fatalities and increases the productivity and quality of the time spent in cars. The transformation of autonomous vehicles into utility requires major factors like vehicle autonomy by varying the vehicle design, perception and human interaction via control, coordination and planning [1].

The operability of autonomous vehicles in the complex environments and the dynamic surroundings requires the use of generalization capability to control, coordinate and plan the predictable and unpredictable situations along the trajectories. The generalization capability requires the operation on time to attain the human-level reliability and safety in a complex rugged and terrain environment. Here, both the informed and uninformed decisions by a prediction model should require an accurate perception [2].

Various computer vision systems are deployed so far on autonomous driving to achieve the error rate in a minimal and acceptable range. However, most systems fail in achieving the acceptable error rates due to incorrect decision making by the computer vision navigation system. In recent times, the approaches combining decision making, control and perception modules obtain promising results [1].

Most recently, the machine and deep learning modules incorporate well with complex planning and decision making, where the definite performance of driving still has been a challenge that requires future solutions. Deep learning models have proven its success in various object tracking systems; however, it requires a lot of supervised labelled training that eventually increases the complexity in tracking [3]. The generation of unlimited labelled data would eventually affect the performance of deep learning classification models with redundant training sets. Hence, the challenge is to limit the training data on a deep learning model for recognizing the surrounding environment considering both predictable and unpredictable objects.

Considering the above challenge, in this paper, a model is designed and developed that plans and makes optimal decisions for autonomous vehicles using its limited datasets. The model further adopts learning pattern for the interaction with its adjacent vehicle(s) and its driving pattern based on past patterns with proper control reaching its destination range.

The main contribution of the paper involves the following

- (a) The authors design an autonomous driving model with a hybrid deep learning model that comprises a convolutional neural network (CNN) and recurrent neural network (RNN).
- (b) The author(s) has planned to execute the deep learning hybrid modules in a path planning behavioral module that involves recognition, prediction and planning behavior of autonomous driving on a rugged and terrain surface. The hybrid

CNN with RNN is developed to equip with the recognition, prediction and planning behavior of autonomous driving.

- (c) The prediction in this behavioral model uses CNN for classification of predictable and unpredictable behavior into classes, and the planning module uses RNN for future prediction of vehicle smooth movement based on CNN input classes.

The outline of the paper is given below: Sect. 2 provides the related works. Section 3 details the proposed method with three different models using CNN-RNN methods. Section 4 evaluates the performance of the proposed model. Section 5 concludes the entire work with possible directions for future scope.

## 2 Related Works

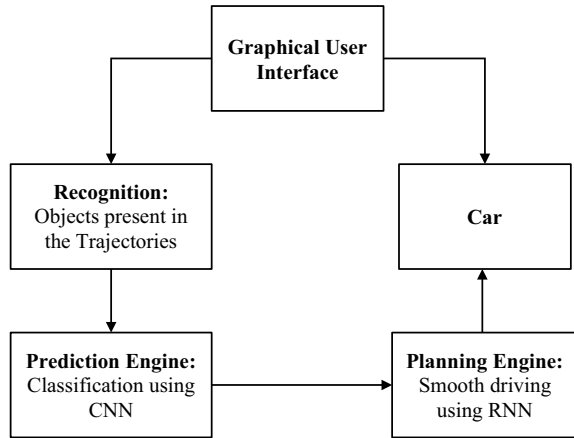
Various existing deep learning classifiers are considered to develop the present study, out of which, YOLO [4] using CNN detects the objects from the input images [5–7]. However, this method uses a road lane detector for the track detection from the input video frames and it provides the decision-making driving behavior to the autonomous cars. CNNs are also used for human tracking [8] but it suffers from the lack of input training data. This limitation is avoided in [9], where the CNNs are trained in prior with a large dataset for the classification and it is fine-tuned with limited tracking data. To study further the temporal correlation between the objects, RNN [10] is also used in automated self-driving cars [11, 12]. The RNNs handle the sequential prediction and it can process on a multi-dimensional image [13]. The RNNs fully utilizes the temporal and semantic information; however, with the increasing frame size, the RNN performance is limited under complex scenarios, especially in traffic conditions [14].

## 3 Proposed Method

The most significant feature of the present study involves the selection of neural network architecture for the proposed deep learning algorithm that can provide similar performance w.r.t real-time driving performance of a car. In this section, to achieve a proper selection of neural network architecture, the autonomous driving model (Fig. 1) is presented using a hybrid deep learning model that involves the use of convolutional neural network (CNN) and recurrent neural network (RNN).

For autonomous driving, the system is illustrated in Fig. 1, which uses CNN to detect the objects and multiple objects in a video frame is selected using the stacked ensemble CNN framework. This is responsible for recognition and prediction of objects. The predicted outputs from CNN are given to inputs RNN planning module that offers path planning based on the objects present in the road segments.

**Fig. 1** Framework of proposed CNN-RNN model for autonomous driving



It is not enough for an autonomous driving agent to recognize its environment; it must also be able to build internal models that predict the future states of the environment. The level of intelligence using the hybrid deep learning model is used to operate the car autonomously on rugged and terrain surfaces.

The hybrid CNN with RNN is developed to equip with the recognition, prediction and planning behavior of autonomous driving.

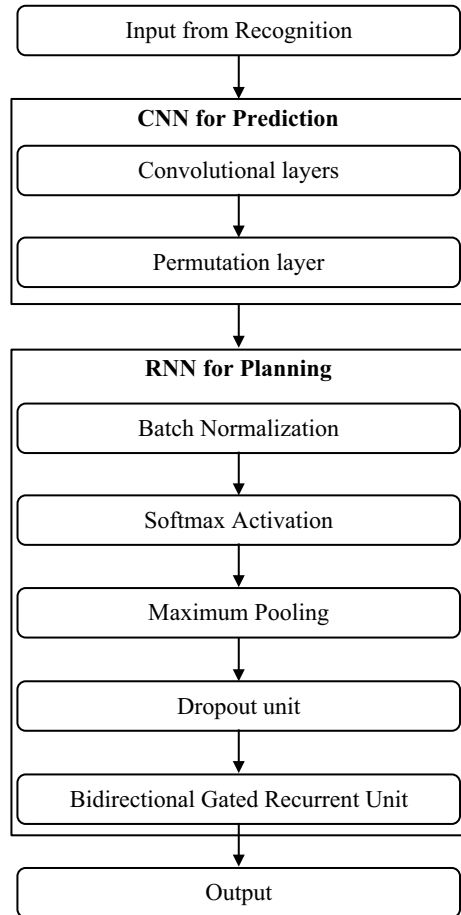
- Step 1. The recognition module estimates the path and accepts the input objects that include both predictable and unpredictable objects from the surroundings, and sends it to the prediction module.
- Step 2. The prediction module using CNN classifies the predictable and unpredictable classes (behavior), and it is then updated continuously with the trajectories in an iterative manner to achieve the task of an accurate autonomous driving.
- Step 3. The planning module using RNN interacts actively to the changes (predictable and unpredictable classes) to the indicated task (i.e., accident-free driving that should enable collision-free driving in cluttered environments). RNN makes the planning of future prediction, where the model incorporates the results of classification to plan the smooth driving action sequences. The planning model handles the objective function in a time-variant mode and it achieves optimal high-quality motion while driving using this hybrid deep learning model (Fig. 2).

#### a. **Recognition**

In the recognition model, the objects in the external environments are extracted during the testing phase. The redundant frames like objects not on the trajectories are considered as redundant and eliminated as input data. The residual data with objects is sent as an input to the CNN classifier.

#### b. **Prediction or classification using CNN**

**Fig. 2** Deep learning model with automated detection module



Objects in the movement are extracted with increasing convolutional layers with more solidity. After the training using ImageNet, the convolution layer extracts the moving objects in motion. In addition, it extracts objects that are static in the input video frames in the foreground.

1. Convolutional layers perform the processing operation that is a linear operation when a series of weights is multiplied using the video input data array. A filter is used to scan the entire image, which is referred to as a translation invariance, to detect the saliency from the input image. Since filters are repeatedly used with the input array, a two-dimensional feature map is created.
2. Permutation layers confirm the object classified (predictable objects and unpredictable objects) by the convolution layer, which lists the reference objects in the order in a permutation-based representation. The similarity is



estimated by using the two permutations instead of using the distance function. Between the reference object and the ground truth. This relationship tends to be closest to that shown by distance.

### c. **Planning using RNN**

The regular operations of RNN for planning to drive the self-driving cars are thus given below:

1. Batch normalization reduces the number of hidden layers required for processing the input features from CNN. It adjusts the scaling and activation function to speed up the learning process.
2. The batch normalization lowers the number of hidden layers necessary for processing CNN inputs. It adapts the activation and scaling function to speed up the process of learning.
3. Softmax Activation produces a vector that shows a list of potential outcomes in the probability distribution.
4. Maximum pooling is a packing process that calculates the maximum or largest value in each patch of every feature map, where the packing and permutation layers are processed.
5. Dropout unit ignores many hidden units on the forward pass, which are likely to drop the individual nodes so that a reduced network is left; the entry and output edges of a dropped node, on the other hand, are also removable. These operations are performed to prevent overfitting.
6. To address the gradient problem, the bidirectional gated recurrent unit is used to make operation more rapid and efficient.

The RNN video frames are first converted to one-dimensional image sequences by directional scanning of a two-dimensional video frame. The regional sequence is then used to train the RL to learn in motion video temporal and spatial objects. With feedback loops, the learning strategy is increased, as the network can remember past data. This, however, leads to a memory space limit, as the amount of video data increases. The corresponding functions are updated and the past sequences are ideally remembered. This shows a significant observation between objects of the past and present. Recurrent layers are employed to learn the relationship between areas in the video frames.

## **4 Results and Discussions**

The code is written in Python using the Pytorch framework, so basic knowledge of the language and the framework is recommended. The coding is executed on a Python programming language and CUDA. The operating system requirements involve the Windows 10 operating system with Intel Core i5 processor, 16 GB of RAM, and NVIDIA GTX 1070 GPU 8 GB.

The datasets are collected from BDD100K: A large-scale diverse driving video database [15]. The entire system is trained involving CNN with RNN over several iterations, such that the proposed model achieves minimum error.

The proposed method is evaluated against existing CNN and RNN models in terrain, traffic and free roads. The recall results in all these three environments show that the proposed hybrid model attains improved self-driving performance than existing CNN and RNN methods. The models are trained with various objects that include traffic symbols, type of cars, road lines, obstacles, and other unpredictable behaviors of vehicles, animal/human crossings and road blockages.

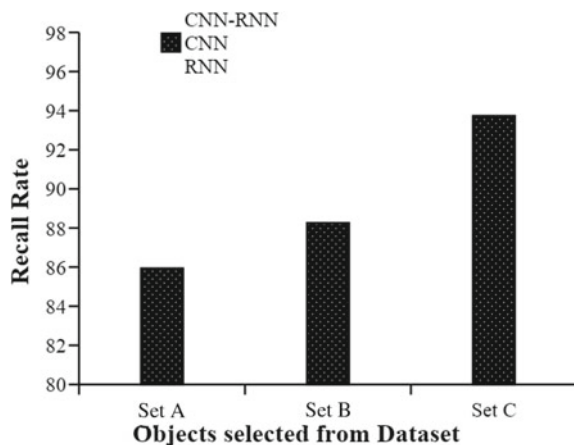
The study is evaluated under three different settings that include **Set A**: Traffic symbols, type of cars and other unpredictable vehicles behavior. **Set B**: Road lines, obstacles, and other unpredictable vehicles behavior and road blockages. **Set C**: traffic symbols, type of cars, road lines, obstacles, and other unpredictable behaviors of vehicles, animal/human crossings and road blockages.

Figures 3, 4 and 5 show the testing recall results in a terrain environment, dense traffic roads and less traffic roads after 1000 runs of training the hybrid model with training datasets. The results show that the proposed method has a higher recall rate than the existing CNN or RNN deep learning classifier. The evaluation of set A, B and C shows that the set C with an increased number of features has higher recall rate, i.e., with increasing features, the recall results show more optimal results in proposed and in existing classifiers. The increased performance in the proposed method is due to the elimination of redundant datasets at the initial screening level and use of individual deep learning classifiers for prediction and planning stages.

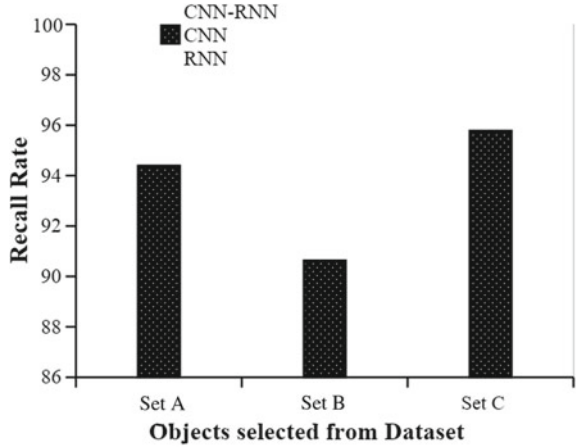
The proposed system is tested over 50,000 test images to predict the performance of the model. The testing process does not consider discrete labels; hence, the study uses root mean square error (RMSE) to evaluate the hybrid model performance on all the set A, B and C.

The results of Figs. 6, 7 and 8 show the RMSE of test results in terrain environment, dense traffic roads and less traffic roads. The results of RMSE between the proposed

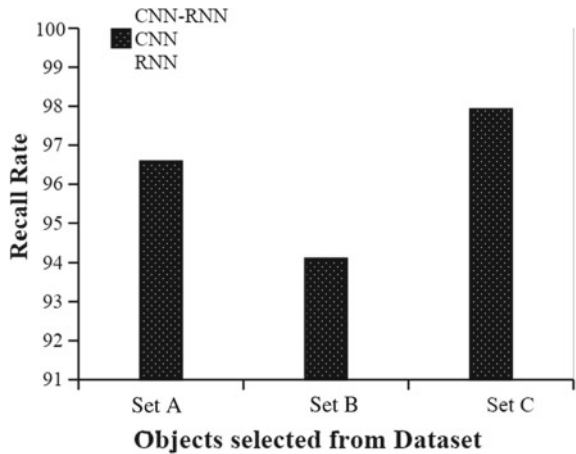
**Fig. 3** Testing recall results in terrain environment after 1000 runs of training the hybrid model with training datasets



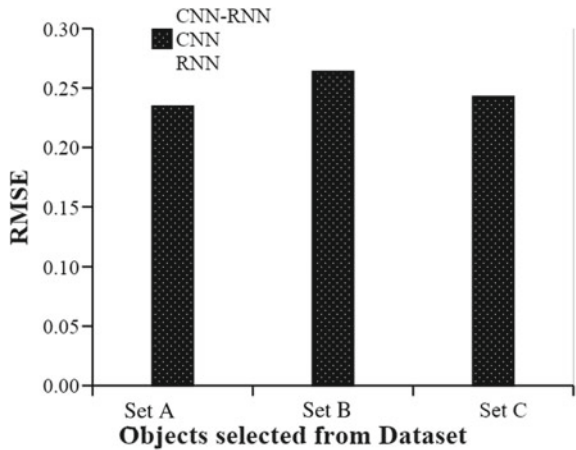
**Fig. 4** Testing recall results in dense traffic roads after 1000 runs of training the hybrid model with training datasets



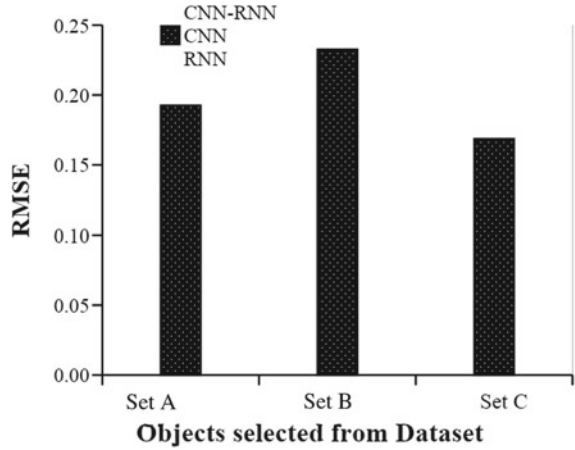
**Fig. 5** Testing recall results in less traffic roads after 1000 runs of training the hybrid model with training datasets



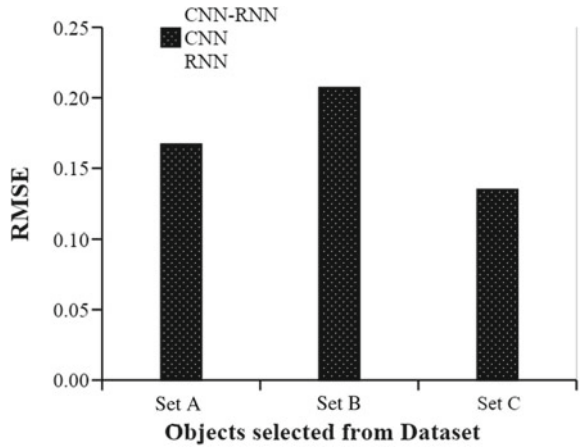
**Fig. 6** RMSE of test results in terrain environment



**Fig. 7** RMSE of test results in dense traffic roads



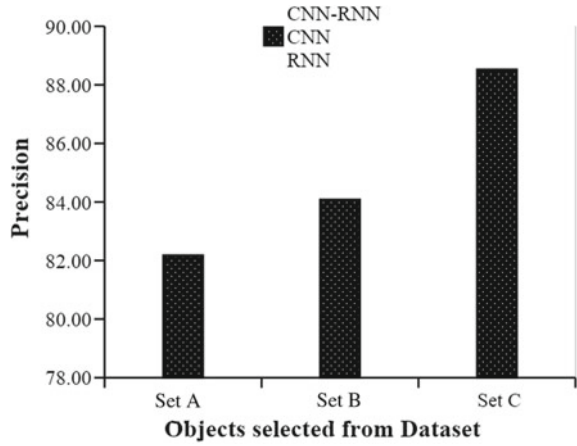
**Fig. 8** RMSE of test results in less traffic roads



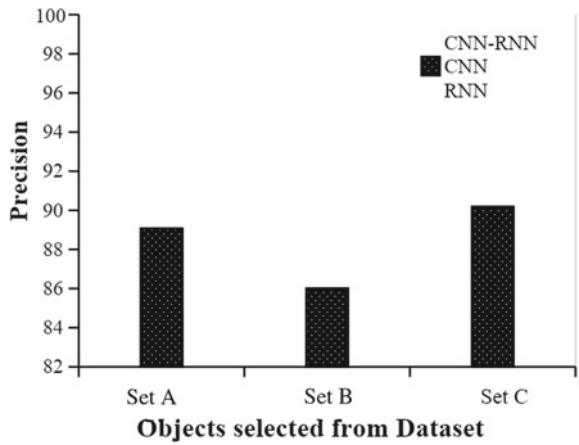
CNN-RNN and existing CNN and RNN show that the proposed method achieves reduced RMSE rate for Set C than Set A and Set B, where Set B has higher RMSE than all the other methods. On the other hand, the proposed method offers reduced RMSE in less traffic roads than high dense roads and terrain environment.

Figures 9, 10 and 11 show the testing precision results in a terrain environment, dense traffic roads and less traffic roads after 1000 runs of training the hybrid model with training datasets. The results show that the proposed method has a higher recall rate than the existing CNN or RNN deep learning classifier.

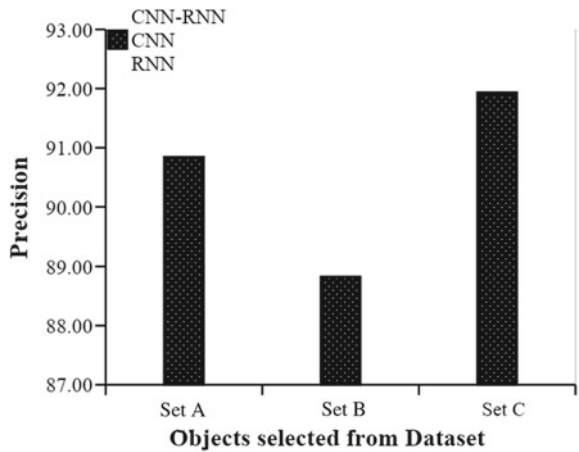
**Fig. 9** Precision of the proposed model on terrain environment



**Fig. 10** Precision of the proposed model on dense traffic roads



**Fig. 11** Precision of the proposed model on less traffic roads



## 5 Conclusions

In this paper, an autonomous hybrid deep learning model is developed to improve the smooth driving of cars in rugged and terrain environment and high traffic conditions. The hybrid deep learning methods using CNN and RNN for optimal planning and prediction in autonomous driving behavior facilitate the improved performance of cars. The experimental results from the available datasets have proven the optimal efficacy of the model with accident-free driving. Further, eradicating numerous or redundant labelled datasets at the initial recognition stage has effectively improved the decision-making ability of both CNN and RNN. The outcome has shown that the proposed method using CNN for planning and RNN for prediction has improved the capability of vehicle driving with optimal detection of the object from trajectories, classifying it into predictable and unpredictable behaviors and finally, the prediction of vehicle movement. In future, the study can be improved by reducing the redundancy at the recognition stage by considering the spatial and temporal features of the video, thereby the labelled datasets can be limited at the initial stage, especially in case of level 5 autonomous vehicles. The design of the present driving behavior on level 5 autonomous vehicles can be considered for future endeavors.

## References

1. Schwarting W, Alonso-Mora J, Rus D (2018) Planning and decision-making for autonomous vehicles. *Ann Rev Control Robot Auton Syst*
2. Surden H, Williams MA (2016) Technological opacity, predictability, and self-driving cars. *Cardozo L Rev* 38:121
3. Van Engelen JE, Hoos HH (2020) A survey on semi-supervised learning. *Mach Learn* 109(2):373–440
4. Nugraha BT, Su SF (2017) Towards self-driving car using convolutional neural network and road lane detector. In: 2017 2nd International conference on automation, cognitive science, optics, micro electro-mechanical system, and information technology (ICACOMIT), pp 65–69. IEEE
5. Cai Z, Fan Q, Feris RS, Vasconcelos N (2016) A unified multi-scale deep convolutional neural network for fast object detection. In: European conference on computer vision, pp 354–370. Springer, Cham
6. Wu B, Iandola F, Jin PH, Keutzer K (2017) Squeezenet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 129–137
7. He S, Lau RW, Liu W, Huang Z, Yang Q (2015) Supercnn: a superpixelwise convolutional neural network for salient object detection. *Int J Comput Vision* 115(3):330–344
8. Fan J, Xu W, Wu Y, Gong Y (2010) Human tracking using convolutional neural networks. *IEEE Trans Neural Netw* 21(10):1610–1623
9. Wang N, Yeung DY (2013) Learning a deep compact image representation for visual tracking. In: Advances in neural information processing systems, pp 809–817
10. Raj JS, Ananthi JV (2019) Recurrent neural networks and nonlinear prediction in support vector machines. *J Soft Comput Paradigm (JSCP)* 1(01):33–40
11. Cui Z, Xiao S, Feng J, Yan S (2016) Recurrently target-attending tracking. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1449–1458

12. Ning G, Zhang Z, Huang C, Ren X, Wang H, Cai C, He Z (2017) Spatially supervised recurrent convolutional neural networks for visual object tracking. In: 2017 IEEE international symposium on circuits and systems (ISCAS), pp 1–4. IEEE
13. Elman JL (1990) Finding structure in time. *Cogn Sci* 14(2):179–211
14. Koresh MHJD, Deva J (2019) Computer vision based traffic sign sensing for smart transport. *J Innov Image Process (JIIP)* 1(01):11–19
15. BDD100K: a large-scale diverse driving video database. Available at: <https://bair.berkeley.edu/blog/2018/05/30/bdd/>. Accessed on 15.03.2020

# Video Enhancement and Low-Resolution Facial Image Reconstruction for Crime Investigation



Joel Eliza Jacob and S. Saritha

**Abstract** In video examination, image or video quality improvement is an emanant area of research. Images and videos acquired by the imaging sensors have rich and detailed information. The viewable representation of low-quality videos can be improved using video enhancement techniques. A video enhancement and reconstruction framework for facial images with low resolution are proposed in this paper. The video enhancement phase improves the appearance of images that are not sharply defined. In the reconstruction phase, the facial regions are selected from the high-quality video frames and the resolutions of these selected facial regions are increased. After reconstructing the facial regions, the aim is to authenticate and verify the human face through a face recognition network. A detailed analysis of the experiments is reported and is observed that the results obtained are significant.

**Keywords** Convolutional neural networks · Video enhancement · Super-resolution · Face recognition

## 1 Introduction

In crime analysis and evidence acquisition, security cameras, mobile devices, CCTV systems, and other imaging sensors are considered as sources for evidence gathering. Videos and images collected from these sources are mostly used in crime investigation. The video surveillance devices are used as security cameras as well as monitoring systems at homes, shopping centers, and road intersections, parking areas, and banks. The precision of data that is acquired from videos and images depends on video texture and quality. Poor quality video footage can lead to information loss, reducing the efficiency of crime investigation. The quality of video

---

J. E. Jacob (✉) · S. Saritha  
Rajagiri School of Engineering and Technology, Emakulam, India  
e-mail: [elizajjoel@gmail.com](mailto:elizajjoel@gmail.com)

S. Saritha  
e-mail: [saritha\\_s@rajagiritech.edu.in](mailto:saritha_s@rajagiritech.edu.in)



footage can affect many evidence extraction techniques such as object recognition and detection. The accuracy of most of the machine learning tasks which is used to perform video analysis can be improved by enhancing the image appearance and quality [1]. Propose a crime evidence collection and culprit identification technique for poor resolution surveillance videos, by enhancing the quality of video footages.

The video enhancement and reconstruction are used in different fields like video surveillance, remote sensing, medical imaging, traffic management, satellite images, and abnormal activity detection. Some of the important techniques used for video enhancement and reconstruction are histogram equalization [2, 3], retinex theory [4], super-resolution [5], and deep learning techniques [6]. Histogram equalization method uniformly distributes or equalizes the histogram corresponding to each image as explained in [7]. Another method for enhancement is retinex theory which follows the same way of the physical image capturing model and how the human eye responds when it sees an object. It is based on how the human eyes and brain communicate with each other. The retinex-based method is based on a contrast enrichment method which is done using the luminance adaptation technique to refine the sharpness of footages in [4, 8]. Video enhancement and reconstruction can also be achieved by applying a super-resolution method. The super-resolution techniques upscale the resolution of poor resolution images. It involves zooming a particular area in the image where there is a need for a high-resolution image.

In this paper, deep learning models are used for video enhancement and human face reconstruction. Deep learning approaches follow the same concept used for the working of the neurons present inside the brain. Deep learning algorithms are formed by using different layers such as the input layer, output layer, and hidden layers that are linked to each other. Each layer contains neurons that are attached to neurons in the next layer. Neurons that exist in these layers will consume the input data and then process it and will pass it to the layers next to it. The activation function, weight, and bias determine the strength of the neurons in each layer. Large amounts of data are consumed by each neuron and then propagate them to multiple layers. In video analysis for crime identification, there is a need to identify and recognize human faces, suspicious objects, and tools. The problems in object recognition are due to the small area of the object, blurry and dim footages occurred due to lighting and environmental conditions (weather, fog, etc.), and position of the imaging sensors. The significant ideas of this paper include:

- A video enhancement module that enhances the entire video frame by frame and thereby extracting the required information.
- A face reconstruction module involves selecting a particular area of interest in each frame (e.g., human face) and applying the super-resolution methodology to refine the resolution of the selected human facial regions.
- A face recognition module which helps to identify or recognize the person.

Here, efficient techniques for video enhancement and human face recognition in low-quality surveillance videos are emphasized. Section 2 contains the review of works used for enhancement and super-resolution techniques for face recognition in low-quality images. The proposed system architecture and the implementation

details are explained in Sect. 3. The results and analysis of this work are discussed in Sect. 4. Future work and conclusion are given in Sect. 5.

## 2 Related Works

Video enhancement is a fundamental step in many computer vision and image processing tasks. An image enhancement algorithm is introduced in [6] using a CNN-based method. To avoid the gradient vanishing problem, a special module is designed which uses multi-scale feature maps. Here, a special convolutional module is designed which is based on the inception and residual models. The model used in [6] only concentrated on enhancing the content details of the image. In [9], hybrid network is proposed to enhance the appearance of low-quality images. The proposed architecture is based on an encoder–decoder network. Sometimes structural details of the image are lost even though an encoder-based network is used here. To remedy this, and to improve the edge details, a spatially variant recurrent neural network (RNN) is used along with a new autoencoder is introduced in [9]. Here, hybrid network architecture which includes both CNN and RNN is used to perform image enhancement of low-quality images. Also, the adversarial loss functions and perceptual loss are combined with the proposed network to increase the appearance and visual quality of the enhanced results. A deep autoencoder-based approach is introduced in [9] to extract features from lowlight images and enhance images without over-brightening the lighter parts of the images are proposed in [9]. A stacked-sparse denoising autoencoder in [10] is applied to enhance and denoise images. In [11], architecture based on discrete wavelet transformation (DWT) and fully convolutional networks are proposed. Further architecture for image resampling and enhancement net is included in the neural network proposed in [11]. Most of the video enhancement algorithms lead to either over enhancement or under enhancement problems. The video enhancement using histogram equalization causes over brightness and was mainly used for grayscale images only. Another method for enhancement was a variation of histogram equalization known as contrast limited adaptive histogram equalization which led to the loss of natural image features. Many algorithms using deep learning was also proposed to upscale the image or video quality, but most of these deep learning algorithms affected the content details of the image. In this, a video enhancement technique based on the residual dense network is proposed which improves the image quality without affecting the natural image features.

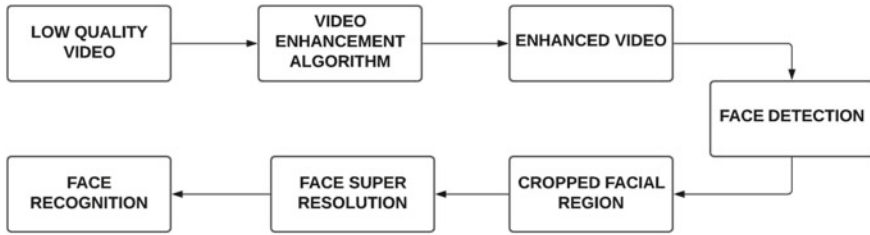
A super-resolution (SR) technique using a deep learning model is detailed in [12]. The model is based on a learning method that includes a mapping of low- and high-resolution images. A connection between the high- and low-resolution images is created using a deep convolutional neural network. The architecture proposed in [12] takes input as low-resolution images. Deep convolutional neural networks are used to improve the performance and accuracy of super-resolution methodologies. But the system proposed in [12] caused computational complexity and cost. [13] introduces a dense deep network using skip connections for image super-resolution. An image

super-resolution mechanism emphasis on a registration algorithm for image reconstruction is proposed in [14]. At first, there is a registration algorithm for the selection of sub-pixel shifted images and then an interpolation-based method is used to attain high-resolution (HR) image. The architecture explained in [14] had a complex structure. A real-time super-resolution of videos is presented using a convolutional neural network (CNN) in [15] extracts feature maps from the low-resolution space. The model introduces the sub-pixel convolution layer, wherein the feature maps of low-resolution images are upscaled to obtain the HR output. This is done by training the upscaling filters. A “sub-pixel motion compensation” layer in a CNN framework proposed in [16] is used in the video super-resolution. In [17], a super-resolution method is introduced using a deep layer that has more performance than those explained in other SR methods such as the method explained in [12]. A video restoration architecture using enhanced deformable convolutions is proposed in [18]. Both the methods described in [17, 18] show better results in upgrading the resolution of the video and image.

Face recognition in low-resolution images is one of the difficulties encountered by the research community. The various factors that affect the resolution of imaging sensors are equipment quality, camera distance, angle, and other environmental factors. Face recognition in low-resolution images is mainly concerned with the identification and verification of facial images in low-resolution videos or images. In surveillance videos, human faces are blurry and are difficult to obtain and thus an alternative technique known as person reidentification is used to match person’s images across cameras. A person reidentification on super-resolution images is proposed in [19]. The model is based on an integrated attention block and residual dense block (RDB). A super-resolution method for face recognition on high dimensional features using unsupervised algorithms is proposed in [20]. A super-resolution technique using interpolation-based methods along with a smooth regression structure for mapping the relationship between low-resolution pixels and missing high-resolution pixels is proposed in [21].

### 3 Proposed System

The architecture of video enhancement and human face recognition in low-quality surveillance videos for crime analysis is explained in this section. The low-quality videos are extracted frame by frame and then applied to a video enhancement model where high-quality frames are generated. These high-quality frames are further combined to produce a video of better quality. The video enhancement model is based on a residual dense network. Residual dense network (RDN) works based on the principle of dense connected convolutional layers. The model based on RDN extracts rich features and enhances the video quality from the acquired low-quality videos. After enhancing the video, target frames are identified and these frames are then used for detecting the suspected human face from the crime scene. For detecting the human face, haar-based classifiers are used. Haar features work on the principle



**Fig. 1** Architecture of the proposed system

of differences of sums of intensity and are used for detecting the face. After identifying the face with the help of a haar classifier and bounding box, the region which includes the face is cropped and given to the next section for further analysis. Then, a super-resolution technique is applied to the cropped facial region to upscale the resolution of the selected area. Super-resolution is a technique that can restore the details of the image and improve image resolution. The basic idea is to reconstruct a target face and then use it for face recognition tasks. The super-resolution techniques can be done using different deep learning models. In this paper, the super-resolution is done by using convolutional layers. After improving the resolution of the selected face, face recognition is done to identify the target. It is a technique that is used for authenticating and identifying an individual from an image. Here, a dataset is created using images of different persons who need to be recognized. Then, this dataset is trained and tested by using a face recognition network. The proposed system is split into three different modules: video enhancement module, face reconstruction module, and the face recognition module. The proposed system is depicted in Fig. 1.

### 3.1 Video Enhancement Module

The video enhancement process includes techniques that increase the visual appearance of the original data before processing. By using the video enhancement method, noise can be removed and increase the sharpness and brightness of the frames. Video enhancement is an image preprocessing method, and the primary objective is to process the image and produce a more enhanced image than the original input image, and this can be applied to several areas such as surveillance systems, medical images, and satellite images. The video enhancement technique used here helps to recognize and identify a criminal or suspect from a low-quality video. The overall architecture of the video enhancement module is given in Fig. 2.

In the video enhancement module, a convolutional neural networks-based model is applied to upgrade the appearance of the given input video. In this, a residual dense network (RDN)-based model is used for enhancing the video appearance. The video enhancement techniques using other CNN-based models cannot extract the

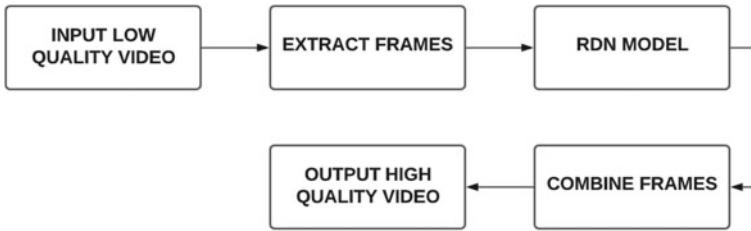


Fig. 2 Video enhancement architecture

hierarchical features of the low-quality videos completely which causes low performance. But, the RDN-based model can extract the hierarchical image features. The network parameters used by the RDN is less in comparison with other techniques, which increased its performance effectively. A video enhancement techniques based on RDN-based model is proposed here for enhancing the video appearance since it provides better performance than other models. Here, six low-quality video datasets are used for performing the experiments. From the low-quality video, the low-quality video frames are extracted. The resolution of the input videos is  $854 \times 48$ . The frame per rate of the videos used here varies from 15 to 25 fps. After extracting the video frames, a pre-trained RDN-based model is loaded for enhancing the video frames. The RDN used here to extract the image features from the low-quality input video frames and then combines the relevant image characteristics from all the layers and produces enhanced video frames. The RDN-based network contains residual dense blocks (RDBs), which contains densely connected layers and local feature fusion. The convolutional layers of current residual dense block connect to the previous residual dense block, resulting in a continuous state pass known as contiguous memory mechanism, which preserves the information that needs to be passed. In each of the RDBs, the information in previous and current residual dense blocks is concatenated using a local feature fusion module to extract local dense features. After identifying the local dense features, the hierarchical features of the image are preserved using a global feature fusion framework. Here, a pre-trained residual dense network for video enhancement is being used. The architecture of the RDN in [22] is given in Fig. 3. The fundamental components of the architecture are (a) D—residual dense blocks (RDB), (b) C—convolutional layers inside the RDBs, (c) G—feature maps of each convolutional layers inside the RDBs, (d) F—extracted features, (e) LQ—low-quality image, and (f) HQ—high-quality image. The RDN-based models generate enhanced

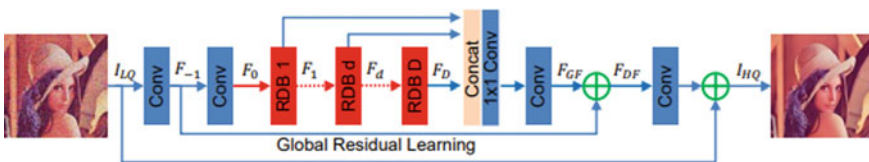


Fig. 3 Residual dense network [22]

high-quality frames from the low-quality frames. After enhancing the video frames, these frames are then combined to form a high-quality video. The high-quality videos obtained are of resolution  $1708 \times 960$ . The high-quality video generated using an RDN-based model will be given as an input to the face reconstruction module.

### 3.2 Face Reconstruction Module

After enhancing the low-quality video, the next process is to reconstruct the human facial image from the video. From the enhanced video, the frames are extracted and target frames are identified. Human faces are detected from the target frames and are then used to recognize the face. The human face detection is done using haar cascade classifiers. The detected facial regions are of low resolution, so the super-resolution technique can be applied for identifying the facial images more accurately. In this work, haar cascade classifiers are used for detecting the face. After detecting the face, the face region is marked using a bounding box and the resolution of the facial region is increased. Super-resolution techniques use deep learning methodologies to add the missing pixels. This project proposes a convolutional neural network layer for image upscaling. After identifying the region which includes the face, the facial region is cropped and given to the model which is created using convolutional layers. The output will be a high-resolution facial image. The entire workflow of face reconstruction is given in Fig. 4.

The model for super-resolution is done using a convolutional neural network which requires mainly three convolutional layers. Then, the model is trained by using a set of images of low and high resolution. The dataset used here is the Celebhq dataset, which includes high-quality images of the human face. Then using this model the high-resolution face is obtained from the low-resolution images. Super-resolution aims to attain a high-resolution output image. The initial step is the preprocessing of the image which includes increasing the resolution of the low-resolution images, using one of the convolutional layers. For extracting the features, a convolution

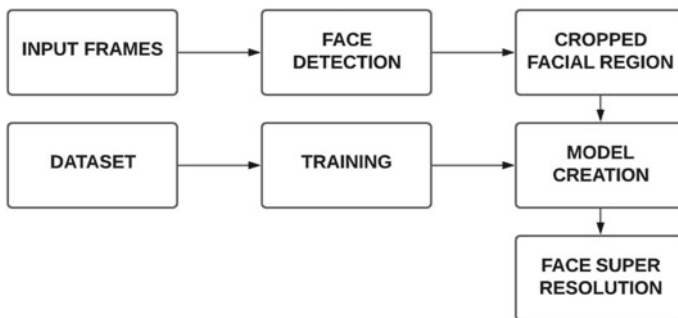


Fig. 4 Workflow of face reconstruction

operation with ReLU is performed. The values in every layer are updated to maximize PSNR or minimize MSE. Finally, in the output phase, the convolution operation is performed to produce a high-resolution image.

### 3.3 Face Recognition

Face recognition systems include identifying or authenticating a human face. The main technique used in the face recognition technology is to compare the distinct facial characteristics of a human image with a custom dataset. The facial landmarks of the facial image can be extracted to identify the unique features of the face and then recognize the face. The facial landmark identifies the position, size, and structure of the nose, eyes, ears, and frontal face features. These features are used for matching the database images. The framework of the face recognition system is given in Fig. 5. The facial recognition method used in this work includes two steps to create encoding and names of faces in the dataset and to recognize faces in the video frames. In this, a dataset is created which includes the facial images. In this, facial images of six persons are used for recognition, and corresponding to each person 10 facial images are taken. The next step is to generate facial encoding using the function `face_encoding()` for each facial image in the dataset. Then for recognizing the faces, from the input video, the facial image is detected and facial encoded are generated. The facial image is compared with the encoded dataset using `compare_faces()`. If facial images in the dataset is a match with the detected face, then return the name of the person along with a bounding box.

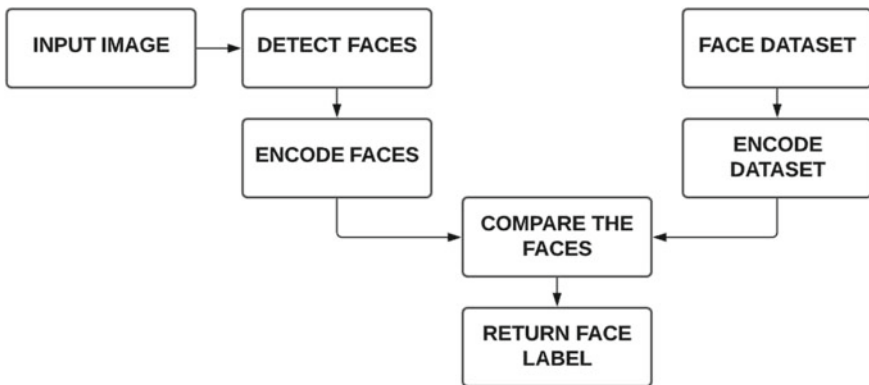


Fig. 5 Face recognition architecture

## 4 Results and Discussions

The results of the work for each module are described in detail in this section. In the video enhancement module, six videos were taken; from these videos, the frames are extracted. The resolution of the videos taken is of resolution  $854 \times 480$ . The extracted frames are given to the residual dense network. The output enhanced frames obtained after passing through the RDN model is of resolution  $1708 \times 960$ . Sample video frames are shown in Fig. 6. The input frames of low quality are depicted in Fig. 6a, b, and the high-quality frames after passing through the residual network are given in Fig. 6c, d.

To establish the merits of the proposed method, the results are compared with techniques like histogram equalization (HE) and contrast limited adaptive histogram equalization method (CLAHE) which are reported in the literature. The comparison of the above said enhancement methods are shown in Fig. 7.

The video enhancement process is further analyzed using image quality metrics. The metrics used for this evaluation are peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), mean squared error (MSE).

Here, six video datasets prepared using camera resolution ranging from 13 to 5 MP are taken for study, and different enhancement techniques are applied to these videos for improving its quality. The frame rate of the videos ranges from 15 to 25 fps. Using a chosen reference frame, the image quality is calculated for each technique. The average PSNR value for histogram equalization is 11.8, contrast limited histogram equalization is 10.1, and RDN is 14.3. The PSNR value using RDN is larger. Therefore, the enhancement technique using RDN has better quality. The average SSIM value for histogram equalization is 0.41, contrast limited histogram equalization is 0.28, and RDN is 0.55. The SSIM value is used to check the similarity of the input frame with the enhanced results. Here, RDN-based technique used for enhancement is more similar to the input frame. The average MSE value for HE is 16,640, CLAHE is 21,509, and RDN is 8726. The MSE value is less for the RDN-based technique, so it gives a better outcome when compared with other techniques. The entire result analysis is presented in Table 1.

The face detection module is used to detect regions that include human faces. Here, haar cascade classifiers are used for identifying the faces. The face super-resolution is done to increase the resolution of the detected face. After cropping the facial region, it is used to apply the super-resolution technique. Here, five video datasets of camera resolution are being chosen ranging from 13 to 5 MP which is enhanced using the RDN, as the input and frame per rate range from 15 to 25 fps. From each input video frame, the face area is detected and cropped and then given to the super-resolution network. The results are presented in Fig. 8.

The image quality metrics are used for analyzing the quality of facial images after applying the super-resolution technique. For evaluating the image quality, a facial image of each person is kept as the reference image. The results of this analysis are given in Table 2. Here, the average PSNR value of the low-resolution faces is 14.38 and the high-resolution face is 25.9. So, the PSNR value of high-resolution face is



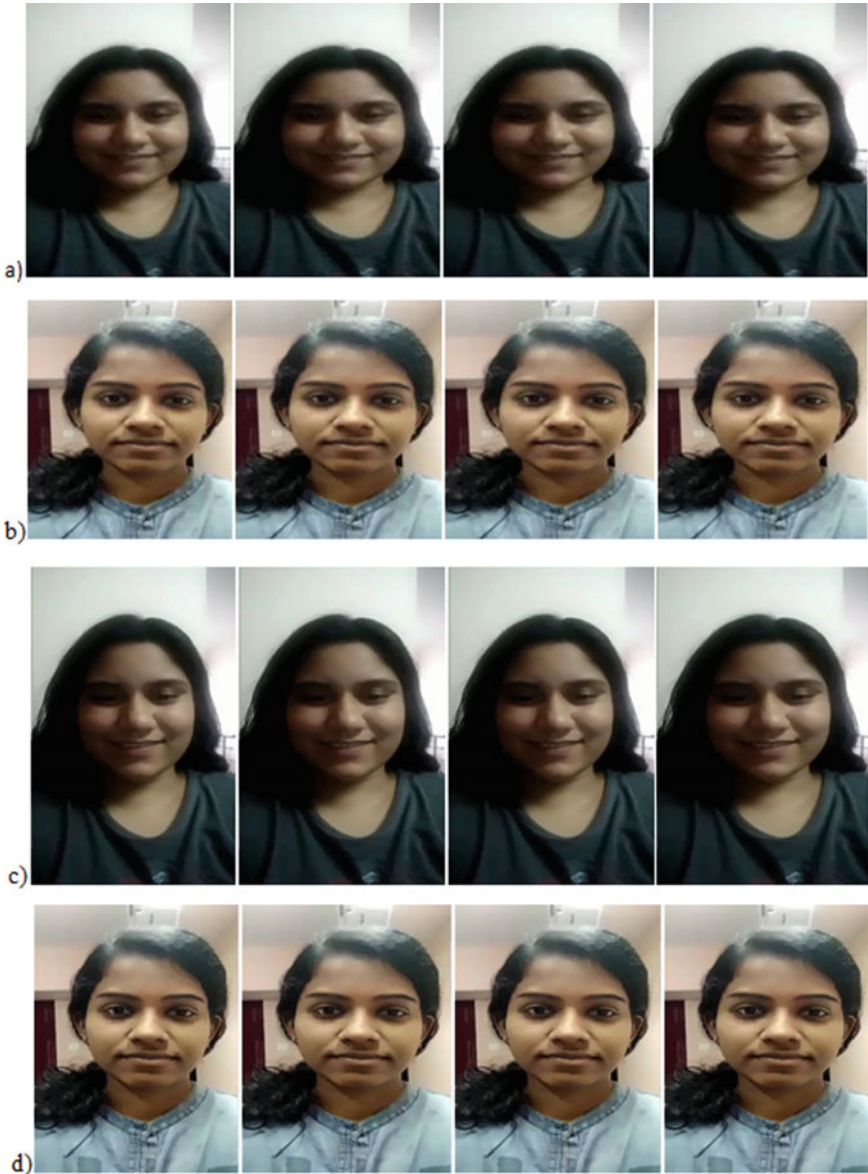
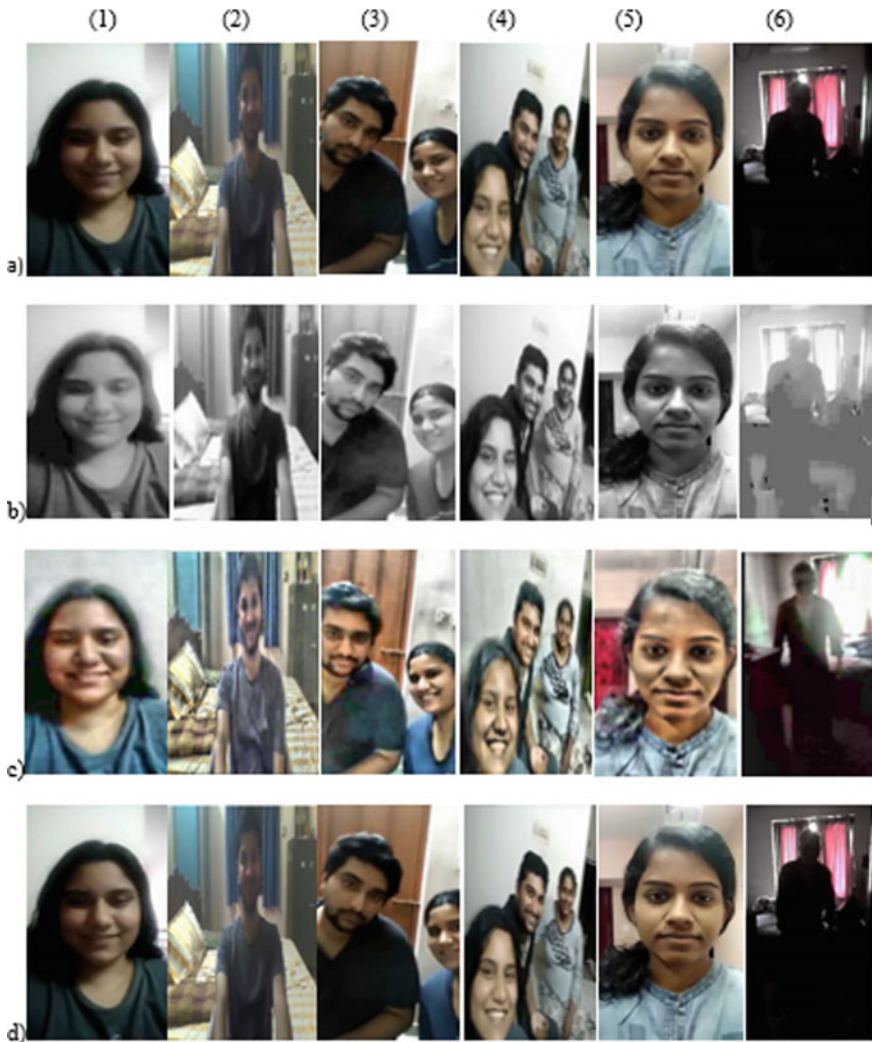


Fig. 6 Video enhancement a, b input low-quality frames. c, d High-quality frames



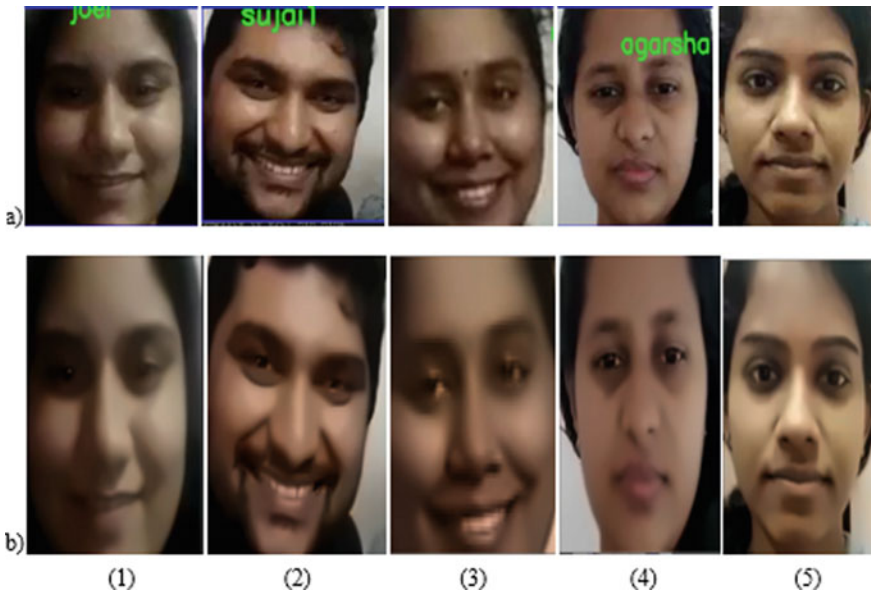
**Fig. 7** Comparison of different video enhancement techniques. **a** Input frames, **b** histogram equalization, **c** contrast limited adaptive histogram equalization, and **d** residual dense networks

larger. Thus, it has more quality. The average MSE value of low-resolution faces is 5449.6 and high-resolution faces are 519.53. The MSE value of the high-resolution face is smaller, so it has better quality. The average SSIM value of low-resolution faces is 0.54 and high-resolution faces are 0.94. So, high-resolution faces have more quality.

The accuracy and loss curves of the super-resolution model are also presented in Fig. 9. The accuracy curve includes training accuracy and validation accuracy. The

**Table 1** Image quality metrics for video enhancement

Video	Frame rate	PSNR			SSIM			MSE		
		HE	CLAHE	RDN	HE	CLAHE	RDN	HE	CLAHE	RDN
1	24	11.9	10.8	19.02	0.37	0.27	0.79	12,313	16,091	2443
2	15	13.9	15.7	15.81	0.53	0.44	0.57	8319	5142	5116
3	20	11.5	7.9	11.93	0.55	0.15	0.56	13,515	31,462	12,503
4	25	13.4	8.6	13.6	0.56	0.11	0.58	8838	26,589	9637
5	22	14.3	9.8	14.57	0.35	0.15	0.36	7246	20,303	6808
6	14	5.9	8.2	10.9	0.13	0.18	0.47	49,611	29,468	15,849
Average		11.8	10.1	14.30	0.41	0.28	0.55	16,640	21,509	8726



**Fig. 8** Face regions for image quality analysis **a** input low-resolution faces, **b** face super-resolution

training accuracy is approximately 92%. The loss curve includes the details about the training loss and validation loss.

For the face recognition module, a dataset is prepared which includes images of different persons. The results are depicted in Fig. 10.

To analyze the face recognition results, measures such as precision, recall, and accuracy can be used. From Table 3, the overall accuracy of the face recognition network is 71.42%. The accuracy is found by calculating the number of correct recognitions divided by the overall recognitions. The precision is found to be 77.65% and recall is 87.95%. The average F1 score of the model is 82.35%.

**Table 2** Image quality analysis

Face	Frame rate	PSNR		MSE		SSIM	
		Low-resolution face	High-resolution face	Low-resolution face	High-resolution face	Low-resolution face	High-resolution face
1	24	15.35	23.07	5686	831.28	0.67	0.94
2	25	12.62	27.44	1066	351.30	0.37	0.94
3	25	13.40	26.12	8906	476.17	0.48	0.93
4	22	14.90	24.92	6311	627.13	0.62	0.93
5	14	15.67	27.96	5279	311.8	0.59	0.96
Average		14.38	25.9	5449.6	519.53	0.54	0.94

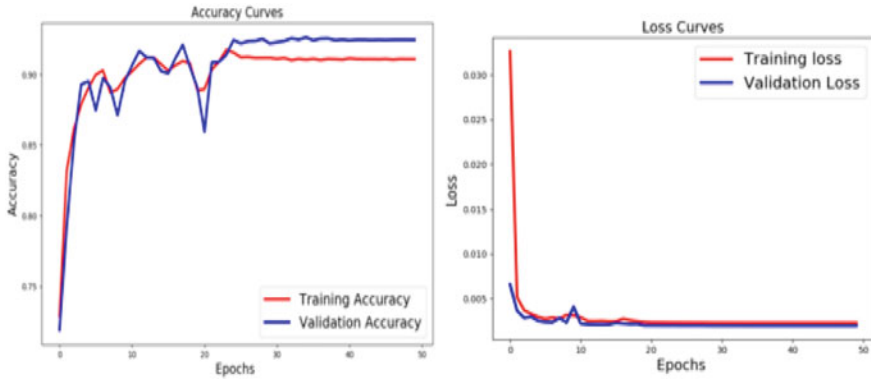


Fig. 9 Accuracy and loss curves

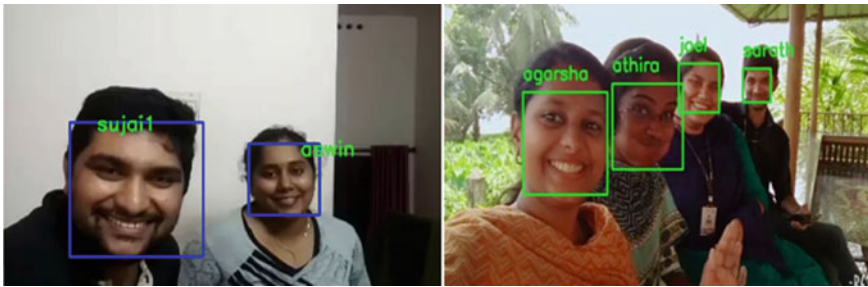


Fig. 10 Face recognition results

Table 3 Face recognition analysis

Video	Frame rate (frames/second)	Precision (%)	Recall (%)	Accuracy (%)	F1 score (%)
1	15	76.19	88.88	69.56	82.04
2	14	76.92	90.09	71.42	82.58
3	15	78.94	88.23	71.40	83.32
4	15	78.57	84.61	73.33	81.47
Average		77.65	87.95	71.42	82.35

## 5 Conclusions

The video enhancement is a preprocessing technique in image processing which is used to refine the video quality. The quality and informative content of original data can be improved before processing using the video enhancement techniques. Here, a technique is proposed to enhance a low-quality video and reconstruct a selected region. The video enhancement technique includes a residual dense network that is

used to generate a high-quality video. Here, the enhancement method showed better results for scenarios that include medium low light images, but the performance is less while enhancing extremely dark images is a drawback of the proposed system. In the future, this module can be extended in enhancing the extremely dark videos, night-vision surveillance videos, and enhancement during unfavorable weather conditions. The enhancement techniques used here can also be used in different applications such as traffic management, autonomous self-driven vehicles, medical images, and satellite imagery. After enhancing the high-quality video, the videos are converted into frames. From these frames, target frames are identified and these frames are used for facial reconstruction. The face reconstruction module includes face detection and super-resolution. The face detection is done using a haar cascade classifiers. Super-resolution is done using a network build using convolutional layers. The super-resolution is used to increase the resolution of the selected area. One of the limitations of this technique is that the efficiency of the super-resolution model is less in smaller regions. Multiple face super-resolutions can be done in the future so that the overall efficiency of this system can be improved. After reconstructing the facial image, this image can be used for face recognition to identify the face of an individual. In the future, face recognition can also be attempted using deep learning models, so that the face recognition process may speed up and performance is improved. Here, a dataset is created which includes images of different persons. Instead of creating a dataset, in future usage, the social network image dataset can be tried. By using a social media dataset, the number of persons whose images can be used for face recognition can be increased.

## References

1. Xiao J, Li S, Xu Q (2019) Video-based evidence analysis and extraction in digital forensic investigation. *IEEE Access*
2. Banik PP, Saha R, Kim K-D (2018) Contrast enhancement of low-light image using histogram equalization and illumination adjustment. In: 2018 international conference on electronics, information, and communication (ICEIC), 24–27 January 2018
3. Sahoo S, Panda J, Mohanty MN (2016) Performance analysis of HE methods for low contrast images. In: 2nd international conference on intelligent computing, communication and convergence (ICCC-2016)
4. Fu Q, Jung C, Xu K (2018) Retinex-based perceptual contrast enhancement in images using luminance adaptation. *IEEE Access*
5. Niu X (2018) An overview of image super-resolution reconstruction algorithm. In: 11th international symposium on computational intelligence and design
6. Tao L, Zhu C, Li Y (2018) LLCNN: a convolutional neural network for low-light image enhancement. *IEEE Vis Commun Image Process (VCIP)*
7. Nithyananda CR, Ramachandra AC (2016) Review on histogram equalization based image enhancement techniques. In: International conference on electrical, electronics, and optimization techniques (ICEEOT)
8. Parihar AS, Singh K (2018) A study on Retinex based method for image enhancement. In: Proceedings of the second international conference on inventive systems and control (ICISC 2018)

9. Ren W, Liu S, Ma L, Xu Q, Xu X (2019) Low-light image enhancement via a deep hybrid network. *IEEE Trans Image Process*
10. Lore KG, Akintayo A, Sarkar S (2017) LLNet: a deep autoencoder approach to natural low-light image enhancement. *ACM Pattern Recogn* 61
11. Guo YH, Ke X, Ma J, Zhang J (2019) A pipeline neural network for low-light image enhancement. *IEEE Access*
12. Dong C, Loy CC, He K, Tang X (2016) Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell* 38(2)
13. Tong T, Li G, Liu X, Gao Q (2017) Image super-resolution using dense skip connections. In: *IEEE international conference on computer vision (ICCV)*
14. Dong L, Jin J, Jiang Y, Zhang M, Xu W (2019) Selection-based subpixel-shifted images super-resolution
15. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *Comput Vis Pattern Recogn*
16. Tao X, Gao H, Liao R, Wang J, Jia J (2017) Detail-revealing deep video super-resolution. *Comput Vis Pattern Recogn*
17. Lim B, Son S, Kim H, Nah S, Lee KM (2017) Enhanced deep residual networks for single image super-resolution. In: *IEEE conference on computer vision and pattern recognition workshops (CVPRW)*
18. Wang X, Chan KCK, Yu K, Dong C, Loy CC (2019) EDVR: video restoration with enhanced deformable convolutional networks. *Comput Vis Pattern Recogn*
19. Qin Z, He W, Deng F, Li M, Liu Y SRPRID: pedestrian re-identification based on super-resolution images. In: *Deep learning: security and forensics research advances and challenges*. *IEEE Access*, vol 7
20. Ma F, Jing X-Y, Yao Y, Zhu X, Peng Z (2019) High-resolution and low-resolution video person re-identification: a benchmark. *IEEE Access*
21. Jiang J, Chen C, Ma J, Wang Z, Wang Z, Hu R (2017) SRLSP: a face image super-resolution algorithm using smooth regression with local structure prior. *IEEE Trans Multimedia* 19(1)
22. Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y (2018) Residual dense network for image super-resolution. In: *IEEE/CVF conference on computer vision and pattern recognition*, December 2018

# Facial Emotion Recognition System for Unusual Behaviour Identification and Alert Generation



P. Sarath Chandran and A. Binu

**Abstract** Facial emotion recognition using expression analysis of human face is one of the important areas in video analysis which uses deep learning concepts to predict the state of mind of a person. Human facial expressions include means of communication through body language, eye behaviour and gestures. Facial emotion recognition discussed in this paper deals with the analysis of the state of mind of a person through his/her facial expressions. The main aim is to propose an unusual behaviour identification and alert generation system using facial expressions. The primary step in facial emotion analysis is to identify faces. After detecting the face, the expression analysis of the person is done by using a convolutional neural network. Through this analysis, unusual behaviour can be identified and an alert is sent to the corresponding authorities so that it can be useful for them in case if any incidents happen.

**Keywords** Video analysis · Deep learning · Emotion classification · Simple mail transfer protocol

## 1 Introduction

Facial expression analysis is the method of classifying one's emotional state based on facial expressions. Facial expression plays a dominant role in communication between two human beings and can also strengthen the relationship between them. The emotional state of the person may affect the decision-making skills, concentration and affect the activities of a person. The frame of mind of a person can be determined by the expressions of that person. Facial expressions are resulted due to movements of the muscles under the skin. The movement of these muscles causes

---

P. Sarath Chandran (✉) · A. Binu  
Rajagiri School of Engineering and Technology, Ernakulam, India  
e-mail: [sarathcn53@gmail.com](mailto:sarathcn53@gmail.com)

A. Binu  
e-mail: [binu\\_a@rajagiritech.edu.in](mailto:binu_a@rajagiritech.edu.in)



folds and lines and causes movement of mouth and eyebrows which create a facial expression. A person's eye plays a vital part in the estimation of their mental state; for example, how happy a person is or how stressed. The analysis of human expressions is applied in different fields such as video surveillance camera, safe driving, intelligent family robot, effective biometric identification, virtual games and expression classification. The emotional states of a person are generalized into seven main categories such as fear, happy, surprise, anger, sadness, neutral and disgust. Deep Learning is a branch of artificial intelligence which tells the computer to do the exact copy of what human does. The different types of deep learning neural networks are convolutional neural networks, feed-forward neural networks and recurrent neural networks. A machine learning technique is discussed in [1] for expression analysis. Here, characteristics such as eyes, eyebrows, nose and mouth are identified from a facial image and then morphological operations and edge detection technique is applied to identify the feature vectors. A feed-forward neural network is used to recognize the expression and classify the emotions into different types [2]. Provide a study of expression analysis based on videos footages. One of the major applications of analysis of mental state of the humans is abnormal behaviour detection. Through this, any abnormal activities can be identified that has occurred. In this project, an ATM scenario is being taken and how a suspect is identified in an ATM based on his/her behaviour. Here, deep learning-based convolutional neural networks are used for abnormal behaviour detection. The main provisions included in this paper are:

- A video pre-processing module uses a heat map for visualizing the area of interest and improves the prediction quality.
- A facial expression analysis module identifies faces of humans and creates a model which extracts facial attributes to classify the emotion.
- Abnormal behaviour alert generation module which sends an alert if any unusual behaviour is detected.

## 2 Literature Survey

Rajesh Kumar et al. [3] described a method which will predict human emotions from videos using deep convolution neural network (CNN). The proposed system in [3] analyses several emotional states (happiness, anger, sadness, surprise, disgust and fear) in a face [4]. Fatemeh Noroozi et al. introduced an emotion classification system, based on audio and visual analysis. In the visual analysis, first, facial landmarks are calculated, i.e. angles and distance. In [4], using a CNN-based model, the emotional states of different individuals can be summarized from the videos. [5] suggests a methodology that is used to classify human behaviour. In [5], the different steps include feature extraction, face identification and classification of emotion. [6] proposes a method which uses a combination of local binary pattern (LBP) features and grey pixel value along with principal component analysis (PCA). In this, a softmax regression classifier is used to classify the emotions based on the six categories. Nimish Ronghe et al. define a CNN-RNN hybrid architecture in [7]

that models the spatiotemporal and spectral characteristics of emotions over variable time and predicts correct emotional reaction. In addition, it also employed a simple support vector machine (SVM) pipeline over time and frequency domain features for emotion recognition from speech in a video. Kah phooi seng et al. [8] classify the emotions into six categories (angry, happy, sad, surprise, fear and disgust) using an audio–visual emotion recognition system.

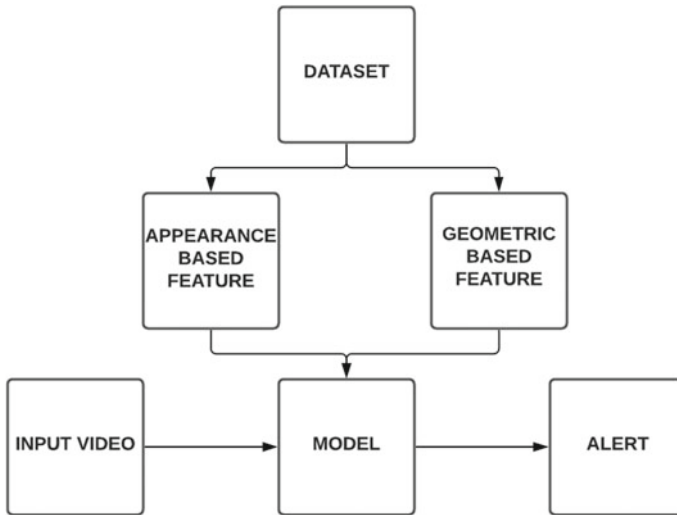
Abdoul Matine Ousmane et al. [9] Propose a system based on deep CNNs. Here, the OpenCV library and Viola and Jones algorithm are applied for detection of faces. Ji-hae-kim et al. [10] introduce a recognition system for emotion analysis based on hierarchical deep learning. The facial attributes are obtained using LBP-based features and facial landmarks. Heechul Jung et al. [11] put forward a deep network based on two different models. From the image sequences, the initial deep network extracts appearance features and the second deep network attains geometry coordinates of a face from facial landmarks. The features from these two models are fused to speed up the efficiency of the facial expression classification [12]. Also, a joint fine-tuning method for combining the two networks to achieve performance improvement in means of recognition rates is presented. Biao Yang et al. [13] Propose a weighted mixture deep neural network which extracts features used for FER tasks, processes greyscale images and local binary pattern of facial images. Different pre-processing mechanisms, like rotation rectification and face detection are implemented. A partial VGG-16 fine-tuning is used in extracting the key points of the facial expression-related attributes of greyscale images. A shallow convolutional neural network (CNN) built based on DeepID extracts LBP-based facial expressions. The result of emotion recognition is calculated using a softmax classification.

In [7], an expression recognition system using a hybrid CNN-RNN architecture for emotion analysis is mainly focused on which can recognize emotion in a video and predict the emotions. The facial emotion recognition system proposed in this paper outperforms most of the existing emotion recognition methods. Most of the existing systems proposed in the literature survey classify the emotions into six basic emotions like anger, disgust, fear, happy, sad and surprise. Here, the emotion is being classified into seven categories such as happy, sad, anger, neutral, surprise, fear and disgust. The existing models used for face recognition do not apply to real-time analysis and have less accuracy compared to the model proposed in this paper.

### 3 System Architecture

In this project, a facial expression analysis-based model has been used to identify an abnormal behaviour from the analysis of that individual's facial expression. The architecture of the proposed work is shown in Fig. 1.

In this architecture, from the data set, the appearance-based and geometric-based features of the human face are extracted. The data set used here is a FER2013 data set which includes labelled emotions and its corresponding pixel values. A model is trained using these features. A CNN-based network is used to create a



**Fig. 1** System architecture

model for emotion classification. From the video, the initial stage is to identify and recognize the facial area. After identifying the target face, it is cropped and used for expression classification. Face detection and identification can be done using haar cascade classifiers. After detecting the facial region, appearance-based feature and geometric-based features are extracted and are given to the CNN structure for training. After training, the CNN model will predict emotion for facial expression recognition. Based on the prediction, unusual behaviour alerts are sent via e-mail if the predicted expressions are sad, anger or fear. The entire framework of the proposed model can be split into three submodules.

### ***3.1 Video Pre-processing***

In this project, the initial step is to load the input video. Here, the input video which is loaded is converted into video frames. Each frame is processed for getting better visualization and analysis results. Video pre-processing is the method of enhancing and processing video frames. In this project, the video frames are enhanced at the time of emotion prediction. The different methods included in the video pre-processing are heat map generation, kernel sharpening and image denoising. A heat map is a representation of two-dimensional information which are represented by using colours. A heat map can be created on the image using the function `cv2.applyColorMap()` available in OpenCV. Kernel sharpening extracts the image features by increasing the contrast of the darker and brighter regions of the image. Image denoising is

obtained by removing the noise from the images which are blurred due to noise and artifacts, thereby improving the appearance of the image.

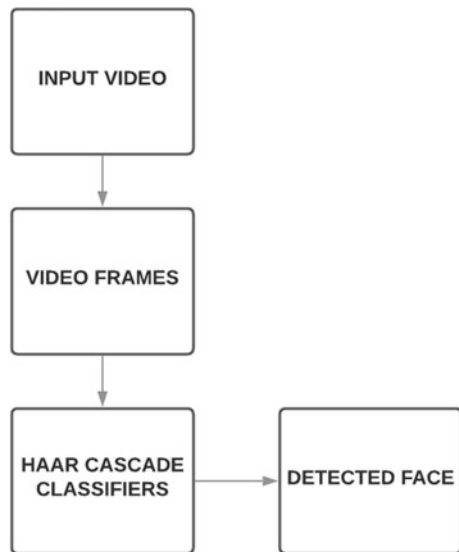
### 3.2 Facial Expression Analysis

In this module, the main aim is to identify the faces from the video and then create a facial expression model to classify the expression.

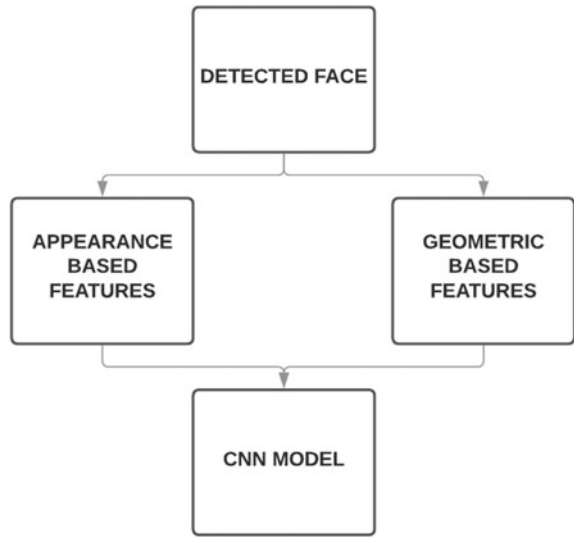
**Face Detection** In this module, from the given input video, the faces are detected using bounding boxes. Detection of the face is done using haar cascade classifier (Fig. 2).

Haar cascade algorithms are pre-trained classifiers which can be used in detecting faces. It is popular for detecting faces and humans; however, it can be learned to identify any object. The AdaBoost learning algorithm is used with haar cascades and selects some relevant features from a large set to provide accurate results. It is the best detector in terms of speed and reliability. The face detection using haar-like features is based on computing the sum of the difference between white pixels and black pixels. The advantage of this method is the face computation of integral image values. Haar-like features-based object detection gives better results when compared to other object detection techniques like LBP-based feature extraction techniques. The haar cascade-based algorithms can also be used for real-time object detection

**Fig. 2** Face detection architecture



**Fig. 3** Facial expression model



and are more robust to illumination changes. Here, a face detection method is being proposed using haar cascade classifiers.

### **Facial Expression Model**

From the region of interest (i.e. the detected facial image), appearance-based features are extracted. The appearance-based features are extracted using local binary patterns. Local binary pattern (LBP) is a texture operator labels each pixel in an image by thresholding each pixel neighbourhood and results in a binary number. LBPs are used to compute a local representation of texture. Then using the histogram of local binary patterns, the final feature vector is calculated. Along with the appearance-based features, geometric-based characteristics are also obtained. Geometric feature learning methods extract unique geometric attributes from facial images and are constructed by a set of geometric coordinates like points, lines and curves or surfaces. Using a pre-trained shape predictor known as 68-point facial landmark detector, the facial coordinates are identified. Then using the features obtained from LBP and facial landmarks, the facial expression classification such as sad, stress, happy, disgust, surprise, fear and neutral are classified using CNN-based network. The architecture of the CNN-based model is given in Fig. 4 (Fig. 3).

### **3.3 Alert Generation**

A model is created using a CNN-based network. Figure 5 shows a detailed architecture of how an alert is generated. The facial expressions are predicted from the given input video using a CNN-based network. Based on the obtained results, unusual

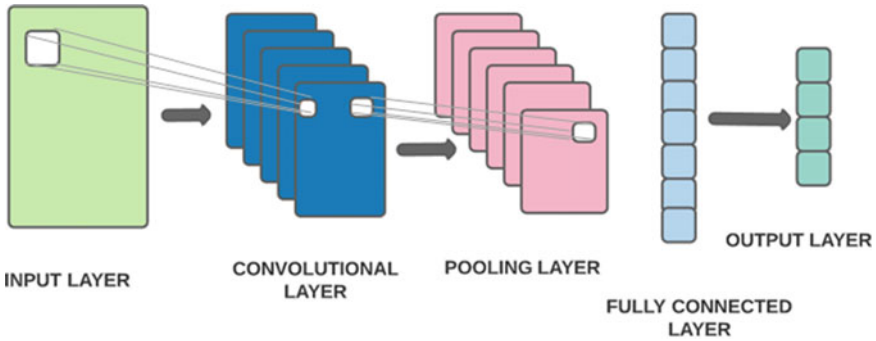


Fig. 4 Convolutional neural network



Fig. 5 Alert generation

behaviour alerts are provided by sending an e-mail to the corresponding authorities. The mail contains the behaviour predicted along with the image of the person. The mail is send using simple mail transfer protocol (SMTP).

## 4 Implementation Details

The implementation details of the project facial expression analysis for abnormal behaviour detection is discussed in this section. Using facial analysis, different emotional condition of the human can be identified and this can be used for identifying the abnormal behaviours. Table 1 shows a comparison of the existing methods for face recognition analysis as mentioned in the literature survey.

### 4.1 Data set Used

The data set used in this project is FER2013. The data set includes facial images of size 48\*48. In this data set, the facial images are categorized and labelled into seven emotions. The csv file used as FER2013 data set includes two columns one for storing the information about emotions and another for storing corresponding pixel values.

**Table 1** Comparison based on existing techniques used for literature survey

	Method	Data set	Advantage	Disadvantage
Rajesh Kumar et al. [3]	CNN	FERC-2013	<ul style="list-style-type: none"> <li>Percentage variations in emotions are useful for exploring micro-expressions</li> </ul>	<ul style="list-style-type: none"> <li>The faces are not detected correctly due to pose variation and alignment of the face</li> <li>Emotions like happy and surprise are often classified as neutral</li> </ul>
Jayalekshmi J et al. [5]	SVM, random classifier, KNN classifier	JAFFE	<ul style="list-style-type: none"> <li>The emotion classifiers based on SVM works well for high-dimensional data</li> </ul>	<ul style="list-style-type: none"> <li>The models used for emotion classification cause overfitting</li> </ul>
Yanpeng Liu et al. [6]	LBP, PCA, softmax regression classifier	Ck+	<ul style="list-style-type: none"> <li>The use of LBP features decreases the computational cost</li> <li>The PCA method used here reduces dimensions, computational and memory cost</li> </ul>	<ul style="list-style-type: none"> <li>Difficult to recognize expression from videos</li> <li>Only classifies six basic emotions</li> </ul>
Ours	LBP, facial landmarks, CNN	FER2013	<ul style="list-style-type: none"> <li>Combination of LBP-based features and facial landmarks increases the accuracy and performance</li> </ul>	<ul style="list-style-type: none"> <li>The emotion 'disgust' cannot be predicted correctly and is often misinterpreted as other emotions</li> </ul>

## 4.2 Video Pre-processing

In video pre-processing, the input video frames are enhanced and appearance details are improved. In this project, a video pre-processing is done during the time of emotion prediction. At the time of prediction, input video is read frame by frame and then each frame is converted into greyscale. Then, heat map normalization, kernel sharpening and image denoising are applied to improve the image quality.

### **4.3 Training**

In this section, the data set is loaded. From the data set, the pixel values of the image corresponding to each emotion are taken as the training data. The facial expression analysis model using CNN-based network is used for emotion prediction. At first, the required packages and data set are loaded. The local binary patterns are obtained by using the function `LocalBinaryPatterns()`. Along with the local binary patterns, the facial landmarks are also extracted. The local binary pattern-based features and facial landmarks are combined and given to the model for training. A CNN-based model is used for model creation. A sequential model is used for adding the layers. A conv2D layer is added with 64 filters, and then batch normalization layer is added and ReLU activation function used. Another 2D convolutional layer, batch normalization and ReLU function are added. Three sets of conv2D with 32 filters, batch normalization and activation function are added. Finally, the softmax layer is used for the prediction. The model created is saved and then used for testing.

### **4.4 Testing**

Here, after creating the model, the next step is to test the model. The CNN-based model is used for emotion prediction. At first, load the required packages and model created. From the given input video, frames are extracted and emotion is detected from the input frames. The model predicts facial expression (angry, sad, fear, happy, surprise, disgust and neutral).

### **4.5 Working**

The working of the project, facial expression analysis, for abnormal behaviour detection is discussed in this section. The data set is loaded, and features are acquired from the data set. The features are extracted based on local binary patterns and facial landmarks. These features are used in training CNN-based model. Finally, the facial expressions are classified as angry, sad, fear, happy, disgust, surprise and neutral. If the facial expressions are identified as unusual, then an alert message is sent via a mail.

### **4.6 Alert Generation**

Through facial expression analysis, different types of emotions are detected. If any unusual behaviour is detected, a mail is sent to alert the authorities. The mail



includes the predicted unusual emotion, and the image of the person whose emotion is predicted. The mail is sent using a simple mail transfer protocol. A smtplib module is used in Python, which is used for sending a mail.

## 5 Results and Discussion

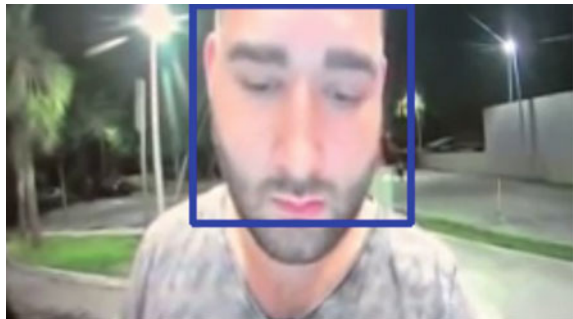
Facial expression analysis for abnormal behaviour detection is a method of analysing human expressions into different categories such as happy, sad, anger, surprise, disgust, neutral and fear. In this project, the ATM-based scenario is being considered. If any abnormal events occurred then the system proposed here will alert the corresponding authorities about the unusual event that has occurred. For this, at first, an input surveillance video is given to the model. From each frame, human faces are identified using haar cascade classifier. After detecting the face, the corresponding emotion of that face is identified. If any unusual behaviour is identified, then an alert message is generated and sent to the corresponding authorities. The human face detection using haar classifiers is given in Fig. 6.

After detecting the human face, the emotions are classified as anger, sad, disgust, surprise, happy and neutral. The video frames are enhanced at the time of prediction. The video chosen for the experiment is of resolution  $640 \times 360$ , and the frame rate is ranging from 7 to 11 fps. The response time analysis for detection is around 30–50 s. The video pre-processing is done using heat map generation that are explained in Fig. 7.

After image pre-processing, the local binary pattern-based feature extractions and landmark-based feature extraction techniques are used to identify the characteristics of the face. The local binary pattern-based feature extraction and landmarks extraction are given in Fig. 8.

The facial features are extracted using local binary patterns and facial landmarks. The model for emotion classification is created by using a CNN-based model. Using this CNN-based model, the emotions are predicted as sad, neutral, happy, angry, surprise and fear (Fig. 9).

**Fig. 6** Face detection





**Fig. 7** Video pre-processing **a** input frame, **b** heatmap and **c** frame after pre-processing

After classifying the emotions, the alerts are generated in case of any abnormal behaviour to send a warning message via mail. The alert mail sent includes the behaviour detected along with the image of the person (Fig. 10).

In this, the efficiency of CNN-based model is evaluated based on the loss and accuracy curves. The accuracy curves and loss curves give an idea about how effectively the model is trained and tested. From these curves, how well the model is learned based on underfit, overfit and good fit can be identified. From the accuracy curve, the training accuracy and validation accuracy can be determined. The training accuracy of the CNN-based model created for emotion classification is 95%. The loss curve gives the details about the training loss and validation loss. The loss and accuracy graph for the model is given in Fig. 11.

The analysis based on precision, accuracy, recall and F1 score is given in Table 2. The assessment of the emotion analysis is based on statistical measures of the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) to determine whether the human emotions can be correctly identified. The aim is to find out the accuracy, precision, recall and F1 score of the face recognition model. The overall accuracy of the emotion classification model is 70.35%, precision is 70.94%, recall is 90.85% and F1 score is 79.61%.

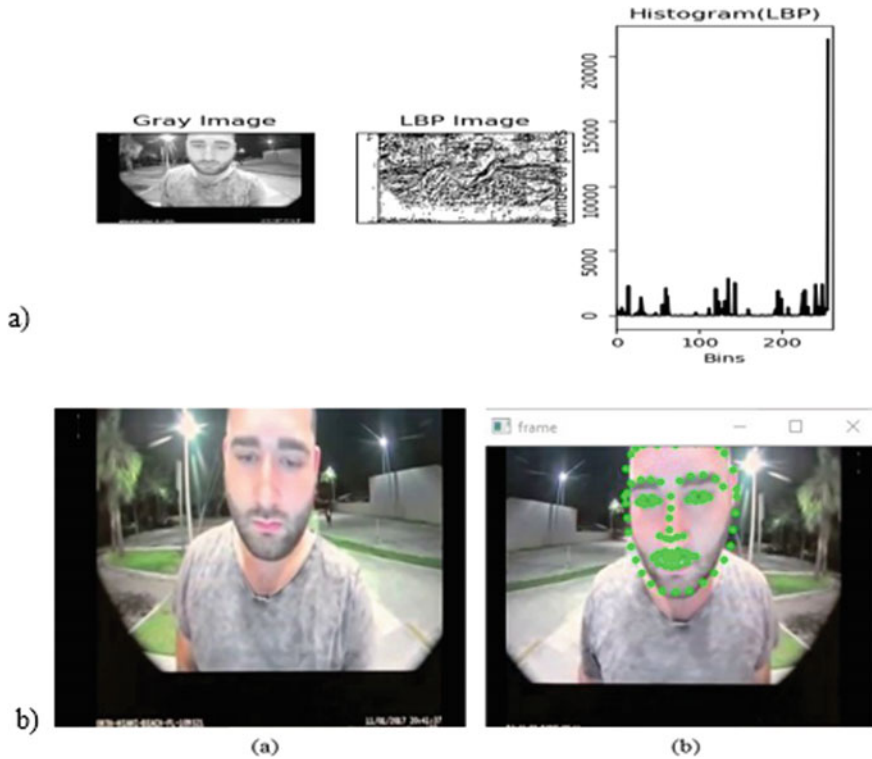


Fig. 8 a Local binary patterns and b facial landmarks

## 6 Future Works and Conclusion

Facial expressions are used for non-verbal communication. The state of mind of a person can be identified from the facial expressions. Emotion recognition or facial expression analysis is one of the emerging areas in artificial intelligence which is used for predicting the emotional behaviour of a human being. This project tried to identify the mental well-being of a human being based on the expressions of the individual. While doing the human emotion recognition, the hand's movements, as well as a change in the posture of the human being, should be considered, so in future, if the human posture is being incorporated along with facial expression, better prediction of emotion can be done. Along with facial expression, audio analysis can also be incorporated to predict the emotions of an individual. In this project, the case of ATM scenarios only is being tried to take, which can be extended to other scenarios also. The emotion predicted can be used in the identification of abnormal behaviour. In this project, the initial step is to process the video frames. After processing the video frames, human faces can be detected with the help of haar cascade classifiers. From the detected faces, the emotional state of the human can be predicted with the help

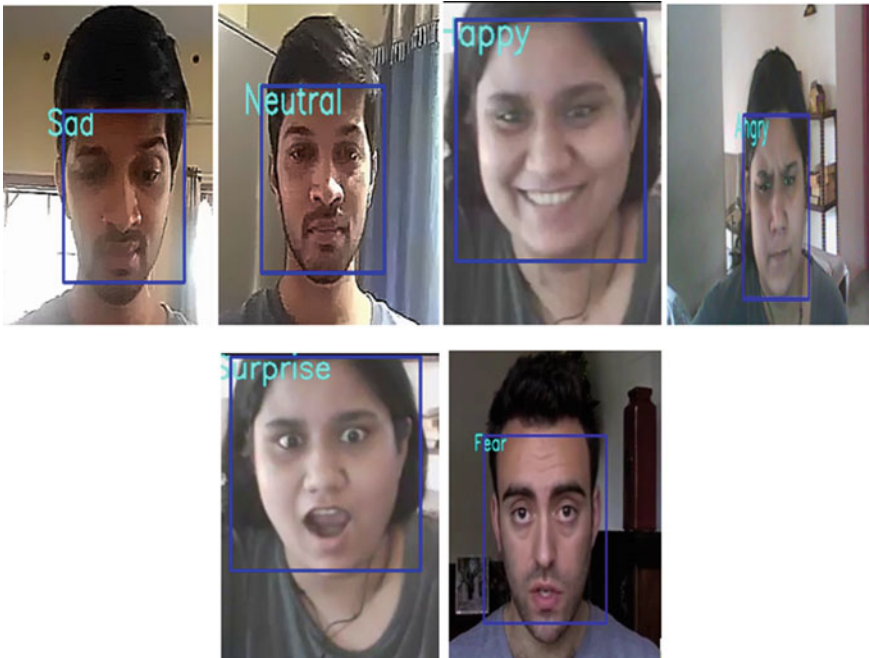


Fig. 9 Emotion classification

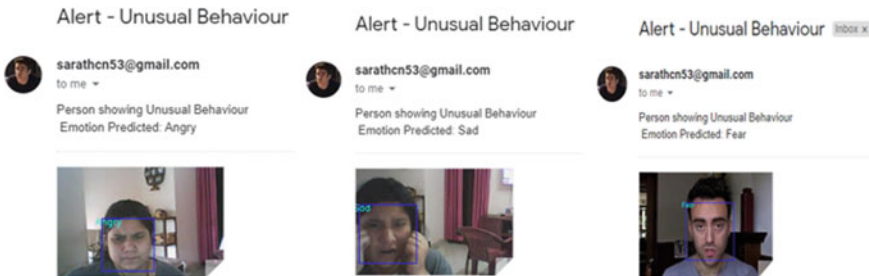
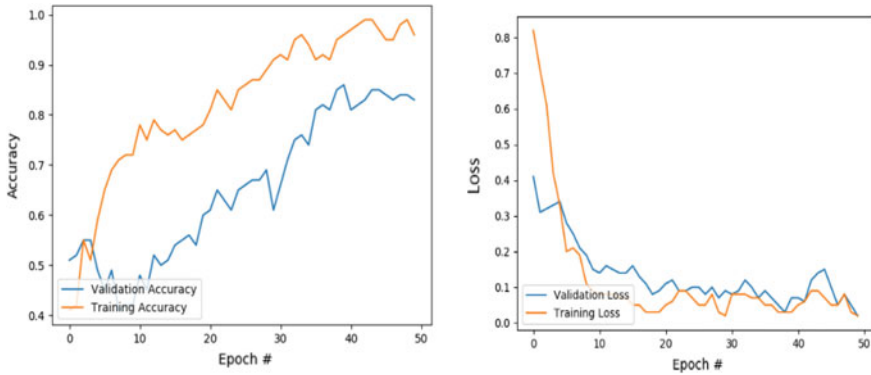


Fig. 10 Alert generation

of a deep learning model. Finally, the emotion is classified as happy, sad, disgust, surprise, fear, anger and neutral. If the emotion is classified as fear, anger or sad, then an alert message is sent to the corresponding authorities via mail. The main drawback of the system is accuracy in classifying the emotion as disgust. In some of the cases, the disgust is not correctly predicted and are misinterpreted as other emotions. Another drawback of the proposed system is that the predicted emotions can vary due to change in the alignment of the facial regions and other lighting variations.



**Fig. 11** Accuracy and loss graphs

**Table 2** Analysis of facial expression model

Video	Frame rate (frames/second)	Precision	Recall	Accuracy	F1-score
Video 1	7	73.68	97.67	72.41	83.99
Video 2	17	71.11	86.48	71	78.04
Video 3	11	68.75	94.28	68	79.51
Video 4	7	70.24	85	70	76.9
Average		70.94	90.85	70.35	79.61

## References

1. Mahmud F, Al Mamun Md (2017) Facial expression recognition system using extreme learning machine. *Int J Sci Eng Res* 8(3)
2. Salih H, Kulkarni L (2017) Study of video based facial expression and emotions recognition methods. In: *International conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC)*
3. Rajesh Kumar GA, Kumar RK, Sanyal G (2017) Facial expression analysis using deep convolutional neural network. In: *International conference on signal processing and communication (ICSPC'17)*, 28–29 July 2017
4. Noroozi F, Marjanovic M, Njegus A, Escalera S (2017) Audio-visual emotion recognition in video clips. *IEEE Trans Affect Comput* 10
5. Jayalekshmi J, Mathew T (2017) Facial expression recognition and emotion classification system for sentiment analysis. In: *International conference on networks and advances in computational technologies (NetACT)*
6. Liu Y, Cao Y, Li Y, Liu M, Song R, Wang Y, Xu Z, Ma X (2016) Facial expression recognition with PCA and LBP features extracting from active facial patches. In: *IEEE international conference on real-time computing and robotics (RCAR)*
7. Ronghe N, Nakashe S, Pawar A, Bobde S (2017) Emotion recognition and reaction prediction in videos. In: *Third international conference on research in computational intelligence and communication networks (ICRCICN)*
8. Seng KP, Ang L-M (2018) Video analytics for customer emotion and satisfaction at contact centers. *IEEE Trans Human Mach Syst* 8

9. Ousmane AM, Djara T, Zoumarou WFJ, Vianou A (2019) Automatic recognition system of emotions expressed through the face using machine learning: application to police interrogation simulation. In: 3rd international conference on bio-engineering for smart technologies (BioSMART)
10. Kim J-H, Kim B-G, Roy PP, Jeong D-M (2019) Efficient facial expression recognition algorithm based on hierarchical deep neural network. *IEEE Access* 7
11. Jung H, Lee S, Yim J, Park S, Kim J (2015) Joint fine-tuning in deep neural networks for facial expression recognition. In: *IEEE international conference on computer vision (ICCV)*
12. Sharma N, Jain C (2019) Characterization of facial expression using deep neural networks. In: 5th international conference on advanced computing and communication systems (ICACCS)
13. Yang B, Cao J, Ni R, Zhang Y (2017) Facial expression recognition using weighted mixture deep neural network based on double-channel facial images. *IEEE Access* 6
14. Kim S, Kim H (2019) Deep explanation model for facial expression recognition through facial action coding unit. In: *IEEE international conference on big data and smart computing (BigComp)*
15. Nair P, Subha V (2018) Facial expression analysis for distress detection. In: *Proceedings of the 2nd international conference on electronics, communication and aerospace technology*

# Secure Digital Image Watermarking by Using SVD and AES



K. M. Sahila and Bibu Thomas

**Abstract** In recent times, digital data is widely used for processing and sometimes the processed data is needed to be transmitted from one organization to other. For such situations, secure transmission of digital data over the network is needed. Digital watermarking provides authority by preventing unauthorized access of multimedia data like audio, video and images. Our research work is based on an SVD-AES-based digital image watermarking scheme to provide secure transmission of digital images over networks. Initially, SVD-based watermarking is carried out. After watermarking, image is encrypted by means of an Advanced Encryption Standard (AES) algorithm. After that, encrypted image is shared over networks and finally AES decryption gives the watermarked image. The proposed method uses AES-128 encryption algorithm. The algorithm shows robust nature towards attacks like content removal, copy and paste attack and also cryptographic attacks like brute-force attack.

**Keywords** LSB watermarking · Arnold scrambling · SVD · AES-128 · Encryption · Decryption

## 1 Introduction

In recent times, due to the great advancement of network technology and computer technology, multimedia data can be transferred through a better way. However, digital data can be easily modified by an attacker. Also, data can be easily duplicated and spread all over the world. So, the security of digital data has become a prime concern. To prevent illegal modifications and duplications, initially, image is encrypted by means of an encryption algorithm [1]. Encryption [2] process involves the use of a secret key that is transmitted along with the data. This same key is used for the

---

K. M. Sahila (✉) · B. Thomas

Department of Electronics and Communication Engineering, College of Engineering Thalassery, Kannur, Kerala, India

e-mail: [sahala.jasmin@gmail.com](mailto:sahala.jasmin@gmail.com)

B. Thomas

e-mail: [bibu.bibuthomas@gmail.com](mailto:bibu.bibuthomas@gmail.com)

decryption purpose at the receiver side. But this method can secure the data until it is in the encrypted form. Once the decryption is carried out, then the data can be easily duplicated. To prove the authority of the owner, some messages are hidden into the image. For that watermarking technology [3] is used. After the watermarking, got the watermarked image that is used for distribution. Through watermarking, then only the owner can extract the watermark with the help of the extraction algorithm. Watermarking can provide authority while encryption provides security.

With the integration of SVD [4], watermarking and AES encryption [5] technique, image authority and security can be assured [6]. This method is widely adopted where both image security and image authority are important [7]. This paper is arranged in five parts. Section 2 gives the methods used for implementation. Section 3 provides the proposed scheme. Performance analysis and experimental results are discussed in Sect. 4. Our research conclusion is given in Sect. 5.

## 2 Implementation Methods

### 2.1 Singular Value Decomposition (SVD)

In matrix algebra, SVD [8] decomposes any  $m \times n$  matrix  $A$  into three new matrices that is

$$A = USV^T \quad (1)$$

where  $U$  and  $V$  are orthonormal matrices, i.e.  $U^T U = I$  and  $V^T V = I$ . Matrix  $S$  is known as a singular matrix which contains singular values that can be arranged diagonally in a decreasing manner. The singular values of  $S$  follow the property,

$$S_1 \geq S_2 \dots S_n \geq 0 \quad (2)$$

In our research, SVD is used for watermarking [9] the digital images. It provides copyright protection and authentication for the multimedia contents [10].

### 2.2 Advanced Encryption Standard (AES)

Advanced Encryption Standard (AES) is one of the most popularly used symmetric key encryption algorithms in cryptography [11]. Comparing to TDES, it is six times faster. Characteristics of AES involves the following,

- Symmetric key



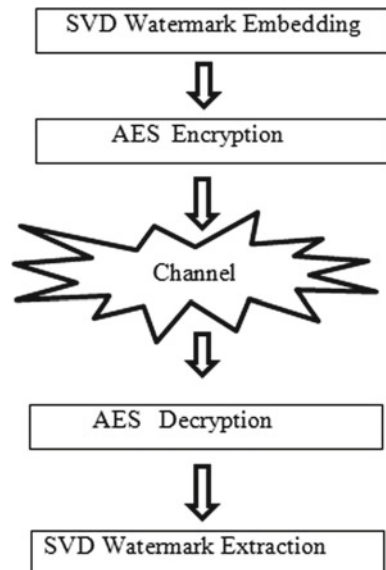
- 128-bit data with variable keys such as 128-bit (10 rounds) or 192-bit (12 rounds) or 256 (14 rounds)-bit keys.
- Complete specifications and design details are provided.
- Stronger and six times faster than TDES.

AES algorithm is an iterative algorithm. It involves a series of linked mathematical operations known as substitutions and permutations networks (SPN). Also, AES performs all its operations on bytes. Hence, 128 bits of the data block are arranged as  $4 \times 4$  matrix (that is 16 bytes). Depending upon the length of the key size, the number of rounds in AES varies like AES-128 with 128-bit key needs 10 rounds, AES-192 with 192-bit key needs 12 rounds and AES-256 with 256-bit key needs 14 rounds. From the input state key, each of the round keys is calculated.

### 3 Proposed Method

To provide a better transmission of the digital images over networks, digital image watermarking is combined by using SVD with AES encryption technique. Initially, the original image is watermarked by using SVD technique. After that, watermarked image is encrypted by using AES algorithm [12]. This cipher image is sent through a communication channel to the receiver. At the receiving end, AES encrypted image is decrypted by using the AES decryption algorithm and finally watermark extraction process by using SVD is carried out. Figure 1 gives the block diagram representation.

**Fig. 1** Block diagram of the proposed method



The main features of our algorithm include the use of Arnold scrambling, which helps to find the tampered block information accurately with the help of self-recovery number. Also, only the greatest singular value has been used from each of the  $4 \times 4$  blocks for the calculation of the block authentication number. It will simplify the calculations. Also, neighbourhood block-based recovery is performed for accurate self-recovery. Another important feature is nothing but the use of the AES-128 algorithm for the encryption purpose. AES-128 shows robust nature towards brute-force attacks.

### ***3.1 Watermark Embedding Process***

In our method, SVD has been directly applied to digital images [13]. The diagonal matrix  $S$  has diagonal elements that are known as singular values, which are kept in descending order. This singular value contains the information of each layer in the original image, which shows robust nature towards different types of attacks. Arnold transformation gives the embedding area, which ensures the security and self-recovery of the input image. Here, LSB watermarking [14] is used. Inserted watermark is made up of two kinds of bits, i.e. block authentication number (BAN) and self-recovery number (SRN). BAN provides image authenticity, and SRN is used to recover the input image if it gets tampered. Initially, the original image is divided into small  $4 \times 4$  image blocks. After that replace LSB of each  $4 \times 4$  blocks with zeros. Then, blockwise singular value decomposition (SVD) on each  $4 \times 4$  blocks is performed. After the decomposition, three new matrices are got from which singular values can be extracted. These singular values are used for the determination of block authentication number (BAN). Then, determine the BAN by simply converting the greatest singular value to 12-bit binary number. Finally, Arnold scrambling is performed over each  $4 \times 4$  blocks to find the self-recovery number (SRN). After Arnold scrambling,  $4 \times 4$  blocks are again decomposed into  $2 \times 2$  blocks. Next step is to find the average values of each  $2 \times 2$  blocks. Then determine the SRN by taking the first 5 MSBs of final average values. Combine this 12-bit BAN with 20-bit SRN to obtain the watermark information. Finally, the watermark information is inserted into the LSBs of the original image by replacing the last two LSBs of each of the  $4 \times 4$  blocks with that of generated watermark information of each  $4 \times 4$  blocks to get the complete watermarked image. Steps involved in our algorithm is given below.

1. Read the original image.
2. Convert it into a greyscale image.
3. Extract  $4 \times 4$  blocks from the greyscale image.
4. Replace the four LSBs of  $4 \times 4$  blocks with zeros.
5. Find singular values of  $4 \times 4$  blocks.
6. Determine the block authentication number by converting the greatest singular value to 12-bit binary number.

7. Perform Arnold scrambling.
8. Compute the average values of each of the  $2 \times 2$  blocks.
9. Determine the self-recovery bits by taking the first 5 MSBs of final average values.
10. Repeat the above steps in order to obtain the complete watermarked image.

### 3.2 AES Encryption and Decryption Process

AES-128 involves series of linked mathematical operations, and each round’s output is given as input to the next round. Each round needs 128-bit input data with 128-bit key, which means in one round it involves 4 words of key, therefore the key size decides the number of rounds. AES [15] simplifies the design by using the same key for both encryption and decryption purpose. Figure 2 shows the flow diagram of AES algorithm. AES transformations are given below.

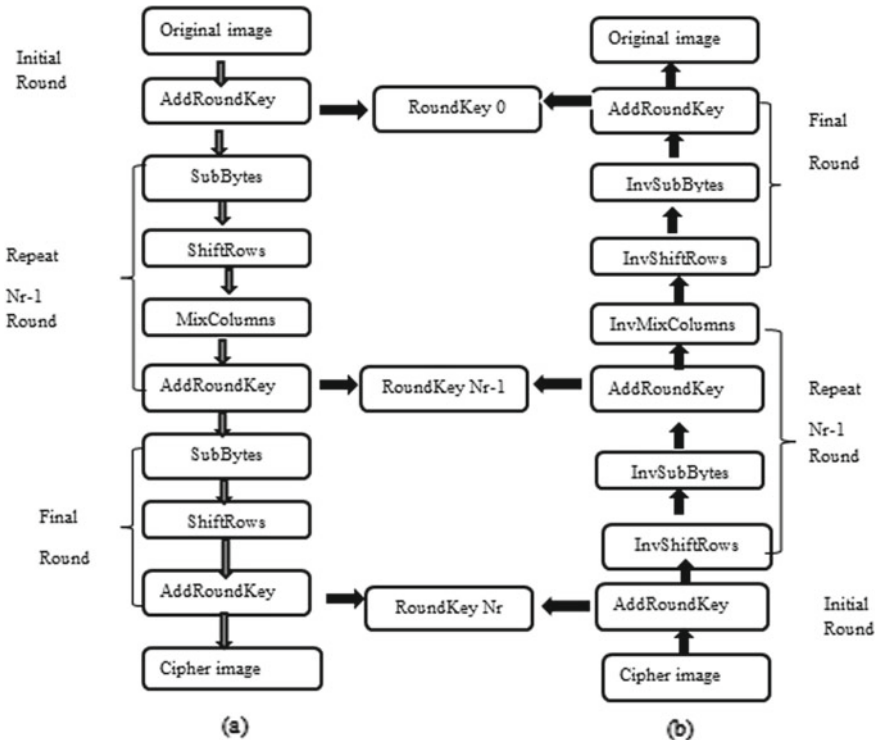


Fig. 2 Flow diagram of AES algorithm (a) Encryption process (b) Decryption process

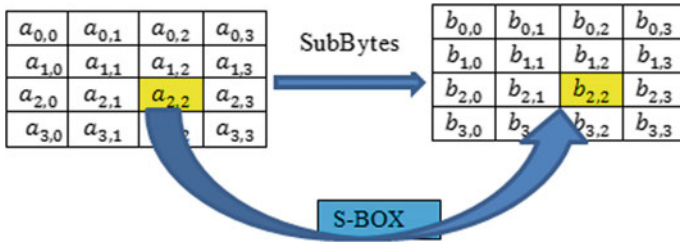


Fig. 3 Byte substitution

**Substitute Bytes (SubBytes)** Here, each of the input state bytes are replaced by another bytes from the substitution table (S-Box). The result is in the form of  $4 \times 4$  matrix. Figure 3 shows the operation.

**ShiftRows** In this round, without changing the first row of the input state, bytes in the second, third and fourth positions are left shifted by 1, 2 and 3 bytes, respectively. Figure 4 shows the result.

**MixColumns** In this transformation, each column of the input state is transformed into a new column by multiplying it with a polynomial  $C(x)$ . This function takes one column of bytes and outputs completely new column of bytes. Mix Column operation is not performed in the final round. Figure 5 shows the result.

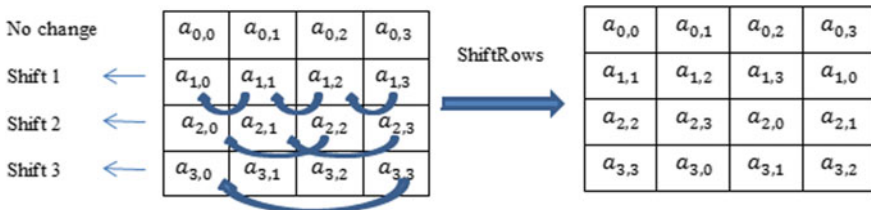


Fig. 4 Shift row

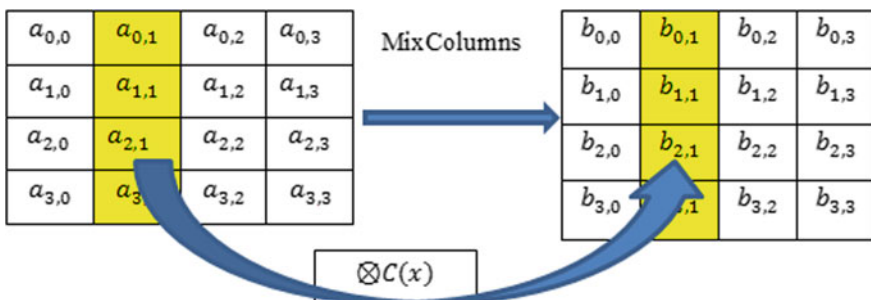


Fig. 5 MixColumn

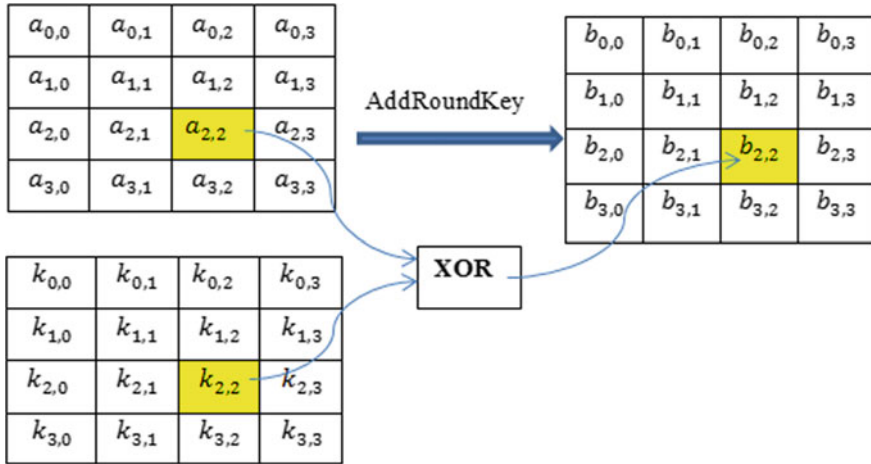


Fig. 6 AddRoundKey

**AddRoundKey** This round is simply an XOR operation of input state bytes with the round key. Key scheduling process derives the round key from the encryption key. Figure 6 shows the result.

Decryption algorithm of AES-128 uses the inverse of transformations of encryption algorithm. Final round does not involve the use of MixColumn operation.

### 3.3 Watermark Extraction Algorithm

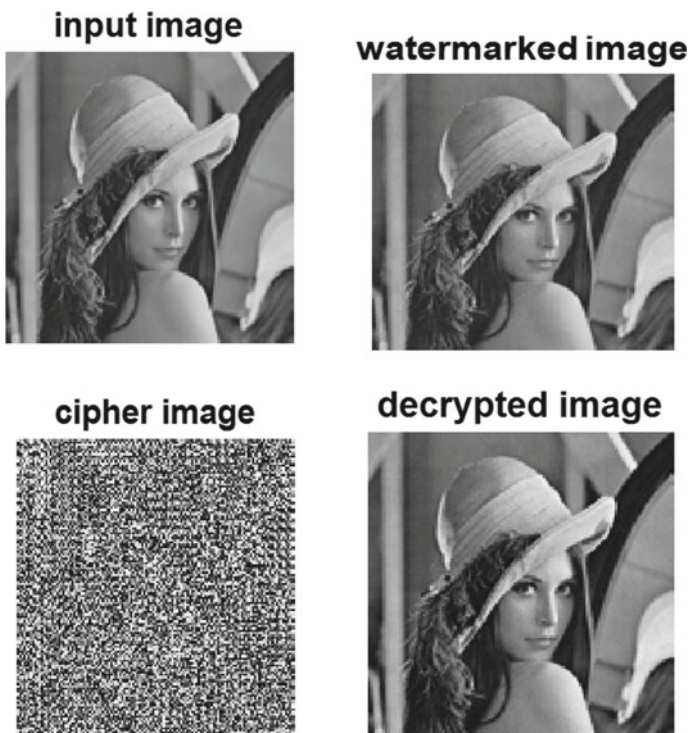
The extraction process is similar to the embedding process in the reverse order. Here, first  $4 \times 4$  blocks are extracted from the watermarked image. Then, BAN and SRN from the  $4 \times 4$  block are extracted. Then, replace the LSBs of each  $4 \times 4$  blocks with zeros. To obtain the singular values, SVD is applied over each  $4 \times 4$  blocks. Then, compute the BAN from the greatest singular value. In the next step, verify whether the LSB extracted BAN and calculated BAN are equal, if not recover the blocks by using self-recovery number. Calculation of SRN is same as in the embedding process. Finally, regenerate the input image through neighbourhood block-based recovery by using SRN. Various steps involved in the watermark extraction process are given below.

1. Read the watermarked image.
2. Extract  $4 \times 4$  blocks from the watermarked image.
3. Extract authentication bits and self-recovery bits from the  $4 \times 4$  block.
4. To obtain the tampered region, authentication bits are calculated from  $4 \times 4$  block.
5. Replace the four LSBs of each  $4 \times 4$  block with zeros.
6. Find singular values using SVD.

7. Compute the block authentication bits from greatest singular value.
8. Verify the recovered BAN and calculated BAN is equal, if not recover the blocks by using self-recovery bits.
9. Determine the self-recovery bits by taking the first 5 MSBs of final average values.
10. Finally, regenerate the 4\*4 block from the 2\*2 blocks and then reconstruct the input image.

## 4 Experimental Results

Performance of the algorithm is evaluated by using different digital images. All the analyses are done using MATLAB version R2016b. Figure 7 shows the experimental results, where (a) shows the input image and (b) shows the watermarked image, respectively, (c) is encrypted watermarked image and (d) gives the decrypted watermarked image. Table 1 shows watermark embedding time, encryption time,



**Fig. 7** Output **a** original image, **b** watermarked image, **c** encrypted watermarked image and **d** decrypted image

**Table 1** Performance analysis of the proposed algorithm

Types of attacks	PSNR value of watermarked image in dB	PSNR value of image after extraction in dB	MSE value of watermarked image in dB	MSE value of image after watermarked extraction in dB
Content removal attack	44.1120	28.7591	2.5228	86.5298
Copy and paste attack	39.1209	38.7815	7.9614	8.6085

decryption time and extraction time. Also, PSNR and MSE values of the input image and watermarked image are calculated.

#### 4.1 Robustness Analysis

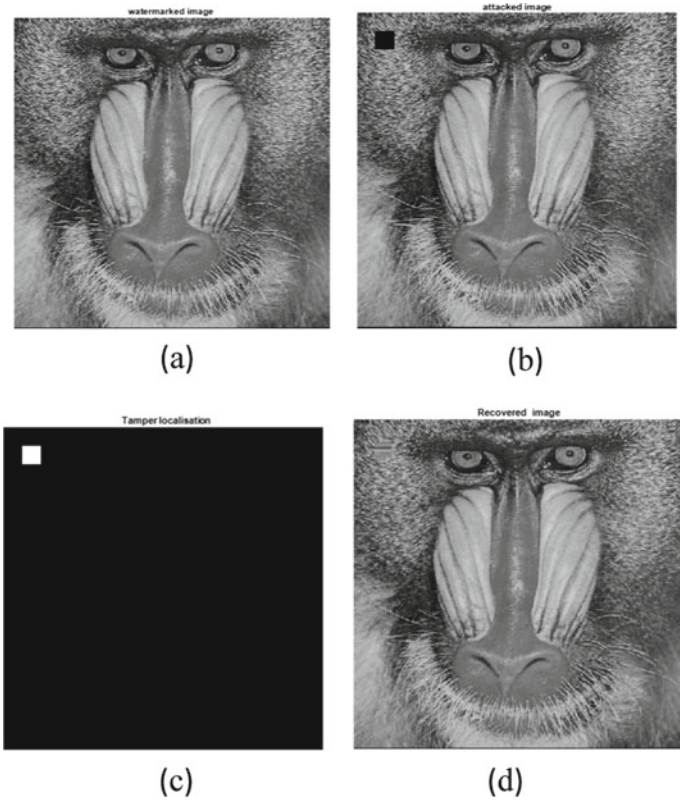
To check the accuracy of our algorithm, two types of image attacks have been used, content removal attack and copy and paste attack. Results of content removal attack are shown in Fig. 8. Some of the contents are removed from the watermarked image after these types of attacks. With the help of watermark extraction algorithm, the tampered region [16] can be located and it also recovers the original image through a neighbourhood block-based recovery method.

**Content Removal Attack** Here, some portion from the image “mandrill” is removed. By using our extraction algorithm, the tampered region can be located. Also, with the help of SRN, original image can be recovered. Results are shown below.

**Copy and Paste Attack** In these types of attacks, some part of the image is copied to the watermarked image. Extraction algorithm provides the exact tampered region and recovers the original image. Results are given in Fig. 9.

#### 4.2 Analysis with PSNR and MSE Values

To check the robustness of the proposed method, PSNR and MSE values are calculated. Table 2 shows PSNR and MSE values against content removal and copy and paste attack. Figure 10 shows variations in the PSNR values of watermarked images before and after extraction against two types of attacks. And Fig. 11 shows variations in the MSE values of watermarked images before and after extraction against the same types of attacks. Analysing these graphs can be seen that there is around 20 dB decrease in the PSNR values after the extraction due to the content removal attack.



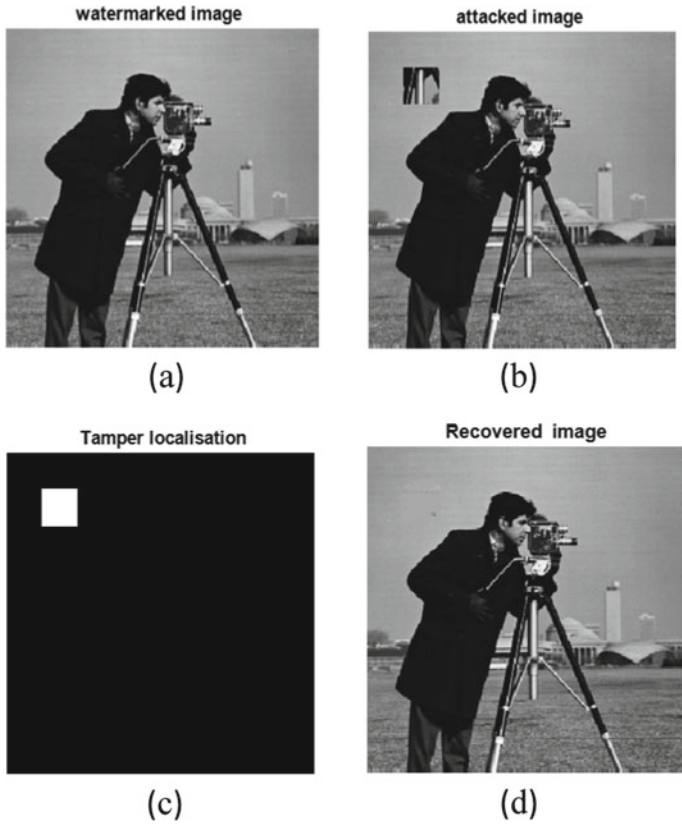
**Fig. 8** **a** Watermarked image, **b** content removal attack, **c** tamper localization and **d** recovered image

Also, MSE values of the watermarked images are increased by some percentage after the attack. But during the copy and paste attack, there are a few dB decreases in the PSNR values. However, the proposed algorithm is acceptably well and provides secure transmission of digital images over networks.

## 5 Conclusion

In this research work, SVD-AES-based digital image watermarking scheme has been proposed which ensure the security of the digital images as well as image authority. Initially, SVD-based watermarking is carried out. After SVD watermarking, image encryption is carried out by using the AES-128 algorithm. This encrypted image





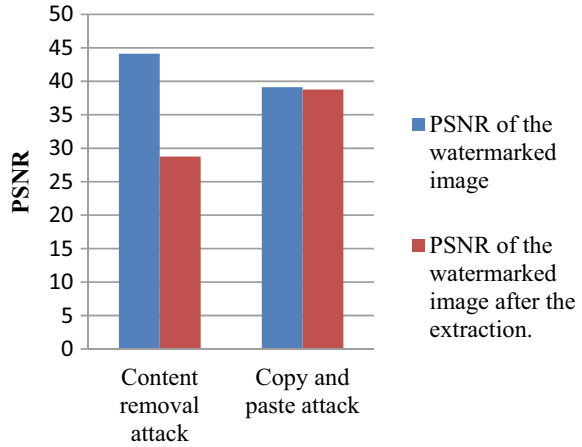
**Fig. 9** a Watermarked image, b copy and paste attack, c tamper localization and d recovered image

is shared over networks and finally, AES decryption is performed to get the watermarked image. The proposed method uses the 128-bit key Advanced Encryption Standard. The algorithm shows robustness towards image attacks like content removal, copy and paste attack and also cryptographic attacks like brute-force attack. In future, the security of the algorithm against different types of image attacks is aimed to improve.

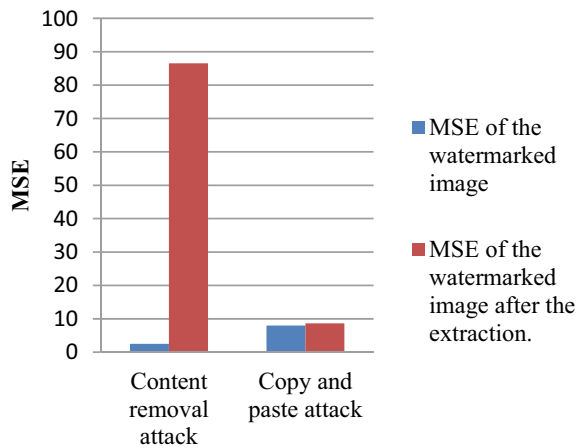
**Table 2** Obtained PSNR, MSE values of watermarked images against various attacks

Images	Watermark embedding time (s)	Encryption time (s)	Decryption time (s)	Watermark extraction time (s)	PSNR of original image and watermarked image	MSE of original image and watermarked image	PSNR of original image and extracted watermarked image	MSE of original image and extracted watermarked image
Cameraman	2.6277	0.2173	0.2923	4.0408	44.1868	2.4797	21.9495	415.0517
Lena	7.2752	0.3889	0.4561	7.7518	44.1120	2.5228	28.7591	86.5298
Peppers	7.1168	0.4489	0.4590	8.5726	44.1684	2.4902	27.3302	120.2427
Coloredchips	6.5052	0.4638	0.4722	6.8500	44.0403	2.5648	26.1896	156.3567
Football	2.9932	0.4585	0.4598	4.1696	43.9938	2.5924	32.7425	34.5807
Rice	3.5056	0.3992	0.3995	4.8415	44.2530	2.4420	23.5801	285.1482
Mandrill	7.3538	0.4220	0.4599	7.5098	44.2639	2.4361	29.6470	70.5308

**Fig. 10** Variations in the PSNR values of watermarked images before and after extraction against two types of attacks



**Fig. 11** Variations in the MSE values of watermarked images before and after extraction against two types of attacks



## References

1. Chang C-C, Hwang M-S, Chen T-S (2001) A new encryption algorithm for image cryptosystems. *J Syst Soft.* [www.elsevier.com/locate/jss](http://www.elsevier.com/locate/jss)
2. Abusukhon A, Talib M (2012) A novel network security algorithm based on private key encryption. In: *IEEE international conference on cyber security, cyber warfare and digital forensic (Cyber Sec)*
3. Larijani HH, Rad GR (2008) A new spatial domain algorithm for gray scale images watermarking. In: *International conference on computer and communication engineering*
4. Liu R, Tan T (2002) An SVD-based watermarking scheme for protecting rightful ownership. *IEEE Trans Multimedia* 4:121–128
5. Metkar SP, Lichade MV (2013) Digital image security improvement by integrating watermarking and encryption technique. In: *IEEE international conference on signal processing, computing and control (ISPCC)*. <https://ieeexplore.ieee.org/xpl/conhome/6636397/proceeding>

6. Bouslimi D, Coatrieux G, Cozic M, Roux C (2012) A joint encryption/watermarking system for verifying the reliability of medical images. *IEEE Trans Inform Technol Biomed* 16(5). 1089-7771/\$31.00 © 2012 IEEE
7. Wong PW, Memon N (2001) Secret and public key image watermarking schemes for image authentication and ownership verification. *IEEE Trans Image Process* 10
8. Shnayderman A, Gusev A, Eskicioglu AM (2006) An SVD-based gray scale image quality measure for local and global assessment. *IEEE Trans Image Process* 15(2)
9. Kumar S et al (2016) A novel spatial domain technique for digital image watermarking using block entropy. In: 2016 fifth international conference on recent trends in information technology, 978-1-4673-9802-2/16/\$31.00© 2016. IEEE
10. Zhang N (2016) Spatial domain digital watermarking of multimedia objects for buyer authentication, 978-1-5090-0654-0/16/\$31.00 ©2016. IEEE
11. Stallings W *Cryptography and network security principles and practices*, 4th Edition
12. Chuman T, Sirichotedumrong W, Kiya H (2015) Encryption-then-compression systems using grayscale-based image encryption for JPEG images. *IEEE Trans Inform Forensics Sec*, <https://doi.org/10.1109/tifs.2018.2881677>
13. Gonzalez RC, Woods RE (2004) *Digital image processing*, 2nd edition, pp 411–514
14. Basu A, Roy SS, Chattopadhyay A (2016) Implementation of a spatial domain salient region based digital image watermarking scheme. In: Second international conference on research in computational intelligence and communication networks (ICRCICN). 978-1-5090-1047-9/16/\$31.00 © 2016 IEEE
15. Zhang Q, Qunding (2015) Digital image encryption based on advanced encryption standard (AES) algorithm. In: Fifth international conference on instrumentation and measurement, computer, communication and control. 978-1-4673-7723-2/15 \$31.00 © 2015 IEEE. <https://doi.org/10.1109/imccc.2015.261>
16. Singh A, Dutta MK (2014) A blind ad fragile watermarking scheme for tamper detection of medical images preserving ROI. In: International conference on medical imaging, m-health and emerging communication systems (MedCom). 978-1-4799-5097-3/14/\$31.00 ©2014 IEEE
17. Kallel M, Kallel IF, Bouhlef MS (2006) Medical image watermarking scheme for preserving the image history. 0-7803-9521-2/06/\$20.00 §2006 IEEE

# BER Analysis of Power Rotational Interleaver on OFDM-IDMA System Over Powerline



Priyanka Agarwal, Ashish Pratap, and M. Shukla

**Abstract** Power line communication (PLC) is a technology of transferring data over power lines. As the power lines were not designed to carry communication signals, noise in the form of impulses and fading corrupt the signals. Researchers have proposed various techniques to make communication over power lines feasible such as interleave division multiple access (IDMA), code division multiple access (CDMA), and orthogonal frequency division multiplexing (OFDM). The most promising results have been obtained by employing a combination OFDM-IDMA scheme which tends to resolve both the issues caused due to impulse noise and fading. But the key factor in the performance is based on the type of interleaver used in the system. In literature, many interleavers have been proposed such as random, prime, power, tree-based, and power rotational out of which the performance of power rotational interleaver has not been implemented in OFDM-IDMA over PLC. The paper aims to bring out the key aspects of power rotational interleaver and analyze the performance in terms of bit error rate by performing simulations in MATLAB and also present a comparative study with the other interleavers.

**Keywords** Power rotational interleaver · OFDM-IDMA · IDMA · Random interleaver · Power line communications (PLC)

## 1 Introduction

Smart grid concept is gaining importance nowadays, as the shortage of electricity is creating havoc in daily life and to fulfill the electricity demands, the grid needs to be

---

P. Agarwal (✉) · A. Pratap · M. Shukla  
Electronics Engineering Department, Harcourt Butler Technical University, Kanpur, India  
e-mail: [priyanka.or.bhartiya@gmail.com](mailto:priyanka.or.bhartiya@gmail.com)

A. Pratap  
e-mail: [ashishpratap152@gmail.com](mailto:ashishpratap152@gmail.com)

M. Shukla  
e-mail: [manojkrshukla@gmail.com](mailto:manojkrshukla@gmail.com)

evolved. The smart term refers to real-time communication with the grid to control the consumption, to monitor the usage, and to rectify the network failures instantaneously [1]. For real-time communication and resolving network issues, communication signals are transmitted over power lines. But as the power lines are designed to carry only electrical signals, lots of issues come up while conveying messages such as noise in the form of impulses and frequency selective fading. Impulsive noise is caused due to switching of appliances over the electrical network which causes transients of very short duration thereby creating burst errors. Frequency selective (FS) fading arises due to branching of the power cable, as when the signal passes through any diversion reflection happens due to the discontinuities in impedance [2]. The multiple reflections when reaching the receiver along with the main signal might cause destructive interference, resulting in the loss of data. But along with some cons of power line communication (PLC), the main benefit of employing this method is the free installation cost as the power line are ubiquitous, and hardware cost is a key component of any system. Therefore, to make communication feasible over power lines, robust modulation techniques must be employed. The modulation scheme must be able to counteract the effects of fading and impulsive noise [3]. One such method widely employed in the wireless medium is orthogonal frequency division multiplexing (OFDM) and interleave division multiple access (IDMA).

OFDM is an upgraded version of frequency division multiplexing (FDM), and it relies on orthogonal carriers for the transmission of data signals. The orthogonality helps in nullifying the effects of FS fading as different bits are transported over the different carrier, and as FS fading impacts signals over a particular frequency, only a few bits are affected and the data can be reproduced with insignificant errors. OFDM has some additional advantages such as less inter-carrier interference, high data rates, and use of complex equalizers are avoided at the receiver which significantly lowers the system cost, but OFDM fails to remove burst errors. To solve the problem of burst errors, interleave division multiple access (IDMA) can be employed. This technique relies on interleaving or scrambling the data, resulting in the random ordering of bits followed by spreading [4]. Thus, even if burst error happens in the signal, on re-ordering of the data at the receiver, burst error gets distributed and leads to almost correct reception of data. Further, the spreading of data nullifies the effect of impulses. Thus, OFDM-IDMA can collectively resolve the issues while communicating over power lines [5]. OFDM-IDMA was implemented in [6, 7] over wireless medium but introduced over power lines in [8].

A major factor in the performance of the IDMA system is the choice of interleaver. An efficient interleaver should be able to generate multiple interleaving patterns for different users in a very short time, the correlation among the various patterns must be low and computational complexity should be less [9]. Many interleavers are proposed in literature such as random, master random, block, matrix, helical, tree, prime, and power rotational [10, 11]. In [12], Shukla has proposed the use of IDMA over power line channel and implemented it using random interleaver. The performance was satisfactory and cheaper as the power lines already exist. The paper [13], proposed a user-specific chip-level interleaver called as power or master random interleaver. It performed almost the same as random interleaver with less

memory and bandwidth requirement. Recently, Yadav [14] analyzed the performance of power interleavers on integrated IDMA, SC-FDMA-IDMA and OFDM-IDMA and proved the enhanced performance of power interleaver as compared to random. In [15], a novel tree-based interleaver was proposed and which proved superior to master random and random interleavers in terms of BER, memory requirement. In [16], Tripathi implemented tree-based interleaver over PLC and the results show that it performs better than random interleaver even on other mediums. Yadav in [17], compared the functioning of conventional IDMA with grouped IDMA, OFDM-IDMA, using tree-based interleaver and plotted the BER graph. The benefits of tree-based interleaver along with the advantages OFDM-IDMA proved to be an excellent choice for communication.

In [18], Sharma proposed power rotational interleaver, an improved version of power interleaver which occupied less memory and had less computation complexity. In [19], Tripathi and Shukla implemented power rotational interleaver in underwater acoustic communication and plotted the results. The performance results were satisfactory. In [20], Agarwal has compared the performance of various interleavers over a power line based on BER, memory required, computation complexity. The three interleavers implemented over power line; i.e., random, power, and tree-based have some shortcomings such as random interleaver has high bandwidth requirement; tree-based has less memory usage but large computation complexity, and power has least bandwidth utilization but high cross-correlation. An interleaver which has less bandwidth requirement, less cross-correlation, and less BER was lacking on the power line channel. As per author knowledge, power rotational interleaver has not been implemented on power lines with OFDM-IDMA. Therefore, in the paper power, rotational interleaver is simulated with OFDM-IDMA over power lines and the result is compared with other interleavers. The paper is organized as follows: Sect. 2 describes the OFDM-IDMA and power line channel model. Next section deals with the mechanism of power rotational interleaver. In Sect. 4, the simulation results of power rotational interleaver are presented along with other interleavers. Lastly, the paper is summarized in Sect. 5.

## 2 System Model

### 2.1 OFDM-IDMA System

Orthogonal frequency division multiplexing (OFDM) employs orthogonal subcarriers for transmission of data over the channel. The orthogonality is a boon as it helps in reducing inter-carrier interference (ICI) and by the addition of cyclic prefix inter symbol interference (ISI) can also be minimized. OFDM also supports by providing collectively high data rates by transmitting symbols over low data rate carriers. Due to many advantages, OFDM is highly preferred for less error transmission. But OFDM is unable to tackle the burst errors. Burst errors are the loss in the actual data in which

consecutive bits are affected; thus, while reproducing at the receiver, the group of bits affected are not able to give the exact transmitted data. To resolve it, IDMA is put to use which relies on interleaving as a mechanism of separation of user data. It scrambles the data and thus burst error impacts different bits, resulting in easy regeneration of original data at the receiver. Therefore, OFDM-IDMA can combat multiple disturbances and is a great choice for modulation technique.

The block diagram of OFDM-IDMA is shown in Fig. 1. At the transmitter, the data generator produces random data for  $N$  users ( $U_1, U_2, \dots, U_N$ ). The data is spread by a basic spreading sequence  $S = (+1, -1, +1, \dots, -1)$  to make it resistant to impulses and is further interleaved. Each user has its own interleaving pattern ( $\pi_1, \pi_2, \dots, \pi_N$ ) in order to distinguish the data. Next data is symbol mapped (SM) according to binary phase shift keying (BPSK), and then, before sending out the signal over medium, IFFT is performed. Inverse fast fourier transform helps in the generation of orthogonal carriers by keeping a separation of  $1/f_c$  between them and converts the frequency domain subcarriers in time domain samples for transmission over channel. Also, IFFT divides the spectrum into multiple segments, one for each carrier and thus lowers the data rate of each carrier which results in less interference. The expression for IFFT is given as:

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{2\pi ink/N} \tag{1}$$

At the receiver, the signal obtained can be written as:

$$Y(k) = h_n(k)x_n(k) + \xi_n(k) \tag{2}$$

where  $h$  is the power line channel transfer function,  $x$  is the transmitted signal, and  $\xi$  is the AWGN noise for  $n$ th user and  $k$ th bit. Then, FFT and symbol demapping are

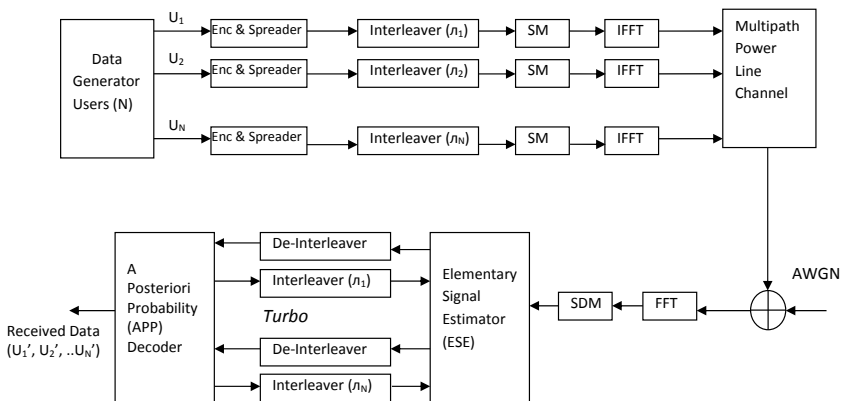


Fig. 1 Block diagram of OFDM-IDMA system



carried out before sending it to turbo processor. The turbo processor forms a major part of IDMA receiver and consists of elementary signal estimator (ESE), blocks of interleavers and de-interleavers, and a posteriori probability (APP) decoders. The ESE works on the principle of minimum mean square error. The ESE starts by predicting the symbol received in the form of log-likelihood ratio which is used to update the values of mean and variance of symbol value. These ESE, bank of interleaver and de-interleaver, APP are executed multiple number of times to decrease the error probability. The output of ESE and APP decoders is formulated as extrinsic log-likelihood ratios (LLRs) and is given as:

$$e(x_n(k)) = \log\left(\frac{p(y/x_n(k) = +1)}{p(y/x_n(k) = -1)}\right), \quad \text{for all } k, n \tag{3}$$

The LLRs from ESE are denoted as  $e_{ESE}$  and from APP decoder as  $e_{DEC}$ . The turbo processor carries out the decoding process by initializing mean and variance and updating them by analyzing the generated LLRs in a series of steps which are mentioned below:

Step 1: Updating the values of mean and variance

Initially,  $e_{DEC}x_n(k) = 0$

Then,  $E(r(k)) = \sum_n h_n E(x_n(k))$

$$\text{Var}(r(k)) = \sum_n |h_n|^2 \text{Var}(x_n(k)) + \sigma^2 \tag{4}$$

$$E(\xi_n(k)) = E(r(k)) - h_n E(x_n(k)) \tag{5}$$

$$\text{Var}(\xi_n(k)) = \text{Var}(r(k)) - |h_n|^2 \text{Var}(x_n(k)) \tag{6}$$

Step 2: Generating LLRs

$$e_{ESE}(x_n(k)) = 2h_n \frac{r(k) - E(r(k)) + h_n E(x_n(k))}{\text{Var}(r(k)) - |h_n|^2 \text{Var}(x_n(k))} \tag{7}$$

Step 3: Updating data

$$e_{DEC}x_n(\pi(k)) = \sum_{k=1}^K e_{ESE} x_n(\pi(k)) \tag{8}$$

The ESE helps in reducing the interference and APP decoder helps in better estimation of bits.

## 2.2 Power Line Channel

The power line can be modeled by using an either top-down or bottom-up approach. The former tries to model the PLC channel based on data collected from frequency responses and employs the best data fitting approach, whereas the latter chooses transmission line fundamentals for deriving the channel model. In this paper, the bottom-up approach has been considered. The transmission line is considered as a pair of two parallel wires which has its resistance ( $R$ ), inductance ( $L$ ), capacitance ( $C$ ), and conductance ( $G$ ). The values of these parameters depend on material and type of wire used and the frequency of the signal to be transmitted [21]. The frequency of the signal is kept high so that it does not interfere with the electrical signal flowing in the wires. After calculating the parameter values, the transfer function of PLC is drawn from the ABCD matrix method as follows:

$$H = \frac{Z_0}{AZ_0 + B + CZ_0Z_s + DZ_s} \quad (9)$$

where  $Z_0$  is the characteristics impedance and  $Z_s$  is the source impedance. As the signal is passed through PLC channel, it gets affected by various types of noise also. In the paper, only AWGN noise is considered.

## 3 Power Rotational Interleaver

At the receiver to decode the data, the interleaving pattern is required. In random interleaver each user has an independent interleaving pattern, and thus, at the receiver, each permuter indices has to be transmitted which occupies loads of bandwidth. In order to save the bandwidth power rotational interleaver is quite helpful. The algorithm of the same is given as follows:

### 3.1 Algorithm

Let  $N$  users are having  $K$  bits of data after spreading process. These are rearranged in the form of row and column having indices as  $r$  and  $c$ , such that ( $r \times c = K$ ). Then, scrambling is performed in two steps: column shuffling and row shuffling.

Step 1: Numbers from 1 to  $K$  are filled in a row fashion in the  $r \times c$  matrix. A random sequence of length ' $r$ ' is generated abruptly by the user. These numbers are divided by the number of columns ( $c$ ), and the remainder is stored. Next, rotation of numbers in the column is done according to the remainder values.

Step 2: The same process as in column permutation is carried out, except the division step. The random number sequence from 1 to  $r$  is stored in the first row.

For the subsequent row permuter indices, the first-row element is shifted right by the value of row index. Next, the numbers in a row are rearranged according to the face values in the same row.

For the next user, the concept of master random interleaver is used. The steps are further explained by an example. The user data bits  $K = 18$ ,  $r = 6$ ,  $c = 3$ . The original matrix is given as:

	<i>r</i>					
<i>c</i>	1	2	3	4	5	6
	7	8	9	10	11	12
	13	14	15	16	17	18

For column shuffling the random sequence generated is:

3	2	5	1	6	4
---	---	---	---	---	---

The numbers are divided by 3 to get the remainder

0	2	2	1	0	1
---	---	---	---	---	---

According to the remainder values, the column numbers are rotated

	<i>r</i>					
<i>c</i>	1	8	9	16	5	18
	7	14	15	4	11	6
	13	2	3	10	17	12

Now, for the row shuffling, again a random sequence is generated:

4	2	6	3	5	1
---	---	---	---	---	---

For subsequent rows, the series in the first row is shifted right according to row indices:

	<i>r</i>					
<i>c</i>	4	2	6	3	5	1
	1	4	2	6	3	5
	3	5	1	4	2	6

Finally, the permuter matrix after row shuffling is:

	<i>r</i>					
<i>c</i>	16	8	18	9	5	1
	7	4	14	6	15	11
	3	17	13	10	2	12

The matrix is reconverted to row form, and interleaving pattern is achieved. Thus, to perform power rotational interleaving only two sequences: column random sequence {3, 2, 5, 1, 6, 4} and row sequence {4, 2, 6, 3, 5, 1} are to be transmitted for a single user instead of one *K* bits sequence. For subsequent users, the given indices are used and master random interleaving concept is applied. Thus, power rotational user occupies lesser memory and bandwidth as compared to random and power interleaver.

## 4 Simulation Results

### 4.1 Comparison Based on Bit Error Rate

The performance of power rotational interleaver was judged against random, tree-based, and power interleaver based on bit error rate (BER) on OFDM-IDMA platform over PLC on MATLAB. The stated interleavers have been implemented over PLC in literature hence are taken as reference for the evaluation of power rotational interleaver. For simulation, number of users  $n = 8$ , blocks = 10, data length = 512, spread length = 16, and the iterations of the turbo processor were taken as 10. The frequency of the signal was chosen as 10 MHz, which is very high to avoid interference with the electrical signal. Some additional parameters of PLC are given in Table 1, and the simulation result is depicted in Fig. 2.

From the graph, it is evident that the power rotational interleaver performs somewhat better than power, random, and tree-based interleavers. The BER decreases more rapidly for power rotational interleavers due to column and row shuffling which results in more randomization and hence less effect of burst error.

**Table 1** Parameters of PLC used for simulation

S. No	Parameter	Value
1.	Frequency	10 MHz
2.	Relative permittivity	3.6
3.	Relative permeability	1
4.	Conductivity	5.8e7
5.	Radius of power line wire	1.12e-3

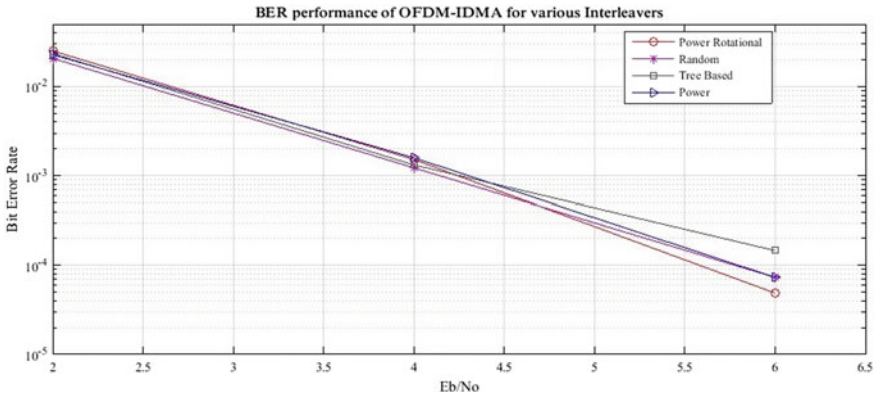


Fig. 2 Simulation results of various interleavers on OFDM-IDMA system over PLC

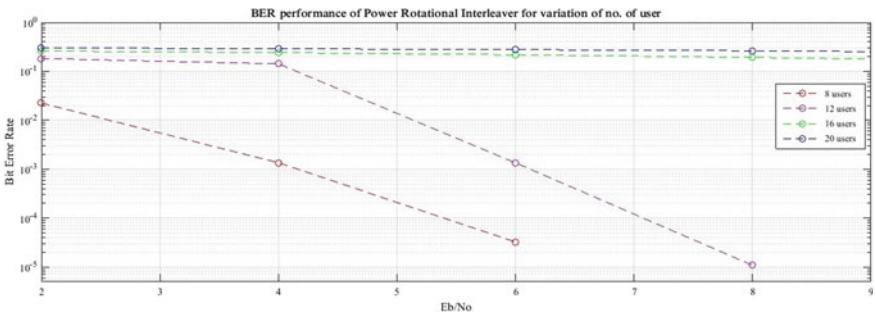


Fig. 3 Simulation results of power rotational interleaver for variation in number of users

Another analysis was done by varying the number of users for power rotational interleaver, and the result is plotted in Fig. 3. From the graph, it can be observed as the number of users is increased the BER increases which are understandable, as it results in multiple simultaneous transmission over the channel which raises the interference for each signal and thus leads to errors (Fig. 4).

Lastly, the performance of power rotational interleaver was compared with random interleaver on OFDM-IDMA and IDMA systems. For both, the power of the technique rotational performs better than random and the results are better for OFDM-IDMA as this method has less influence of interference as compared to IDMA. Thus, overall power rotational interleaver performs substantially better than other interleavers.

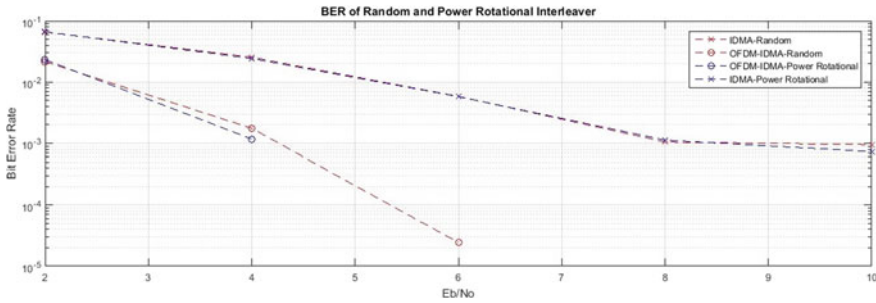


Fig. 4 BER comparison of power rotational and random interleaver on IDMA and OFDM-IDMA

### 4.2 Comparison Based on Other Parameters

In the communication system other than BER, various other parameters such as bandwidth utilization, memory requirement, computation complexity, cross-correlation also play a significant role. The four interleavers random, power, tree-based, and power rotational are compared on the grounds of stated factors. Random interleaver has the highest bandwidth and memory requirement as each user has separate permuter indices and thus has the least cross-correlation among various users. In tree-based interleaver, two master random interleavers are generated, which are later employed to create permuter indices for various clients and thus has higher computation complexity and cross-correlation than random but requires less bandwidth, as only two masters random interleaver is required at receiver for regeneration. Power interleaver needs one master random interleaver for the generation of further permuter indices, and thus, spectrum utilization is less but cross-correlation is very high as all the other permuter indices are powers of master random interleaver. Lastly, power rotational interleaver has same features as power interleaver but has improved cross-correlation as the permuter indices are further rotated both row and column-wise to further randomize the permuter indices. The same has been represented in Table 2.

Table 2 Comparison of interleavers based on various parameters

Parameters	Interleavers			
	Random	Power	Tree based	Power rotational
Bandwidth utilization	High	Least	Low	Least
Memory requirement	High	Least	Low	Least
Computation complexity	Low	High	Medium	High
User-specific cross-correlation	Low	High	Medium	Medium

## 5 Conclusion

The smart grid necessitates communication over power line for real-time services such as fault repairing and tracking of consumption of electricity. To make transmission feasible, modulation technique OFDM-IDMA is quite useful as OFDM combats interferences and IDMA is successful in suppressing burst error. IDMA relies on interleavers for its superior performance. In this paper, power rotational interleaver is implemented over PLC and is compared with random, power, and tree-based interleavers on the MATLAB platform and the results indicate the superior performance of power rotational interleaver. Next simulation points out that power rotational interleaver performs better than random interleaver over IDMA system also. In less memory and bandwidth applications, power rotational interleaver is a viable option. But, the computation complexity of power interleaver is high so applications requiring very high speed cannot be implemented by it. In future, the performance of power rotational can be studied over different modulation and mapping techniques. Also, a new interleaver with lesser computation complexity can be designed.

## References

1. Farhangi H (2009) The path of the smart grid. *IEEE Power Energ Mag* 8(1):18–28
2. Hauser CH, Bakken DE, Bose A (2005) A failure to communicate: next generation communication requirements, technologies, and architecture for the electric power grid. *IEEE Power Energ Mag* 3(2)
3. Bannister S, Beckett P (2009) Enhancing powerline communications in the “Smart Grid” using OFDMA. In: 2009 Australasian universities power engineering conference. IEEE, pp 1–5
4. Ping L, Liu L, Wu K, Leung WK (2006) Interleave division multiple-access. *IEEE Trans Wirel Commun* 5(4):938–947
5. Zhou S, Li Y, Zhao M, Xibin X, Wang J, Yao Y (2005) Novel techniques to improve downlink multiple access capacity for beyond 3G. *IEEE Commun Mag* 43(1):61–69
6. Ping L, Guo Q, Tong J (2007) The OFDM-IDMA approach to wireless communication systems. *IEEE Wirel Commun* 14(3):18–24
7. Dang J, Zhang W, Yang L, Zhang Z (2013) OFDM-IDMA with user grouping. *IEEE Trans Commun* 61(5):1947–1955
8. Chen X, Qu F, Yang L (2011) OFDM-IDMA for power line communications. In: 2011 IEEE international symposium on power line communications and its applications. IEEE, pp 24–28
9. Shukla MK, Srivastava VK, Tiwari S (2011) Implementation of interleavers for iterative IDMA receivers. *Res J Inform Technol*
10. Marne HD, Mukherji P (2015) Comparative study of multiuser detection techniques in OFDM-IDMA systems. In: 2015 international conference on pervasive computing (ICPC). IEEE, pp 1–4
11. Sharma J, Mishra S, Sharma N (2017) An analysis of interleaver types for better BER performance in linear turbo equalizer. In: 2017 international conference on innovations in information, embedded and communication systems (ICIIECS). IEEE, pp 1–5
12. Shukla M, Sharma N, Singh A (2012) Random interleaver with iterative IDMA mechanism for power line communication systems. In: 2012 IEEE international conference on signal processing, computing and control. IEEE, pp 1–6
13. Wu H, Ping L, Perotti A (2006) User-specific chip-level interleaver design for IDMA systems. *Electron Lett* 42(4):233–234

14. Yadav M, Shokeen V, Singhal PK (2017) Uncoded integrated interleave division multiple access systems in presence of power interleavers. *Radioelectron Commun Syst* 60(11):503–511
15. Shukla M, Srivastava VK, Tiwari S (2008) Analysis and design of tree based interleaver for multiuser receivers in IDMA scheme. In: 2008 16th IEEE international conference on networks. IEEE, pp 1–4
16. Tripathi S, Dwivedi JK, Shukla M (2013) Power line communication with tree based interleaver in iterative IDMA systems. In: 2013 5th international conference and computational intelligence and communication networks. IEEE, pp 286–290
17. Yadav M, Shokeen V, Singhal PK (2016) BER versus BSNR analysis of conventional IDMA and OFDM-IDMA based systems with tree interleaving. In: 2016 2nd international conference on advances in computing, communication, and automation (ICACCA) (Fall). IEEE, pp 1–6
18. Sharma S, Ahirwar P, Chauhan S, Upadhyay S, Shukla M (2013) Power rotational interleaver on an IDMA system. *J Innov Syst Des Eng* 4:33–38
19. Tripathi P, Shukla MK (2015) Performance analysis of power rotational interleaver in acoustic environment. In: 2015 international conference on green computing and internet of things (ICGCIoT). IEEE, pp 1266–1271
20. Agarwal P, Shukla M (2020) Effect of various interleavers on uncoded and coded OFDM-IDMA over PLC. In: 2020 5th international conference on communication and electronics systems (ICCES). IEEE
21. Meng H, Chen S, Guan YL, Law CL, So PL, Gunawan E, Lie TT (2002) A transmission line model for high-frequency power line communication channel. In: Proceedings. international conference on power system technology, vol 2. IEEE, pp 1290–1295



# Offline 3D Indoor Navigation Using RSSI



S. Vivek Sidhaarthan, Anand Mukul, P. Ragul, R. Gokul Krishna,  
and D. Bharathi

**Abstract** Indoor positioning and navigation systems are used to track entities in indoor spaces using Global Positioning System (GPS) and other satellite technology. The objective of this research is to demonstrate a semi-dynamic offline 3D indoor navigation setup using Received Signal Strength Indicator (RSSI) values of WiFi routers in the building with the support of an Android application. The data collection module in the application collects RSSI data along with the user's co-ordinates (Online Phase) and a TensorFlow model is trained on the collected data, converted to TensorFlow Lite format and hosted online. In the Offline Phase, RSSI data is recorded, processed with the machine learning model in the device and passed to a module in the app developed with Unity that visualizes the user's co-ordinates in a three-dimensional model of the building. This paper is a bare-bones implementation of the above mentioned ideologies of an indoor navigation system and can be fine-tuned to use in specific applications.

---

S. Vivek Sidhaarthan and Mukul Anand contributed equally to the work.

---

S. Vivek Sidhaarthan · A. Mukul (✉) · P. Ragul · R. Gokul Krishna · D. Bharathi  
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita  
Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India  
e-mail: [mukulanand254@gmail.com](mailto:mukulanand254@gmail.com)

S. Vivek Sidhaarthan  
e-mail: [supersaiyan8vivek@gmail.com](mailto:supersaiyan8vivek@gmail.com)

P. Ragul  
e-mail: [pragul1999@gmail.com](mailto:pragul1999@gmail.com)

R. Gokul Krishna  
e-mail: [gokulkrishna338@gmail.com](mailto:gokulkrishna338@gmail.com)

D. Bharathi  
e-mail: [d\\_bharathi@cb.amrita.edu](mailto:d_bharathi@cb.amrita.edu)

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021  
J. Hemanth et al. (eds.), *Intelligent Data Communication Technologies and Internet  
of Things*, Lecture Notes on Data Engineering and Communications Technologies 57,  
[https://doi.org/10.1007/978-981-15-9509-7\\_67](https://doi.org/10.1007/978-981-15-9509-7_67)

# 1 Introduction

This section gives a brief introduction to indoor navigation and some of the associated challenges. A summary of the work done so far in this topic is explained in Sect. 2. Section 3 describes the proposed solution, steps involved and the system architecture of the solution. The implementation of the system and results are discussed in Section 4. Section 5 concludes the proposed solution and suggests use-case specific applications that can be built having the system proposed in this paper as the base.

## 1.1 Background

Maps application aids in travelling from Place A to Place B but fails when the user needs to go from Point A to Point B inside a complex building. In such a scenario, GPS may not provide an accurate representation as the signals get scattered and attenuated by the building's walls, roofs, furniture etc. [1].

Indoor Positioning and Navigation has been an interesting problem from the early 2000s till date [2]. There is no absolute solution that solves all challenges posed by Indoor Localization and Navigation systems. The development of new technology often changes the perspective that is usually associated with this problem such as the widespread usage of smartphones in navigation.

## 1.2 Challenges

- **Scalability**—The proposed solution has to be repeated for each building in which the navigation system is to be implemented.
- **Cost**—In some solutions, state-of-the-art equipment is required and that involves high cost of initialization and/or operation. It also depends on the availability of the equipment in that area. Such a solution might not be viable for everyone.
- **Signal Noise**—An indoor environment usually involves a lot of human movement and objects. This results in distortion of the radio signals (noise) such as WiFi and Bluetooth. Moreover, the layout of the objects in the building could also be changed, altering the propagation of the signals and leading to inaccurate results.
- **3D Navigation**—3D navigation is also a challenging aspect of indoor navigation. It isn't as straightforward as 2-dimensional navigation since the height-factor proves to be a challenge to compute. It also involves redefining the User Experience (UX) so that the solution is easy to interpret.

## 2 Literature Survey

Indoor navigation can be carried out through various approaches such as with Sound (audible and ultrasound), Light (visible and infrared), Radio Frequency (RF) signals (WiFi, Bluetooth), Magnetic Fields etc. [3]. Bluetooth-based and WiFi-based solutions are the most commonly used due to low cost, ease of availability and compatibility with smartphones. Hence, WiFi-based solution is chosen for our research.

RF solutions are based on Angle of Arrival (AoA), Time of Arrival (ToA), Received Signal Strength (RSS) or combinations or variations of the above approaches [4]. AoA method makes use of the angle at which the signal is recorded by the receiver. ToA method utilizes the time taken by the signal to travel from source to destination. RSS-based solutions involve recording the intensity of the signal that is received.

Based on the above mentioned aspects of signals, positioning of the user is done. The main algorithms that help us position the user are: Triangulation, Trilateration and RSS fingerprint-based [5]. Triangulation uses the properties of triangles to estimate the position of the user by calculating the angles relative to two known reference points. Trilateration is similar to Triangulation but instead uses three known reference points. Fingerprint-based solutions involve the storage of Received Signal Strength Indicator (RSSI) and position information in a database and later retrieving the closest matching record to denote the position of the user.

Lately, the simplicity of RSSI-based solutions have been combined with Machine Learning (ML) to provide good results [6]. k Nearest Neighbours (kNN) is a popular algorithm that has been applied to RSSI fingerprint records, mainly owing to its ease of understanding and implementation [7]. Clustering techniques have also been used on the gathered RSSI data to predict the location of the user [8]. Principal Component Analysis (PCA) is used for pre-processing the data and Ensemble Extreme Learning Machine (ELM) can be used to locate the user [9]. Breakthroughs are also being made with the usage of Deep Learning and Neural Networks for RSSI-based localization [10].

The signal data obtained in an indoor environment is often noisy due to the presence of obstacles that scatter these signals. To overcome this, filtering techniques are used. Particle filter and Kalman filter have shown promising results with regard to RSSI data [11]. Few other filtering methods have also been tested, with varying success rates [12]. Another interesting observation is that usage of Channel State Information (CSI) rather than RSSI could yield better results [13].

The functionality of a smartphone and the ability to record RSSI makes it a great tool for indoor localization [14]. Also with the help of gyroscopic sensor, accelerometer and other sensors present in smartphones, indoor navigation can be enabled [15]. Different aspects of indoor positioning can be fused to provide excellent results. One such novel idea is the combination of images of lights with particle filters acting on sensor data [16].

Another aspect of indoor navigation using smartphones is the evolution of 2D navigation into 3D navigation. One approach involves changing the circles associated

with Trilateration to spheres, with WiFi routers acting as the reference points [17]. Visual data, sensor data and dead reckoning could also aid in creating a 3D indoor navigation system for smartphones [18]. Of late, many phones with inbuilt barometric sensors have emerged. This could be key in identifying the height that the user is located in, in addition to RSSI data [19].

An intriguing proposition for 3D navigation is the utilization of Unity, a game engine, to generate a 3D layout of the building in which indoor navigation is to be implemented [20].

Based on the literature studied, a solution based on RSSI fingerprints and ML is propounded, owing to ease of implementation. Ease-of-implementation is picked as the primary criterion since the focus of this paper is on the overall framework of the solution and not on mere accuracy metrics. The default Android libraries help leverage RSSI information of WiFi signals and open-source ML libraries help in calculating the position of the user based on RSSI. The above mentioned concepts are then combined with a Unity-based building model to complete the indoor navigation system.

### **3 Methodology**

#### ***3.1 Proposed Solution***

The proposed system aims to solve the issue of navigation inside a building by creating a semi-dynamic offline mobile application that helps the user navigate from one point in a building to another point through a 3D representation of the building on the phone screen, along with the path to be taken. The offline positioning of the user is implemented by feeding the values of RSSI of the WiFi routers in the building to an ML model in the application and plotting the resulting co-ordinates in the 3D building model.

Our solution is somewhat scalable since we modularize the application into reusable and modifiable components, involves zero additional cost, takes Height into consideration in terms of floor number and provides a user-friendly visual output.

#### ***3.2 Specific Objectives***

For the building taken into consideration for trial (Academic Block 3 in Amrita University, Coimbatore, India):

1. To record the RSSI of the routers in the building from various points, along with their corresponding Latitude, Longitude and Floor Numbers (Online Phase).
2. To develop a 3D model of the building and map the co-ordinates associated with various locations in the building.

3. To implement a Machine Learning model for Offline 3D Localisation using RSSI.
4. To find the best route from the current point to the target point and visualize the same in the 3D model.
5. To integrate the above mentioned features into a single mobile application and test for accuracy, ease of use and other metrics.

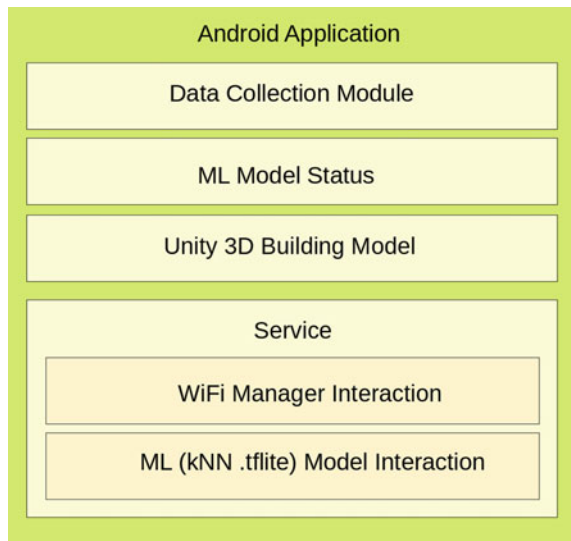
### 3.3 System Architecture and Specifications

The proposed system mainly consists of an Android application with the following architecture:

- A data collection module which helps collect RSSI data using the Android device’s WifiManager and BroadcastReceiver libraries.
- A Machine Learning model page which denotes the availability of the TensorFlow Lite model trained on the collected data and provides an option to download the same for offline usage.
- A 3D building model developed with Unity and integrated into the Android application providing live visual output to the user.
- A custom-built Service that helps obtain RSSI data and converts it to co-ordinates by passing it to the TensorFlow Lite model downloaded and stored locally (Fig. 1).

In the initial Online Phase, RSSI data is collected along with the device’s co-ordinates using WifiManager and BroadcastReceiver libraries and stored in a Comma-Separated Values (.csv) file. This is used to train a Machine Learning

Fig. 1 Android application architecture



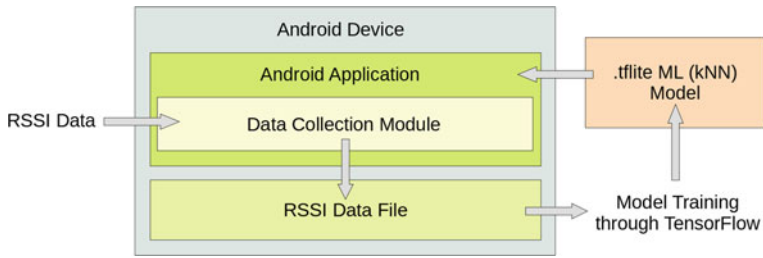
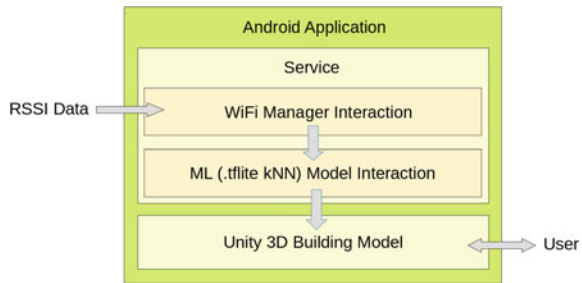


Fig. 2 Online phase

Fig. 3 Offline phase



(TensorFlow) model. The model is then converted to TensorFlow Lite (.tflite) format and saved to the device (Fig. 2).

An Android Service is a component without a User Interface (UI) performing long tasks in the background. In the Offline Phase, a custom-built Android Service helps detect RSSI values, passes them to the trained .tflite model in the device and directs the predicted co-ordinates to the Unity-based 3D Building module that has been integrated into the application (Fig. 3).

The proposed solution requires WiFi routers in the target building, a computer for building the ML model and an Android (version ≥6.0) phone with WiFi functionality. During runtime of the application, the following permissions are requested:

- WiFi access
- High accuracy location access
- Storage read and write.

## 4 Implementation, Results and Discussion

### 4.1 Application Overview

The app consists of 4 primary components—3 different Activities (UI pages) and a Service that enables realtime indoor positioning using RSSI. The home screen has

been designed with the step-by-step procedure for indoor navigation in mind and consists of the sections:

- Data Collection
- Machine Learning (ML) Model
- 3D Building.

First, data is collected through the Data Collection screen, then a model is trained based on the data and hosted online. The second screen enables the download of the model for offline usage. The third screen is the one that the end-user interacts with for offline indoor navigation (Fig. 4).

## ***4.2 Data Collection***

First, in the ‘Access Points’ section of the page, the name of the WiFi network is entered. On clicking ‘Search’, MAC addresses (BSSIDs) corresponding to that name will be obtained. This is to identify all the Access Points (AP) in the building. Once all the APs in the given name are identified, it is saved in a CSV file. The CSV file contains the columns: Latitude, Longitude, Floor and BSSIDs of every router. Once it is completed, collection frequency and floor number are entered in the ‘Collect RSSI Data’ section of the application. RSSI values are appended to the CSV file every ‘x’ seconds until ‘Stop’ is pressed.

The WifiManager API (Application Program Interface) of the Android library is used to implement our system. Apart from this is a BroadcastReceiver that receives updates on the WiFi scan request. BroadcastReceiver is the Android class that listens for updates regarding specific requests. The BroadcastReceiver implementation registers to listen for updates regarding the device’s WiFi state such as change in available APs and change in their signal strengths, and triggers the functions that need to act when the associated buttons are pressed.

## ***4.3 Data Processing and Machine Learning***

The RSSI data obtained in the previous section is then analysed. The data was collected over 17 routers placed in the building. The dataset has 3 target variables—Latitude, Longitude and Floor Number. Each row corresponds to a point in the building where the RSSI values from the routers were recorded (visualized in Fig. 5).

The input values (features) of our ML model are the RSSI values obtained from the 17 different routers in the building. In case there is no RSSI reading for certain routers at specific locations,  $-100$  is inputted in their place. The target variables predicted by the model are Latitude, Longitude and Floor Number (Fig. 6).



Fig. 4 Screenshots from the app in order of usage

Sample input:  $[-100.0, -87.0, -100.0, -80.0, \dots]$

Sample output:  $[10.9061475, 76.8974424, 1.0]$

Using Python’s Scikit-Learn [21] library, ML models like Linear Regression, Decision Tree Regressor, kNN Regressor and Random Forest Regressor are tested. These algorithms have been chosen due to ease of understanding and implementation and also due to their proven success in similar works published in the past. Since the takeaway of the paper is the framework for indoor navigation and not the accuracy



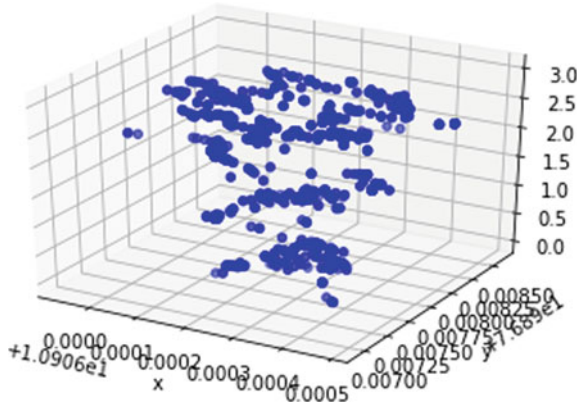


Fig. 5 Points in the building where data was collected

	Latitude	Longitude	Floor	44:31:92:af:c5:50	44:31:92:b0:10:90	44:31:92:9a:44:d0	44:31:92:af:c3:f0
1020	10.906195	76.898072	0	-79.0	-81.0	-64.0	-60.0
1021	10.906199	76.898096	0	-100.0	-81.0	-76.0	-79.0
1023	10.906199	76.898096	0	-100.0	-81.0	-80.0	-79.0
1024	10.906223	76.898126	0	-100.0	-81.0	-80.0	-79.0
1026	10.906223	76.898126	0	-100.0	-63.0	-80.0	-79.0

Fig. 6 Few columns of the collected dataset

of the solution, accuracy (mean error) isn't our primary criterion for choosing the 4 mentioned algorithms among the lot.

Mean, Standard Deviation and Five-Point Summary of the Error (in metres) are the metrics that have been chosen to evaluate the performance of our models. From the observations (refer Fig. 7), it can be concluded that kNN is the best-suited for our application. It outperforms the other algorithms in all evaluation metrics taken into consideration. Boxplots showing the Five-Point Summary and Lineplots showing how the different models predicted each record have also been utilized for a clear understanding of the performance of different algorithms on the dataset.

kNN algorithm is probably the most widely used ML algorithm for localization due to its similarity to the working of a positioning system that predicts based on known (past) data. It works by storing all records of the training set in an  $n$ -dimensional structure (where  $n$ —number of features) and then predicts/classifies the new record based on the target variables of the nearby  $k$  neighbouring records in the training set.

The main parameters that determine the accuracy of the kNN model are 'Number of Neighbors' ( $k$ ) and 'Distance metric'.  $k$  is the number of neighbouring records that the algorithm considers to compute the target variables. The best value here is chosen as 1 after iterating the value of  $k$  from 1 to 10 and running the algorithm. Distance metric is the method of calculation that is followed to obtain the distance between

	Linear Regression	Decision Tree	KNN Regressor	Random Forest
mean	5.741048	6.313830	1.827746	6.430078
std	5.113682	5.986571	2.839027	5.679037
min	0.149072	0.000000	0.000000	0.039986
25%	1.689743	1.817757	0.022123	2.363121
50%	4.149239	4.615785	0.608377	4.964046
75%	8.200180	8.885487	2.322895	8.799984
max	27.479852	26.437676	16.691658	34.957840

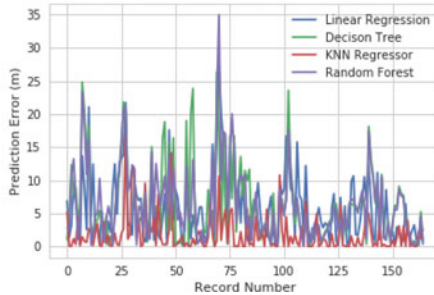
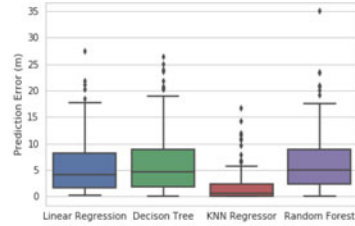


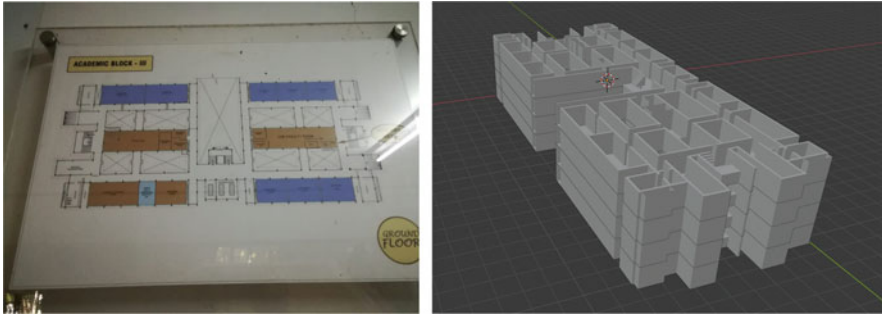
Fig. 7 Results of the tested algorithms

Table 1 kNN model fine-tuning

Distance metric	Formula	n_neighbours	Mean positioning error (m)
Manhattan	$d = \sum  P_i - Q_i $	1	1.942
Euclidean	$d = \sqrt{\sum  P_i - Q_i ^2}$	1	1.827
Chebyshev	$d = \max  P_i - Q_i $	1	2.214
Canberra	$d = \sum  P_i - Q_i  / (P_i + Q_i)$	1	1.810
Bray-Curtis	$BC_{ij} = \sum \frac{ n_{ik} - n_{jk} }{(n_{ik} + n_{jk})}$	1	1.917

two points (records) i.e. the unknown record (to be predicted) and a known record in the dataset. Metrics such as Euclidean, Manhattan, Canberra, Chebyshev and Bray-Curtis have been tried out, with Canberra providing the least mean positioning error for the data (Table 1).

Since kNN proves to be the best algorithm (from the tested ones) for the given dataset, the same is then built from scratch using TensorFlow, so that it can be saved as a TensorFlow Lite (.tflite) model and then used in the Android application. The saved TensorFlow Lite model is then hosted on Google Firebase and linked with the app through Firebase’s ML Kit library. This helps in downloading the model to the device. In case there are changes, the model hosted on Firebase can be swapped with the new one and the model can be downloaded for offline usage again.



**Fig. 8** (left) Available floor plan of Academic Block 3 (2020), Amrita University and (right) model built with Blender

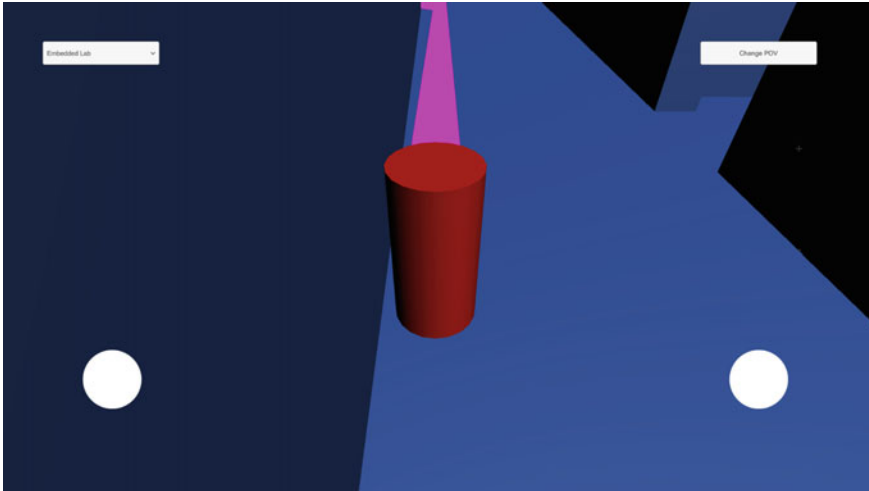
#### ***4.4 3D Building Model—Blender Design***

Blender is an open-source software which is mostly used for modelling objects and animating films and sometimes for MRI scans. With a photo of the building's floor layout that was available, a rough layout of the top view of the building is created using Blender. With the said top view, the walls of the building are designed by projecting it to a plane perpendicular to the top view. Furthermore, planes are added to the base of the walls to make up for the flooring of the building and stairs are modelled from cubes. This results in the completion a floor of the building. The rest of the floors are then duplicated from the initial ground floor which results in making the required building. On completion of the design, it is imported to Unity as a FilmBox (.fbx) model (Fig. 8).

#### ***4.5 3D Building Model—Unity***

Unity is a cross-platform game engine used to develop video games for web plugins, desktop platforms, consoles and mobile devices. GameObjects in Unity are building blocks of the application that help one visualize various objects and setting of our building. The following are the GameObjects used in the solution:

- **The User**—Represents the person using the app. It shows the location of the user in the building. The user is represented by a cylinder. The user is tuned to move towards a destination represented by a Sphere.
- **Cameras**—There are 2 cameras—One for the 3rd Person Point of View (PoV) (Fig. 9) and the other for an Eagle Eye View (Fig. 10). In 3rd Person PoV, the user can follow the path laid out in front of him/her to reach the destination. This mode shows the location of user while traversing the assigned path. The Eagle Eye PoV gives an aerial view of the building. In this mode, the person can look at different



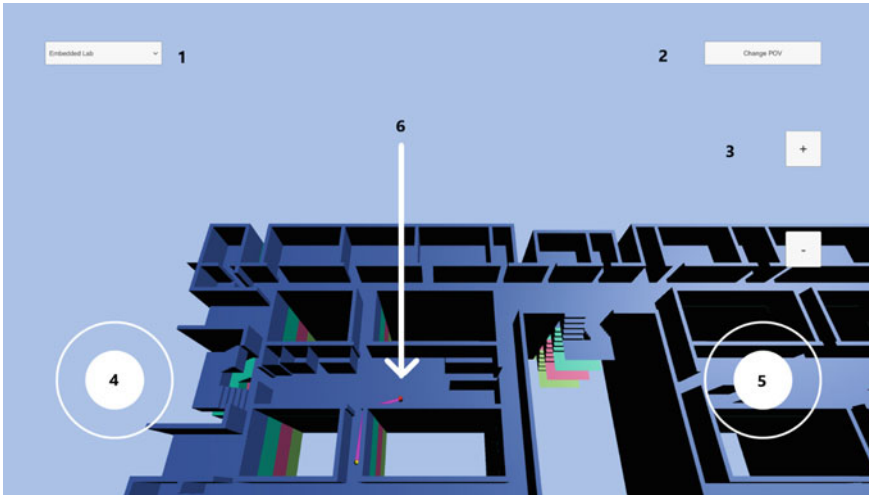
**Fig. 9** User in third person PoV

rooms of interest in the building via the dropdown menu located in the top left corner.

- Sphere—A spherical object which represents the destination to be reached. Its location is assigned based on the choice selected from the drop down menu.
- AB3—Represents the model of the building that was chosen for the experiment. In our case, it's the Academic Block 3 located in Amrita Vishwa Vidyapeetham, Coimbatore, India. It was initially modelled in Blender and then later on imported to Unity. Different colours have been used to represent the different floors of the building.
- EventManager and Manager—Co-ordinate in managing events and the state of control in Unity. Here, it helps in the transition between Aerial PoV and 3rd Person PoV, a zooming functionality and listing the rooms of interest to the user. Both act as a link between independent GameObjects and prevent them from malfunctioning when executing the Components (mesh renderer, scripts, transform) attached to them.
- Canvas—It is the region where all UI elements reside. The Canvas is a GameObject with a Canvas component in it, and all UI elements are children of the canvas and

Components are elements that can be attached to GameObjects for functionality. Custom Components can be created with the help of scripts that assist in achieving the desired functionality. The following are the scripts written for the solution:

- AerialScript.cs—Responsible for the aerial view provided by the two joysticks “joymovement” and “joyrotation”. The “joymovement” joystick is responsible for moving in a 2D plane and The “joyrotation” joystick is used for not only



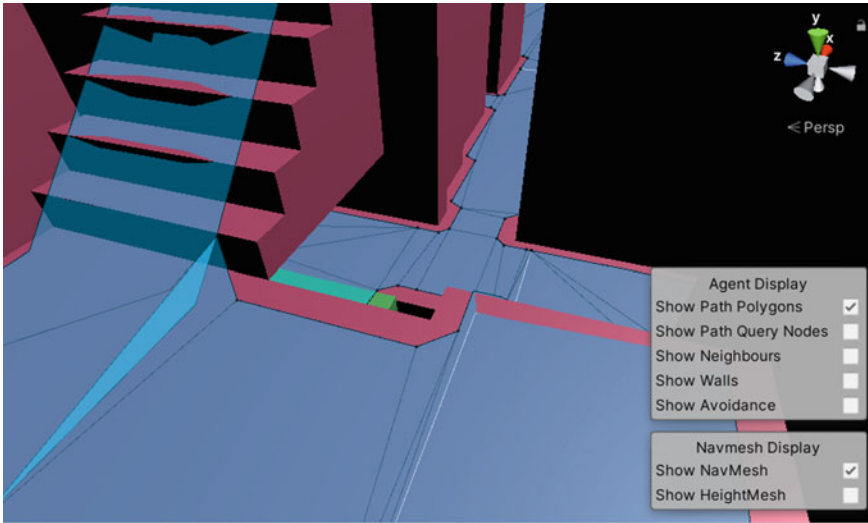
**Fig. 10** Various GameObjects—1. List of rooms, 2. Changing PoV, 3. Zooming functions, 4. Joystick for translation, 5. Joystick for rotation and changing floors, 6. User location

rotating around but also viewing the different floors which are provided by moving vertically.

- CameraChange.cs—Allows one to switch between the cameras (aerial and first person view) whenever the Change PoV button is pressed.
- Zoom.cs—Responsible for zooming in and out of the Aerial View. This helps in tailoring users’ needs such as viewing where they are in the building, what they’re surrounded by, what all facilities are available in the building etc.
- DropDown.cs—Populates the list of locations (of interest) in the building in the dropdown menu present in the app.
- FPScript.cs—Accounts for the 3rd Person’s PoV and shows both the user and the path he/she is following.
- Navigation.cs—Responsible for the navigational aspect. Firstly, a NavMesh (Path for the object to move on) is baked out of the AB3 Model (Fig. 11). A NavMeshAgent is an object capable of traversing through the NavMesh. The path is calculated by Dijkstra’s Shortest Path algorithm whenever a destination is assigned.

The light blue mesh shows navigable regions of the user. After baking the mesh, a NavMeshAgent has to be assigned. In our case, it is the Cylinder (User). Its destination is always assigned where the Sphere resides.

The finalized Unity project is then built as a Gradle project and then integrated into the Android application as an individual Activity (refer Fig. 4).



**Fig. 11** Building NavMesh

## 5 Conclusion and Future Work

The aim of this paper is to illustrate a semi-dynamic and offline setup for 3D indoor navigation. This semi-dynamic setup that is proposed overcomes a major limitation seen in conventional indoor navigation systems—Repeating the same task over and over again for different scenarios.

The goal of the paper isn't to provide the most accurate model for indoor positioning but to devise a modular framework that helps achieve indoor navigation. But then again, the accuracy of an indoor navigation system is still a vital factor and hence, various Machine Learning algorithms have been tested to find the best suited one for the task at hand.

The objective of a mobile framework, easy-to-use UI, readily available and zero-cost solution for offline indoor navigation has been completed. Devices that a common man possesses (smartphone, computer) are used to make the framework as accessible as possible but this could be improvised in future. For example, if all phones turned out to have barometric sensors, height could easily be calculated in metres rather than floor number. Accelerometer and gyroscopic sensors could be used to provide a completely offline solution.

Making a few changes to the infrastructure involved, the framework proposed in this paper can be implemented in specific domains such as:

- Shopping—An AI-powered shop assistant that locates the required items
- Hospitals—To monitor the whereabouts of the patients
- Game Industry—Augmented Reality (AR) games that transform real indoor spaces into virtual playgrounds

- Information Kiosks—In exhibitions and similar spaces to provide the required help without any manual labour.

The scope for indoor navigation systems has not reached an end since technological advancements in the fields of pocket devices and location-based services are still going strong. The primary tool for indoor navigation could switch from smartphones to smart bands or smart watches soon as the pocket devices are gaining users. It is no wonder that indoor navigation has been a popular topic for years, with tons of new ideas and approaches being generated every now and then.

## References

1. Kjærgaard MB, Blunck H, Godsk T, Toftkjær T, Christensen DL, Grønbæk K (2010) Indoor positioning using GPS revisited. In: Floréen P, Krüger A, Spasojevic M (eds) *Pervasive computing. Pervasive 2010. Lecture notes in computer science*, vol 6030. Springer, Berlin, Heidelberg
2. Mendoza-Silva GM et al (2019) A meta-review of indoor positioning systems. *Sensors (Basel, Switzerland)* 19(20):4507
3. Brena RF, García-Vázquez JP, Galván-Tejada CE, Muñoz-Rodríguez D, Vargas-Rosales C, Fangmeyer J (2017) Evolution of indoor positioning technologies: a survey. *J Sens* 2017:1–21
4. Yang C, Shao H (2015) WiFi-based indoor positioning. *IEEE Commun Mag* 53(3):150–157
5. Sakpere W, Adeyeye Oshin M, Mlitwa NB (2017) A state-of-the-art survey of indoor positioning and navigation systems and technologies. *South Afr Comput J* 29(3)
6. Liu Z, Dai B, Wan X, Li X (2019) Hybrid wireless fingerprint indoor localization method based on a convolutional neural network. *Sensors* 19(20):4597
7. Ge X, Qu Z (2016) Optimization WIFI indoor positioning KNN algorithm location-based fingerprint. In: 2016 7th IEEE international conference on software engineering and service science (ICSESS), Beijing, 2016, pp 135–137
8. Cramariuc A, Huttunen H, Lohan ES (2016) Clustering benefits in mobile-centric WiFi positioning in multi-floor buildings. In: 2016 international conference on localization and GNSS (ICL-GNSS), Barcelona, 2016, pp 1–6
9. Qi G, Jin Y, Yan J (2018) RSSI-based floor localization using principal component analysis and ensemble extreme learning machine technique. In: 2018 IEEE 23rd international conference on digital signal processing (DSP), Shanghai, China, 2018, pp 1–5
10. Barnwal S, Peng W (2019) Crowdsensing-based WiFi indoor localization using feed-forward multilayer perceptron regressor. In: 2019 international conference on computational intelligence in data science (ICCIDS), Chennai, India, 2019, pp 1–6
11. Zafari F, Papapanagiotou I, Hacker TJ (2018) A novel Bayesian filtering based algorithm for RSSI-based indoor localization. In: 2018 IEEE international conference on communications (ICC), Kansas City, MO, 2018, pp 1–7
12. Halder SJ, Giri P, Kim W (2015) Advanced smoothing approach of RSSI and LQI for indoor localization system. *Int J Distrib Sens Netw* 2015
13. Hsieh C, Chen J, Nien B (2019) Deep learning-based indoor localization using received signal strength and channel state information. *IEEE Access* 7:33256–33267
14. Maduskar D, Tapaswi S (2017) RSSI based adaptive indoor location tracker. *Sci Phone Apps Mob Dev* 3(1):1
15. Chai W, Chen C, Edwan E, Zhang J, Loffeld O (2012) 2D/3D indoor navigation based on multi-sensor assisted pedestrian navigation in Wi-Fi environments. In: 2012 ubiquitous positioning, indoor navigation, and location based service (UPINLBS), Helsinki, 2012, pp 1–7

16. Landa V, Ben-Moshe B, Hacoheh S, Shvalb N (2018) GoIn—an accurate 3D indoor navigation framework for mobile devices. In: 2018 international conference on indoor positioning and indoor navigation (IPIN), Nantes, 2018, pp 1–8
17. Kim H-H, Kang B-G (2014) A study on the implementation of a 3-dimensional positioning system on indoor environments. In: International conference on advanced technologies for communications (ATC), 2014, pp 750–753
18. Dong J, Noreikis M, Xiao Y, Ylä-Jääski A (2019) ViNav: a vision-based indoor navigation system for smartphones. *IEEE Trans Mob Comput* 18(6):1461–1475
19. Bisio I, Sciarrone A, Bedogni L, Bononi L (2018) WiFi meets barometer: smartphone-based 3D indoor positioning method. In: 2018 IEEE international conference on communications (ICC), Kansas City, MO, 2018, pp 1–6
20. Liu K, Motta G, Tunçer B, Abuhashish I (2017) A 2D and 3D indoor mapping approach for virtual navigation services. In: 2017 IEEE symposium on service-oriented system engineering (SOSE), San Francisco, CA, 2017, pp 102–107
21. Pedregosa et al (2011) Scikit-learn: machine learning in Python. *JMLR* 12, pp 2825–2830
22. Rithin Paul Reddy K, Sai Srija S, Karthi R, Geetha P (2020) Evaluation of water body extraction from satellite images using open-source tools. *Intelligent systems, technologies and applications*, pp 129–140
23. Loganathan A, Bharathi D (2015) Sparsification of graph Laplacian for image indexing using multi-dimensional spectral hashing. *Int J Imaging Robotics* 43–56



# Meanderline Pattern Wearable Textile Antenna for Position Identification in Military Applications



Pruthvi Tenginakai, V. Keerthana, Sowmini Srinath, Fauzan Syed, and P. Parimala

**Abstract** In defence applications, tracking of every individual is necessary. The devices and techniques used in the present world are manual and externally equipped. The automation of real-time tracking of every soldier without any external hindrance is the need of the hour. This task can be accomplished by designing an antenna with high frequency. The designed antenna is fabricated onto a textile (jeans) material which will be worn by the soldiers during training or emergency periods. The antenna is designed to operate at Wi-Fi frequency (2.4 GHz) for simplicity of sensor (GPS) integration and utility. It is interfaced with the Arduino board which acts like a transceiver to send the real-time location of the soldier in the form of latitudes and longitudes. This setup is completely automated and integrated on the clothing of the soldier, which eliminates carrying any extra device for tracking himself. The designed meander line patch antenna on the ‘jeans’ textile material has the specifications which are optimised to produce maximum directivity and gain despite the human torso.

---

P. Tenginakai (✉)

Department of Telecommunication Engineering, Ramaiah Institute of Technology, #E 601 HRC Ibbani Apartment, Jakkuru, Bangalore 560064, India  
e-mail: [pruthvijayadev@gmail.com](mailto:pruthvijayadev@gmail.com)

V. Keerthana

Department of Telecommunication Engineering, Ramaiah Institute of Technology, No.534, 8th Main, Vijayanagar, Bangalore 560040, India  
e-mail: [keerthanavijay01@gmail.com](mailto:keerthanavijay01@gmail.com)

S. Srinath

Department of Telecommunication Engineering, Ramaiah Institute of Technology, 3#C 808, Vaishnavi Ars Gardenia, SM Road, T. Dasarahalli Near Jalahalli Cross, Bangalore 560057, India  
e-mail: [sowminisrinath6@gmail.com](mailto:sowminisrinath6@gmail.com)

F. Syed

Department of Telecommunication Engineering, Ramaiah Institute of Technology, 4#214, 3rd Main, 10th Block, Talacauvery Layout, Amruthahalli, Byatarayanapura, Bangalore 560092, India  
e-mail: [funfauzan@gmail.com](mailto:funfauzan@gmail.com)

P. Parimala

Department of ETE, MSRIT, Mattikere, Bangalore 560054, India  
e-mail: [parimalap@msrit.edu](mailto:parimalap@msrit.edu)

Serial communication is used to transmit and receive sensor data via a pre-coded Arduino board. The project can also be extended to several other applications such as heart monitoring of soldiers, etc.

**Keywords** Microstrip antenna · Meanderline antenna · SPI communication · Edge line feed technique · HFSS

## 1 Introduction

Antennas have been the main communication component over the years. The miniaturisation of antenna technology is the need of the present time. When an antenna is etched on a board using microstrip techniques, it is called a microstrip antenna. They are used at microwave frequencies, ranging from 1 to 1000 GHz. A thin metal foil of various shapes on the surface of a custom substrate like ‘jeans’ with dielectric constant 1.67 and loss tangent of 0.025, with a metal foil ground plane on the other side of the board. An RF transmitting wearable antenna for real-time position identification of soldiers, working at 2.4 GHz is designed. An arising category of textile-based communication systems can be used. A wearable device is a technology that is worn on the human body. The signals or pulse is sensed by the wearable device can be mounted on the person to be monitored.

Bending a wire in a calculated geometrical configuration, such as the structure of the meander line is a useful way to shorten the conventional linear wire antennas [1]. A 20–50 mm ground plane and 10–20 printed meander-line antenna comprise the RF section.

Experimental calculation, fabrication, and detailed simulation [2] of a 2.4 GHz printed dipole antenna is presented for wireless communication applications. The perfectly matched Berenger layer absorbing boundary condition is used for the parameter computation. The characteristics of the antenna radiation include input standing-wave ratio (VSWR), radiation patterns, and diversity in polarisation are measured.

A suitable design for high data rate application is reported by a slotted rectangular patch microstrip antenna [3]. The antenna is designed for a frequency of 2.4 GHz and is fed by a microstrip line. The high-frequency structural simulator (HFSS) software is used to design and simulate this antenna.

Internet of things (IoT) applications use printed inverted F antenna (PIFA) with meandering line [4] and meandering shorting strip under 2.4 GHz in the industrial, scientific, and medical (ISM) band. Commercial and medical applications use printed circuit board (PCB) because of its compactness, low profile, and cheaper cost compared to low temperature co-fired ceramic (LTCC) technology. The efficiency and bandwidth of the PIFA increase when conventional PCB line is replaced by PIFA.

The complete module of position identification system using meanderline antenna consists of transmitter and receiver devices. Transmitter end as shown in Fig. 1 of

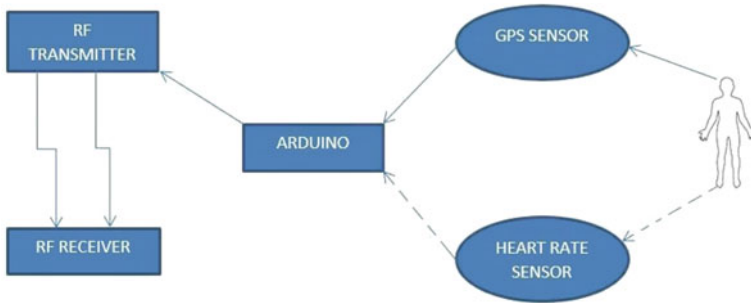


Fig. 1 Block diagram—transmitter end

the communication includes the designed transceiver, a GPS sensor setup bridged by an Arduino microcontroller. The GPS sensor picks up the latitude and longitude readings and the controller analyses this data and this is transmitted through the RF module. Based on the same orientation, an extension of this project is that various health parameters of a soldier such as temperature, blood pressure can be monitored too.

The block diagram of the receiver end is illustrated in Fig. 2 that consists of the designed transceiver, Arduino microcontroller. The data from the GPS module is collected in CSV format which in turn is plotted on a pre-loaded map of the war zone/training zone.

The overall functional block diagram can be designed as shown in Fig. 3. Every individual soldier has a meanderline patch antenna on this textile (clothes) which is wirelessly connected to the GPS module and other optional setups. The data from every GPS setup is received by the microcontroller and processed further. This collective data is handled by the master microcontroller at the military base station to setup and track the strategies and movement of the entire contingent.

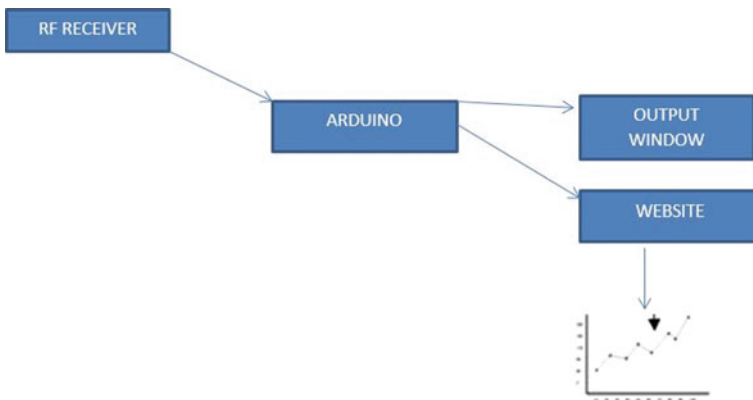
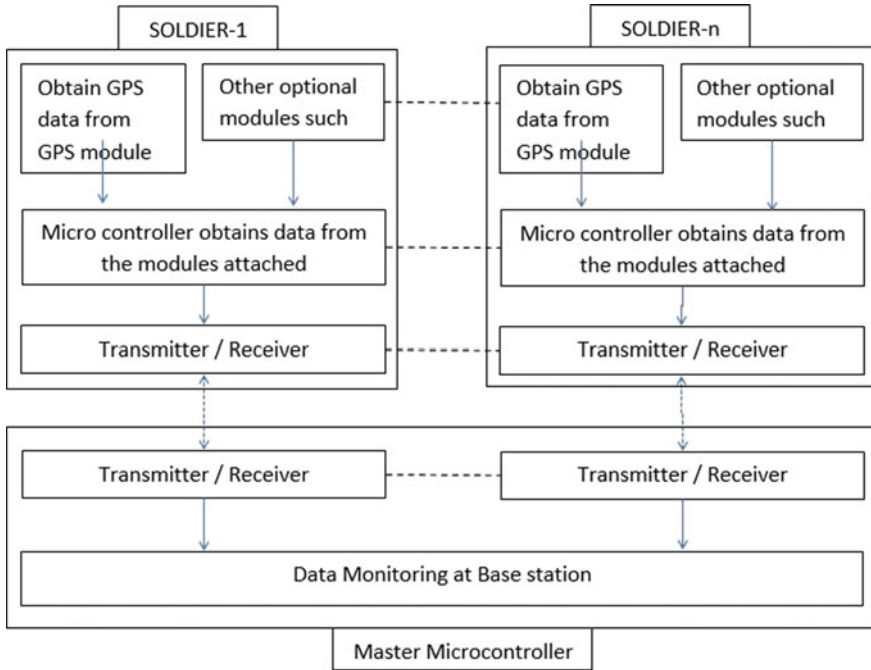


Fig. 2 Block diagram—receiver end



**Fig. 3** Functional block diagram

In Fig. 3, other optional modules include heart rate monitoring, pulse monitoring, and various other parameters of the soldier.

## 2 Design of Meander Line Antenna

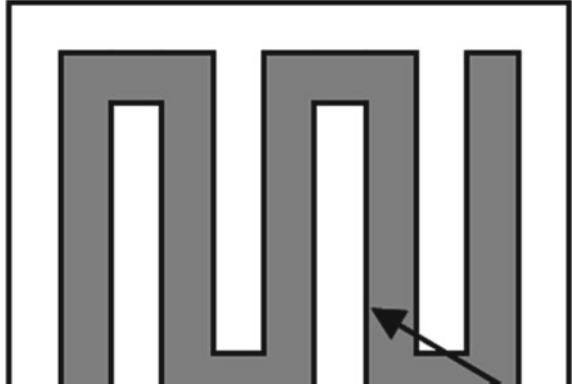
A meanderline antenna is a straight conductor line folded into horizontal and vertical lines which form turns to reduce the total length of antenna smaller than the original length. It looks like a zig-zag pattern as shown in Fig. 4. The number of conductive turns is directly proportional to the output efficiency of the meanderline antenna. This makes the antenna affordable, weightless, more efficient, better electrical characteristics.

The width of the antenna is found using Eq. (1).

$$\text{Width of patch } (W) = \frac{C}{2 \times F_r} \times \left( \frac{\epsilon_r + 1}{2} \right)^{-\frac{1}{2}} \tag{1}$$

where,

**Fig. 4** Meanderline structure



$c$  = speed of light ( $3 \times 10^8$  m/s)  
 $F_r$  = operating frequency  
 $\epsilon_r$  = dielectric constant.

The length of the antenna is found using Eq. (2)

$$\text{Length of patch } (L) = L_{\text{eff}} - (2 * \Delta L) \tag{2}$$

where

$L_{\text{eff}}$  = effective length calculated by Eq. (3)

$$\text{Effective length } (L_{\text{eff}}) = \frac{C}{2 * (F_r) * \sqrt{\epsilon_e}} \tag{3}$$

where

$\epsilon_e$  = effective permittivity.

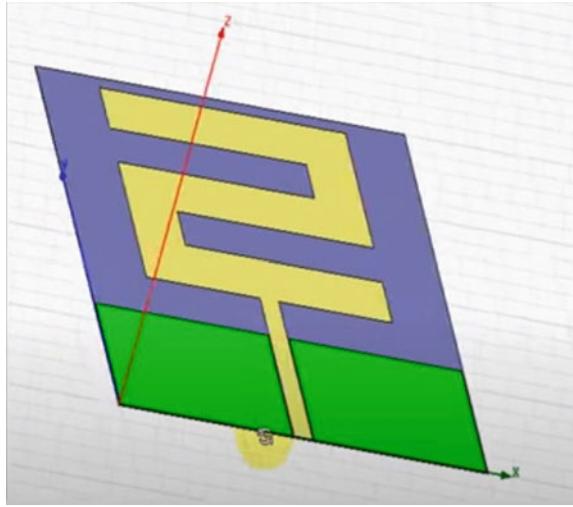
The effective dielectric constant is given by Eq. (4)

$$\epsilon_e = \left( \frac{\epsilon_r + 1}{2} \right) + \left( \frac{\epsilon_r - 1}{2} \right) \times \left( 1 + 12 * \frac{h}{w} \right)^{-\frac{1}{2}} \tag{4}$$

where

$h$  = height of dielectric material (4 mm)  
 $w$  = calculated width of the antenna.

**Fig. 5** Rectangular antenna with meanderline pattern



The length of extension ( $\Delta L$ ) is found using Eq. (5),

$$(\Delta L) = 0.412 \times \left( \frac{(\epsilon_e + 0.3)\left(\frac{w}{h} + 0.264\right)}{(\epsilon_e - 0.258)\left(\frac{w}{h} + 0.8\right)} \right) \tag{5}$$

The ground plane dimensions are calculated using Eqs. (6) and (7),

$$\text{Length of ground}(L_g) = 6h + L \tag{6}$$

$$\text{Width of ground}(W_g) = 6h + w \tag{7}$$

The width of the designed patch antenna is 54.09 mm, and length is 55.4346 mm with the ground dimensions being (79.4346 mm  $\times$  78.09 mm). The antenna is powered by an edge line-feed cable. The meanderline is considered up to the second iteration. The dielectric material used is a custom ‘jeans’ with the dielectric constant of 1.67, loss tangent of 0.025 and thickness 4 mm as shown in Fig. 5.

### 3 Software Overview

Soldier identification module can be supported by the antenna simulation tool ANSYS HFSS and Arduino UNO, user interface.

- A. Antenna simulation—ANSYS high-frequency system simulator (HFSS) is a 3D electromagnetic simulation software for designing high-frequency electronic

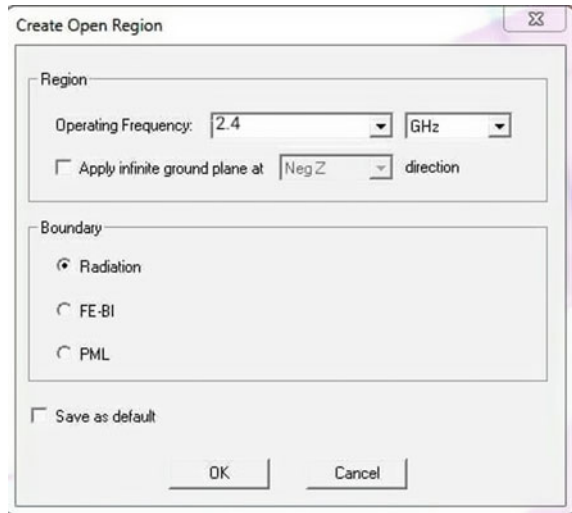
- products such as antennas, antenna arrays, RF or microwave components, high-speed interconnects, filters, connectors, IC packages and printed circuit boards.
- B. User Interface—Arduino UNO is a widely used and affordable open-source microcontroller board based on the ATmega328P microcontroller. The board consists of 14 digital pins and 6 analogue pins. The Arduino IDE can be used via B type USB cable for programming the board accordingly.
  - C. Arduino libraries—Include **RF24 libraries** and create an RF24 object. Usage of reading, listen, set PA level and channel pipe functions are exploited to bridge the communication gap between transmitter and receiver.

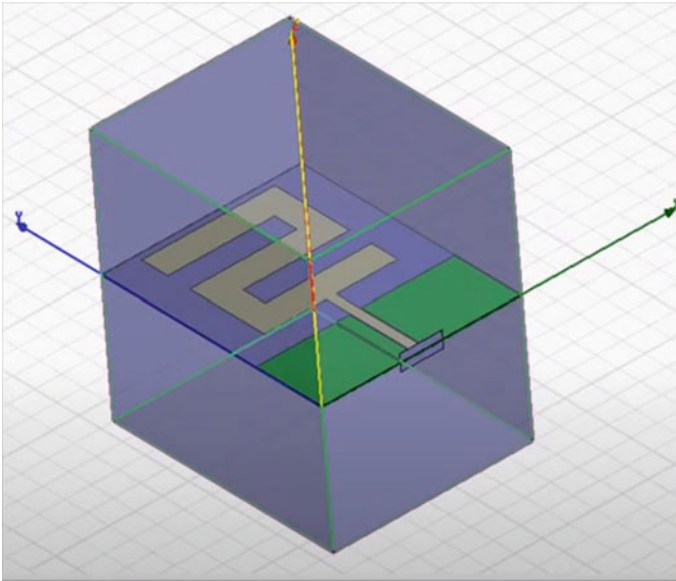
### 4 Antenna Simulation

The antenna simulation is designed as single element meanderline antenna to operate at the frequency of 2.4 GHz. After creating all the variables required initially, assign variables, create a dielectric substance (jeans) as shown in Fig. 6 dropdown box in HFSS tool.

By truncating infinite free space to a finite calculation domain, the radiation boundary is used to mimic free space. A bugger distance of  $\lambda/4$  from the radiating surface is taken from all directions. Figure 7 represents the antenna with a visible radiation box.

**Fig. 6** Radiation box dropdown box with operating frequency 2.4 GHz





**Fig. 7** Meanderline antenna design with radiation box

## 5 Results and Discussion

The simulation of the meanderline pattern antenna was successful using the HFSS v18 software. The impedance matching, VSWR,  $S$ -parameters, gain, directivity, 3D radiation pattern and other such parameters were analysed.

Figure 8 representing the  $S$ -parameters has a major lobe and minor lobe at 2.4 GHz and 1.86 GHz, respectively. The minor lobe has to be eliminated to avoid loss, this can be achieved by using a reflector in the  $Z$  direction.

Figure 9 represents the optimised  $S$  parameter graph after inclusion of the reflector plate. The antenna is optimised to 2.4 GHz only.

The voltage standing wave ratio (VSWR) is a measure of the efficiency of radiation which should be 1 for ideal conditions. Figure 10 shows the VSWR is 1.27 which is almost equal to ideal conditions.

The antenna is given a frequency sweep of the desired duration and set to 2.4 GHz functionality.

Table 1 shows the important parameters of antenna design like operating frequency, VSWR, return loss obtained with the design values calculated.



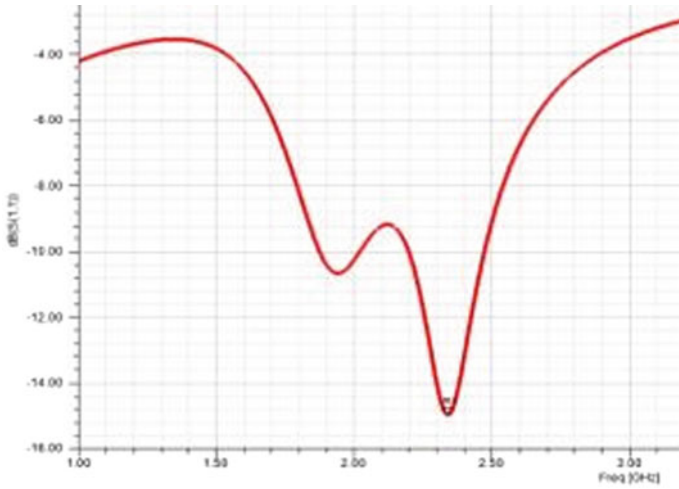


Fig. 8 S-parameters with jeans substrate

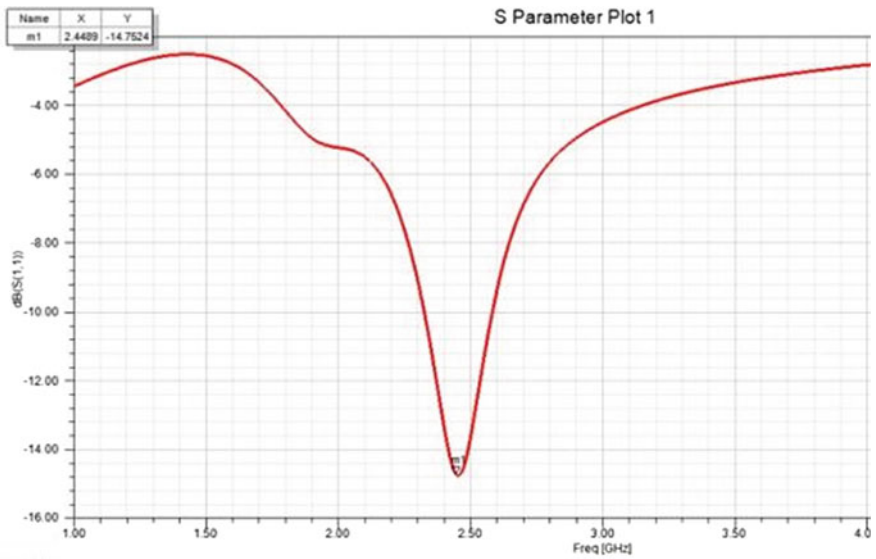
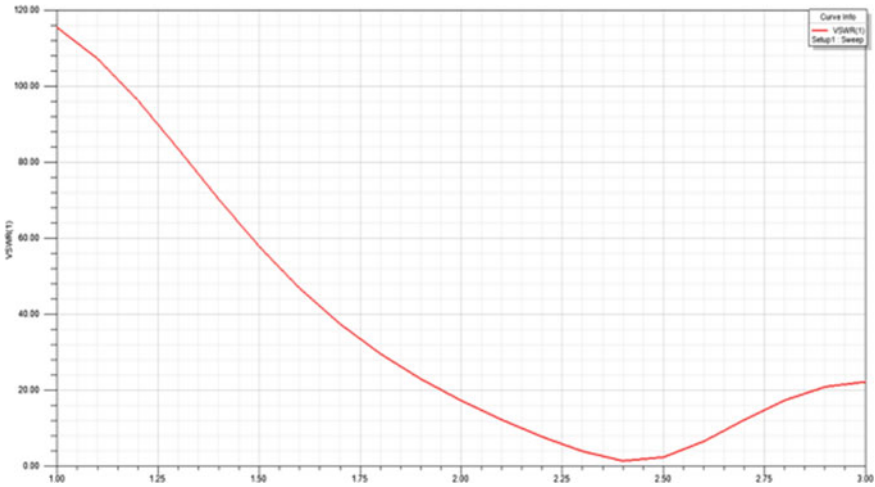


Fig. 9 S-parameters with jeans substrate and reflector

The antenna was simulated, and the radiation patterns are as follows:

Figure 11 represents the directivity of the designed antenna, on using a reflector. An increased directivity is observed and is more focused.



**Fig. 10** VSWR plot of antenna designed for 2.4 GHz

**Table 1** Meanderline antenna parameters obtained

S. No.	Parameters	Simulation results of the antenna
1.	Operating frequency	2.4 GHz
2.	VSWR	1.27
3.	Return loss	18 dB

Figure 12 represents the loss in dB in and around the antenna, the loss is highest at the centre of the antenna and decreases gradually resulting in more radiation in all other directions.

Figure 13 shows the 3-D representation of the current emitted by the designed antenna. The radiation is in the form of a well-directed beam in the X and Y directions can be observed but obsolete in the Z due to the reflector.

Figure 14 represents the antenna gain to distance from the epicentre of the antenna. Hence, the gain continues to increase away from the centre of the antenna until the effective maximum reach (range) has been achieved.

The data thus received is imported to a CSV file with latitude and longitudinal values as shown in Fig. 15. These values are further used to pinpoint the real-time position of the soldier.

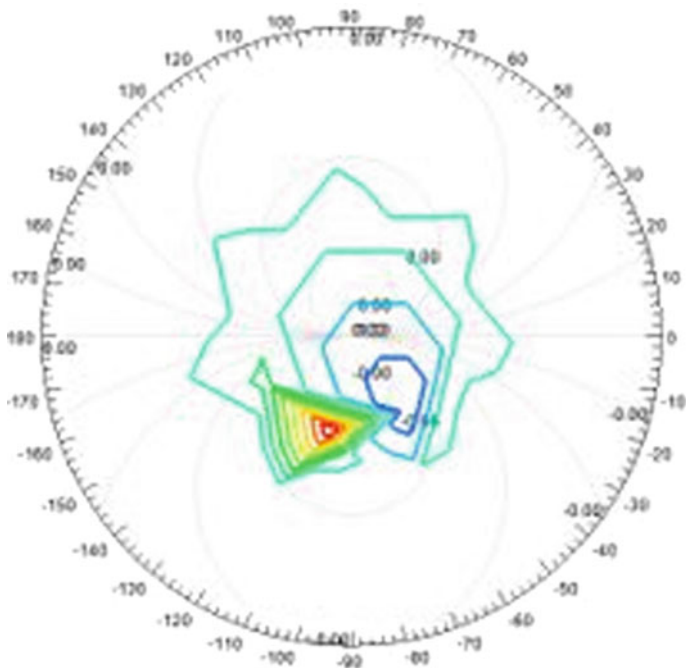


Fig. 11 Designed meanderline antenna directivity plot

Latitude	Longitude	(x, y)
-77.14456892	35.988677026	(1, 1)
-77.26936334	36.176730488	(2, 2)
-78.00000000	36.768578404	(3, 3)
-78.45372920	37.638400000	(4, 4)
-78.78574847	37.988364738	(5, 5)
-78.98335728	38.000000000	(6, 6)
-79.00003527	38.374872809	(7, 7)
-79.37874873	38.562730937	(8, 8)
-79.63807465	38.876347363	(9, 9)
-79.83746374	39.000034733	(10, 10)
-80.00000000	39.383483748	(11, 11)
-77.14456892	35.988677026	(12, 12)
-77.26936334	36.176730488	(13, 13)

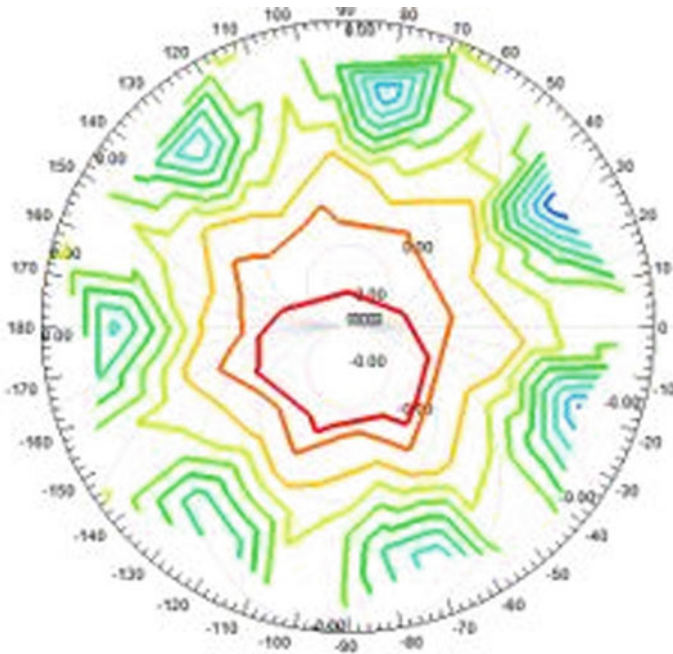


Fig. 12 dB loss representation of designed by meanderline antenna

### S Parameter Plot 1

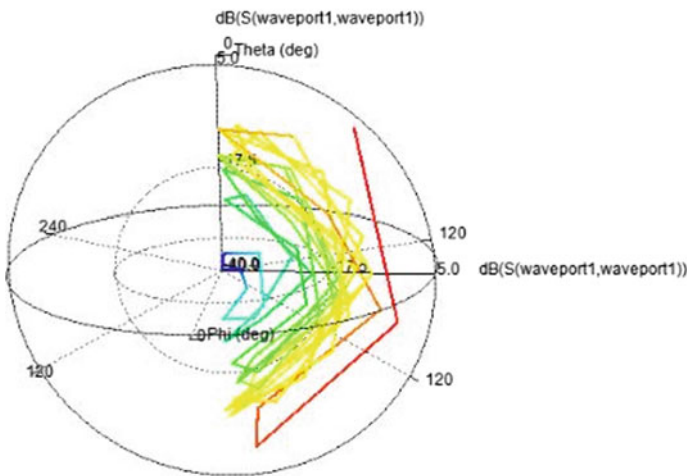


Fig. 13 Surface current distribution

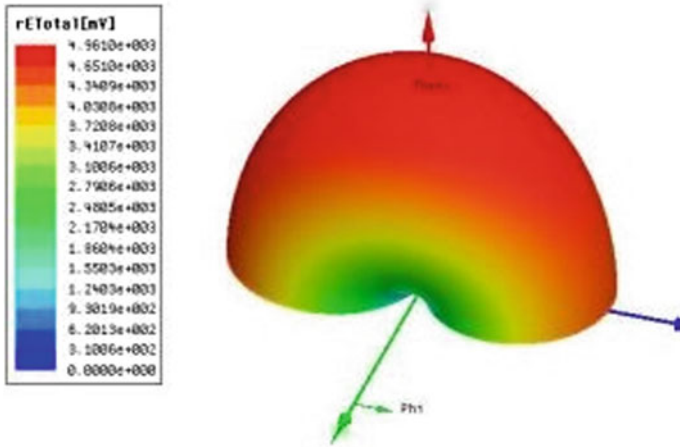


Fig. 14 Meanderline antenna gain plot

Fig. 15 CSV data containing lat/long data from the GPS module

Latitude	Longitude	(x,y)
-77.14456892	35.988677026	(1,1)
-77.26936334	36.176730488	(2,2)
-78.00000000	36.768578404	(3,3)
-78.45372920	37.638400000	(4,4)
-78.78574847	37.988364738	(5,5)
-78.98335728	38.000000000	(6,6)
-79.00003527	38.374872809	(7,7)
-79.37874873	38.562730937	(8,8)
-79.63807465	38.876347363	(9,9)
-79.83746374	39.000034733	(10,10)
-80.00000000	39.383483748	(11,11)
-77.14456892	35.988677026	(12,12)
-77.26936334	36.176730488	(13,13)

## 6 Conclusion

Meanderline pattern antenna simulated using HFSS software has the results tabulated and plotted. The sensed data was successfully transmitted through the designed transmitting antenna to the receiving antenna which was designed to work at 2.4 GHz, which is standard Wi-Fi frequency. The functioning frequency of the antenna can be varied in accordance with its dimensions, to secure the transmission line from an external threat by using the calculations explained in (II). The bandwidth of the designed meanderline antenna is found to be 100 MHz, and the range of transmission was approximately 100 m. Non-line of sight transmission was achieved successfully.

The size of the entire system can be made compact by restricting the number of pins on the Arduino and RF module based on their usage. The antenna size can also be minimised by using fabric as a substrate and by implementing embroidery radiating materials.

This large amount of data collected can be used in predicting the future strategies and performance of the soldiers individually in the war field.

## References

1. Chuang H-R, Kuo L-C (2003) 3-D FDTD design analysis of a 2.4 GHz polarization-diversity printed dipole-antenna with integrated balun, polarization-switching circuit for WLAN and wireless communication applications. *IEEE Trans Microwave Theor Tech* 51(2):374–381
2. Soras C et al (2002) Analysis and design of an inverted-F antenna printed on a PCMCIA card for the 2.4 GHz ISM band. *IEEE Antennas Propagat Mag* 44(2):37–44
3. Lin C-C, Kuo S-W, Chuang H-R (2004) A 2.4 GHz printed Meanderline antenna for WLAN applications. *IEEE AP-S Int Dig* 3:2767–2770
4. Marrocco G (2003) Gain-optimized self-resonant meander line antennas for RFID applications. *IEEE Antennas Wirel Propagat Lett* 2:302–305

# Authentication of Robots Using ECC to Access Cloud-Based Services



Saurabh Jain and Rajesh Doriya

**Abstract** The advancement of internet technologies has led to the appearance of cloud robotics. Besides the numerous benefits of cloud robotics, security threats affect the growth of this field. Therefore, a secure and efficient authentication mechanism are needed to prohibit numbers of outsider and insider attacks present in the system model. Authentication of the robots plays an important role in safely access remote resources and remove the many vulnerabilities present in the network. Elliptic curve cryptography (ECC) gives better results in scarce computing resources and limited energy systems such as cloud robotics. In this paper, an ECC-based authenticating scheme has been proposed to facilitate access to cloud-based robots. The proposed scheme uniquely identifies the robots before using cloud-based robotics services. The security analysis using the proverif tool establishes that the proposed scheme resists many well-known attacks.

**Keywords** Cloud robotics · Authentication · ECC · PKI · Robotic services · Security

## 1 Introduction

For more than five decades, robots have been successfully employed in lieu of humans to perform hazardous and tedious tasks. They have been used in a variety of applications such as health care, defense, space operations, education, etc. Cloud-based robotics provides robot units with privileges like access to big data and shared knowledge, and the ability to send computation-intensive tasks to the higher processing

---

S. Jain (✉) · R. Doriya

Department of Information Technology, National Institute of Technology Raipur, Raipur, India  
e-mail: [sjain.phd2017.it@nitrr.ac.in](mailto:sjain.phd2017.it@nitrr.ac.in)

R. Doriya

e-mail: [rajeshdoriya.it@nitrr.ac.in](mailto:rajeshdoriya.it@nitrr.ac.in)

units in the cloud [1]. Moreover, the knowledge collected by one robot can be used by other robots and the heavy computational tasks can be offloaded to the cloud.

Cloud robotics services are now extensively used in healthcare services, security and crisis management services for smart cities, military missions and rescue operations, etc., resulting in the transaction of sensitive data that should solely be retrieved by authenticated robots. The authors have raised a concern about the security and trust challenges regarding the cloud provider, in the case of a robot trying to connect to the cloud or while executing a task over the cloud. Moreover, storing confidential or sensitive data over the cloud can be dangerous in a low-security environment. Authentication plays an important role in the successful integration of robots and cloud computing services. For identifying the legal users over the insecure communication network, remote user authentication is one of the best mechanisms. Mutual authentication is one of the important aspect, robots, and cloud server both need to authenticate each other, respectively, before using services.

In this paper, the proposed scheme that can be used to authenticate the robots using ECC when they want to access the cloud-based robotic services like simultaneously localization and mapping (SLAM) [2], object detection and recognition, and navigation [3]. ECC provides robust security compared to the other public-key cryptography (PKI) with a small key size and verifies the good result with low-cost.

The rest of this paper is organized as follows; Preliminaries of ECC have been described in Sect. 2. In Sect. 3 of the paper, describes related work. In Sect. 4, the system model has been discussed. The proposed scheme using ECC has been discussed in Sect. 5. In Sect. 6, explains the security analysis. Lastly, Sect. 7 concludes the proposed scheme.

## 2 Preliminaries of ECC

ECC is a perspective to public-key cryptography, based on the algebraic system of elliptic curves over finite fields. ECC used the elliptic curve equation [4] to generate the keys. ECC requires a smaller key size as compared to other public-key cryptography algorithms, such as 256-bit ECC provides equivalent security to 3072-bit RSA.

The Elliptic curve over real numbers is defined as given in the equation:

$$y^2 = x^3 + ax + b, \text{ where } 4a^3 + 27b^2 \neq 0. \quad (1)$$

The set of points on the elliptic curve together with an infinity point P which serves as an identity element forms an abelian group. Where a and b are the elements that define the shape of the elliptic curve.

In ECC, scalar multiplication is very excessively used operation which can be used for different purposes such as encryption/decryption of data, verification of digital signatures, and key generation. Three arithmetic operations namely scalar, point, and field are performed in an elliptic curve. Many researchers have been concerned



with the scalar and point arithmetic for fast solving the complex scalar multiplication operations.

### 3 Related Work

Keeping the aspect of security as a major concern in developing technology, password-based authentication is proposed by Lamport [5] in the year 1981. A different version of this scheme is being introduced in the later year [6–8]. Juang et al. [6] reduce computation cost in limited computation resource environment on the smart card using elliptic curve cryptosystems-based user authentication. Sun et al. [7] found the weaknesses in the work of Juang et al. [6] and improved the work by reduces the storage and computation costs on the smart card. An efficient dynamic key management scheme is presented by Lu et al. [8] for VANETs that achieve user's privacy in the vehicle and also resist the possible collusion from other vehicle users. One of the most economical ways to implement security in embedded devices is to use minimal key size, ECC is one among which offers the same specifically in the case of a constrained environment [9]. Day by day ECC protocol is widening acceptability in multiple security-based scenarios. ECC is one of the vibrant protocols being tested in recent times. An intensive survey is being done by Carla and Sood [10] based on ECC protocol in the year 2011. Thus, in the age, where severe security requirement is in need, though with the availability plethora of public-key cryptographies techniques ECC stands out to be the first pick.

The feature of achieving mutual authentication between smart devices and the server makes it more acceptable. Smart devices using ECC-based authentication were proposed by Abichar et al. [11], Tian et al. [12], and Wu et al. [13] had various limitations. There are protocols proposed by Abichar et al. [11], Tian et al. [12], and Wu et al. [13] which mainly focuses on server-side user authentication. This technique is not too safe as their a high chance that an intruder can play the role of middle man a pretend to become a server to retrieve secure data and information from the user. There is another method of providing mutual authentication with the issuance of a certificate; this particular method is comparably costly and it is being proposed by many researchers [14–17]. This technique is also called certificate-based mutual authentication which sometimes also used as timestamps whose range may vary as per the requirement and need of the instance. In 2014, Moosavi et al. [18] were developed an authentication scheme for IoT networks that are based on ECC for RFID systems.

In 2018, Kumari et al. [19] found that scheme [20] is not achieving mutual authentication as well as vulnerable to other attacks such as offline password-guessing attack, insider attack, and session key agreement. Therefore, they proposed a method that overcomes these well-known attacks. Challa et al. [21] also found that scheme [22] is vulnerable to many attacks such as stolen smart card attacks, user impersonation attacks, and it is also not achieving mutual authentication. Therefore, they proposed

the ECC-based user authentication scheme and it claims to have overcome the known network attacks.

In 2020, Sahoo et al. [23] explored the need for a better authentication scheme for lightweight IoT networks. Therefore, they proposed an ECC-based interoperable lightweight authentication scheme that minimizes the number of well-known attacks, and comparison analysis shows that the proposed scheme minimizes the computation cost as well. Fang et al. [24] also proposed an authentication scheme for the resource-constraint IoT devices which do not only provides mutual authentication but provides forward and end-to-end security too. The proposed scheme works in heterogeneous IoT environments and gives better results in terms of computation and communication costs.

### 4 System Model

In cloud robotics, services are stored in the cloud that can be accessed by robots from the remote location. Figure 1 describes the conceptual authentication model. The different services like SLAM, object recognition, path planning, and navigation, etc., hosted in the cloud. The following steps describe the process:

In this step, client robots send the request to a cloud server for accessing the service hosted by the cloud. Robots should be registered within the cloud for accessing that service. After registration, login credential sends to client robot. In this step, after receiving the login credential the client robot sends a login request, where the authentication server authenticates the client robot. Mutual authentication is one of the important aspect, robots, and cloud server both need to authenticate each other, respectively, before using services. They should agree on a session key that will be used for encryption. In this step, after a successful login, the client robot sends

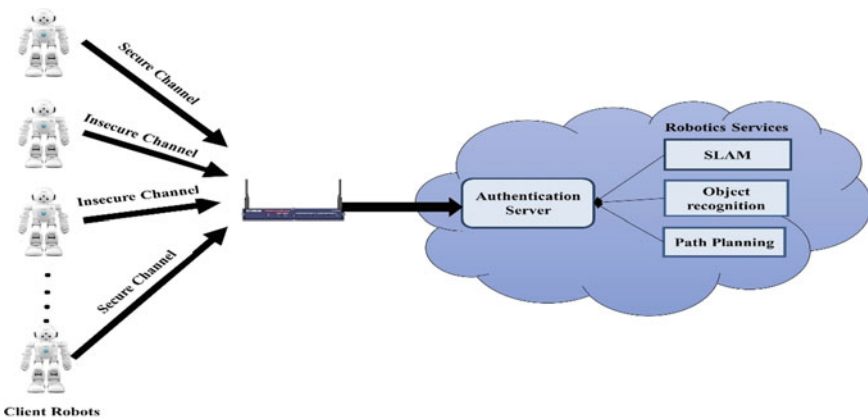


Fig. 1 Conceptual authentication model

the service request to the cloud server in encrypted form. The authentication server forwards a request to the robotic service interface. The interface decrypts the request message and forwards it to an appropriate service host. Finally, robot gets a response in an encrypted form.

## 5 Proposed Scheme

In this section, the proposed scheme will present the importance of authentication of robots to secure access to cloud-based robotics applications and services in the cloud has been shown. Authentication plays an important role in the context of secure access to the resources over the network. The entire work is based on a private cloud setup. The application is deployed over it. The proposed scheme is based on this particular setup. Notations used in the paper are presented in Table 1. The proposed scheme is illustrated in Fig. 2.

### 5.1 Registration Phase

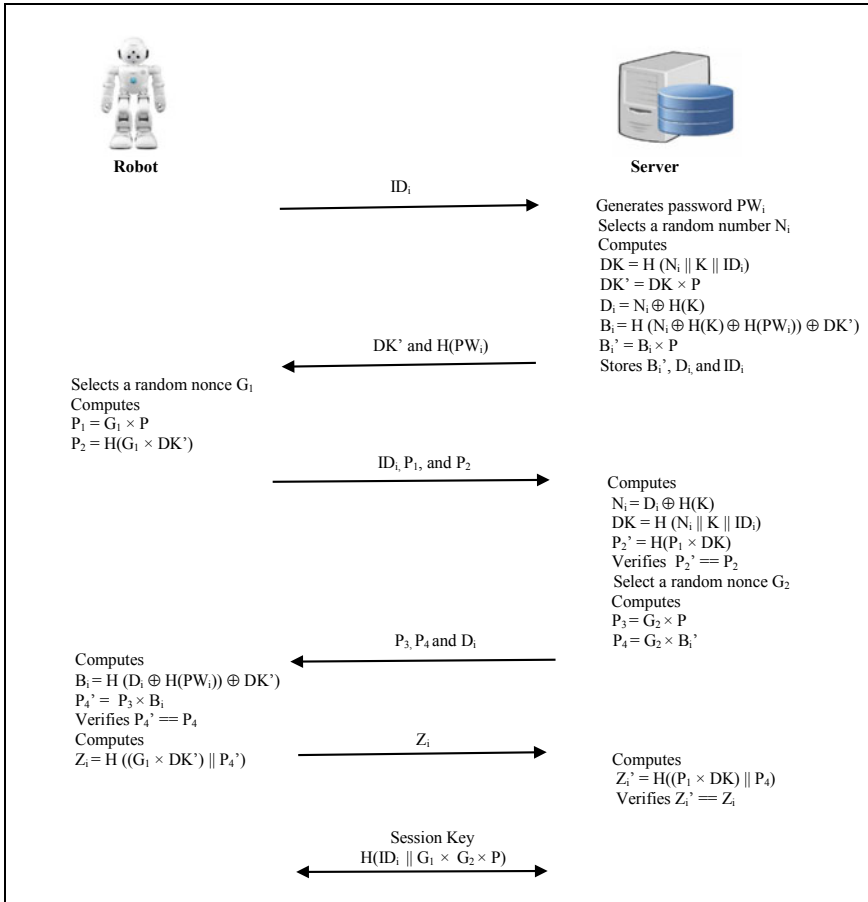
When robots  $R_i$  wants to attain service from the server  $S$ , first  $R_i$  register itself to  $S$  in this phase. For this,  $R_i$  sends its unique id  $ID_i$  to  $S$ . After receiving the unique ids of the robots,  $S$  generates the unique passwords  $PW_i$  and also random nonce  $N_i$  for every  $R_i$ . For this,  $S$  computes

$$DK = H(N_i || K || ID_i),$$

$$DK' = DK \times P,$$

**Table 1** Notation guide

ID	Identity of the robot
$R_i$	Robot
PW	Password of the robot
$P$	Generator point on the elliptic curve with order $n$
$K$	Private key of the server
$S$	Cloud server
$N_1$	Random nonce generated by server
$G_1, G_2$	Random nonce generated by ECC parameters
$\oplus$	XOR operator
$\parallel$	Concatenation operator
DK'	Authentication token
DK, $D_i, B_i, B_i', Z_i$	Variables



**Fig. 2** ECC-based authentication scheme for robots in cloud

$$\begin{aligned}
 D_i &= N_i \oplus H(K), \\
 B_i &= H(N_i \oplus H(K) \oplus H(PW_i) \oplus DK'), \\
 \text{and } B'_i &= B_i \times P.
 \end{aligned}
 \tag{2}$$

where  $K$  is a random number that selects  $S$  as a private key. Finally,  $S$  sends the  $DK'$  and hashed password  $H(PW_i)$  to  $R_i$  and stores  $\{ID_i, D_i, \text{ and } B_i\}$  in its database.

## 5.2 Login Phase

Robot  $R_i$  gets the authentication token  $DK'$  and hashed password  $H(PW_i)$  after successfully registering in the above phase.  $R_i$  can use this authentication token  $DK'$  to get the authentication required message. For getting this,  $R_i$  selects the random nonce  $G_1$  and use  $DK'$  to compute

$$\begin{aligned} P_1 &= G_1 \times P \text{ and} \\ P_2 &= H(G_1 \times DK') \end{aligned} \quad (3)$$

Finally,  $R_i$  sends the login credentials  $\{ID_i, P_1, \text{ and } P_2\}$  to the  $S$  for getting the authentication required message.

## 5.3 Authentication Phase

In this phase,  $S$  receives the login credentials  $\{ID_i, P_1, \text{ and } P_2\}$  from  $R_i$ . Firstly,  $S$  checks the  $ID_i$  of  $R_i$  in its database. If  $S$  finds the records  $\{ID_i, D_i \text{ and } B_i\}$  in its database, then  $S$  uses  $D_i$  and private key  $K$  to compute

$$\begin{aligned} N_i &= D_i \oplus H(K), \\ DK &= H(N_i || K || ID_i), \\ \text{And, } P_2' &= H(P_1 \times DK) \end{aligned} \quad (4)$$

Then,  $S$  verifies the  $P_2' == P_2$ , in case  $P_2'$  is not equal to  $P_2$  which implies Robot  $R_i$  is not a legal user and terminates the authentication process. If the condition is true then proceed further  $S$  chooses the random nonce  $G_2$  and compute

$$P_3 = G_2 \times P, \text{ and } P_4 = G_2 \times B_i' \quad (5)$$

Then,  $S$  sends the parameters  $\{P_3, P_4, \text{ and } D_i\}$  as a response message to  $R_i$ .  $R_i$  uses these parameters to compute

$$B_i = H(D_i \oplus H(PW_i) \oplus K') \text{ and } P_4' = P_3 \times B_i \quad (6)$$

Later on,  $R_i$  verifies  $P_4' = P_4$ . In case,  $P_4'$  is equal to the  $P_4$  which implies  $S$  is a legal server. So,  $R_i$  computes

$$Z_i = H((G_1 \times DK') || P_4') \quad (7)$$

and sends to the  $S$ .  $S$  receives the  $Z_i$  and computes

$$Z'_i = H((P_1 \times DK) || P_4). \quad (8)$$

Finally,  $S$  verifies the  $Z'_i = Z_i$ , if both are equal, then Robot  $R_i$  will share the session key SK with the server  $S$  that will be computed by using the formula:

$$H(ID_i || G_1 \times G_2 \times P). \quad (9)$$

## 6 Security Analysis

The adversary can join in several outlaw ways to gain susceptible information to a user. The following steps are defined as attack model to acquire information from the system used by an adversary:

- i. Messages between client and server travel through an insecure channel that can be captured by adversary via eavesdropping.
- ii. An adversary can transmit eavesdropping messages during the authentication process.
- iii. An adversary can impersonate the legitimate robot to transmit the response to the server.
- iv. An adversary can public the login credential of the legitimate robots.

Informal security analysis of the proposed scheme shows that the presented scheme is robust against many well-known attacks. The following steps describe the procedure to analyze the security of the proposed scheme:

- i. **Provides mutual authentication:** In our proposed scheme, the client robot can receive the  $DK'$ ,  $H(PW_i)$ , and  $D_i$  from the server. Then, the client robot can successfully calculate the  $B_i$  and verifies the equality  $P_4' == P_4$  hold or not. Thus, the proposed system fulfills the requirement of mutual authentication between client robots and servers.
- ii. **Provides user anonymity:** User anonymity says that an adversary should not know the identity ( $ID_i$ ) of the robot in time of transmitting the message during the login phase. In our proposed scheme, adversary also needs to know the value of  $P_1$  and  $P_2$  for successful login in the system that is very difficult to find out without knowing the  $G_1$ ,  $DK'$ , and  $P$ . Hence, anonymity is maintained in the proposed scheme.
- iii. **Resistance to a MITM attack:** Mutual authentication is achieved between the client robot and the server in the proposed scheme. Therefore, the presented scheme resists this type of attack.
- iv. **Resistance to an offline password-guessing attack:** An adversary collects the communication messages between legitimate robot and server then try to find out the password from them. In our proposed scheme, it is very difficult to guess the three secret parameters that are not possible to compute in polynomial time. Therefore, the proposed scheme resists the offline password-guessing attack.

- v. **Known session key security:** If the server successfully computes  $P_2' == P_2$ , that shows client robot is legal and client robot also computes  $P_4' == P_4$  successfully that shows server is legal. After getting both are legal, then, the session key will share between the client robot and server in our proposed scheme. Thus, our proposed scheme achieves session key security.
- vi. **No clock synchronization:** In the proposed scheme, there is no clock synchronization is used between the client robot and server. For saving the resource consumption maintained by the random nonce in time of mutual authentication.
- vii. **Resistance to an impersonation attack:** An adversary will have to generate the login credentials for access to the system. Similarly, if the adversary wants to act as a server, then it needs security parameters such as random nonce and also needs to know the secret key of the server. Hence, the proposed scheme resists the impersonation attack on both client and server.
- viii. **Provides forward secrecy:** It is required that an adversary cannot be detected before transmitting information on the bases of the current transmitted information of the robot. So, it is very difficult for an adversary to computes the secret key without knowing  $G_2$  and  $G_3$ . Therefore, the proposed scheme provides forward secrecy.
- ix. **Resistance to brute force attack:** An adversary applies the brute force attack on the system and gather information such as parameter  $P$ , but is still unable to obtain the client robot's password. Because there is no way to find random nonce  $G_1$  and  $G_2$ , and it is not possible to guess the secret key of the server. Therefore, our system opposes brute force attacks.

## 7 Conclusion

This paper has proposed an authentication scheme to explain the secure way of interaction between the robots and the cloud server to access the robotic services deployed in the server. The proposed scheme prohibits unauthorized access to cloud-based robotics applications. ECC-based authentication scheme which does not require high cost and gives guaranteed security with a small key size. ECC is a lightweight cryptography algorithm that reduces the power consumption of the robot with the minimum computation. The proposed scheme provides authentication as well as protection against security threats. The security analysis shows that the proposed scheme resists many well-known attacks.

## References

1. Russo L, Rosa S, Maggiora M, Basilio Bona B (2016) A novel cloud-based service robotics application to data center environmental monitoring. *Sensors* 16(8):1–18
2. Durrant-Whyte H, Bailey T (2006) Simultaneous localization and mapping: part I. *IEEE Robot Autom Mag* 13(2):99–110

3. Hu G, Tay WP, Wen Y (2012) Cloud robotics: architecture, challenges and applications 26(3):21–28
4. <https://www.searchsecurity.techtarget.com/definition/elliptical-curve-cryptography>
5. Lamport L (1981) Password authentication with insecure communication. *Commun ACM* 24(11):770–772
6. Juang W-S, Chen S-T, Liaw H-T (2008) Robust and efficient password-authenticated key agreement using smart cards. *IEEE Trans Industr Electron* 55(6):2551–2556
7. Sun D-Z, Huai J-P, Sun J-Z, Li J-X, Zhang J-W, Feng Z-Y (2009) Improvements of Juang’s password-authenticated key agreement scheme using smart cards. *IEEE Trans Industr Electron* 56(6):2284–2291
8. Lu Y, Li L, Peng H, Yang Y (2015) An enhanced biometric-based authentication scheme for telecare medicine information systems using elliptic curve cryptosystem. *J Med Syst* 39(3):32
9. Imamoto K, Sakurai K (2005) Design and analysis of diffie-hellman-based key exchange using one-time ID by SVO logic. *Electron Notes Theor Comput Sci* 135(1):79–94
10. Kalra S, Sood SK (2011) Elliptic curve cryptography: survey and its security applications. In: *Proceedings of the international conference on advances in computing and artificial intelligence*. ACM, vol 6, pp 102–106
11. Abi-Char PE, Mhamed A, Bachar E (2007) A fast and secure elliptic curve based authenticated key agreement protocol for low power mobile communications. In: *The 2007 international conference on next generation mobile applications, services and technologies (NGMAST 2007)*. IEEE, pp 235–240
12. Tian X, Wong DS, Zhu RW (2005) Analysis and improvement of an authenticated key exchange protocol for sensor networks. *IEEE Commun Lett* 9(11):970–972
13. Wu S-T, Chiu J-H, Chieu B-C (2005) ID-based remote authentication with smart cards on open distributed system from elliptic curve cryptography. In: *2005 IEEE international conference on electro information technology*. IEEE, 5
14. Debiao He, Jianhua C, Jin Hu (2012) An ID-based client authentication with key agreement protocol for mobile client–server environment on ECC with provable security. *Information Fusion* 13(3):223–230
15. Ray S, Biswas GP (2012) Establishment of ECC-based initial secrecy usable for IKE implementation. In: *Proceedings of the world congress on engineering*, vol 1, pp 1–6
16. Granjal J, Monteiro E, Silva JS (2013) End-to-end transport-layer security for Internet-integrated sensing applications with mutual and delegated ECC public-key authentication. In: *2013 IFIP networking conference*. IEEE, pp 1–9
17. Jiang R, Lai C, Luo J, Wang X, Wang H (2013) EAP-based group authentication and key agreement protocol for machine-type communications. *Int J Distrib Sens Netw* 9(11):304601
18. Moosavi SR, Nigussie E, Virtanen S, Isoaho J (2014) An elliptic curve-based mutual authentication scheme for RFID implant systems. *Proc Comput Sci* 32:198–206
19. Kumari S, Karupiah M, Das AK, Li X, Wu F, Kumar N (2018) A secure authentication scheme based on elliptic curve cryptography for iot and cloud servers. *J Supercomput* 74(12):6428–6453
20. Kalra S, Sood SK (2015) Secure authentication scheme for iot and cloud servers. *Pervas Mobile Comput* 24:210–223
21. Challa S, Das AK, Odelu V, Kumar N, Kumari S, Khan MK, Vasilakos AV (2018) An efficient ecc-based provably secure three-factor user authentication and key agreement protocol for wireless healthcare sensor networks. *Comput Electr Eng* 69:534–554
22. Liu C-H, Chung Y-F (2017) Secure user authentication scheme for wireless healthcare sensor networks. *Comput Electr Eng* 59:250–261
23. Sahoo A, Sahoo SS, Sahoo S, Sahoo B, Turuk AK (2020) An interoperable ECC based authentication and key agreement scheme for IoT environment. In: *2020 international conference on communication systems and networks (COMSNETS)*. IEEE, pp 419–426
24. Fang D, Qian Y, Hu RQ (2020) A flexible and efficient authentication and secure data transmission scheme for IoT applications. *IEEE Internet Things J* 7(4):3474–3484



# Intelligent Web of Things Based on Fuzzy Neural Networks



Zahraa Sabeeh Amory and Haider Kadam Hoomod

**Abstract** Web of thing is modern technology and a subset of the internet of things (IoT) which brings many application and new possibilities to improve usability and the interoperability of IoT. WoT has many challenges spatially data analytics and storage due to the increasing number of sensing devices capable of acquiring huge amounts of data. This paper presents two techniques of analytics sensors data which distributed in smart home to measure temperature and humidity of rooms: firstly, clustering analysis to sensors measurements of room's temperature and humidity in the smart home as one solution to the challenge of data analytics and storage and to determine the temperature and humidity patterns which help in making better decisions at right time, secondly, prediction model of room's temperature and humidity as a safety system in the smart home when compared the current value with predictive value depends on historical data to detect the deviation of the measured data from the sensors, then the output of two techniques are shown in designed web pages to provide better services for the citizens living in the home. Those techniques are implemented by using intelligence neural networks and fuzzy logic, fuzzy adaptive resonance theory neural network (Fuzzy ART—NN) for clustering model and long-short term memory recurrent neural network (LSTM—RNN) for the prediction model. The performance of the clustering model evaluated by training time and the accuracy when the results are shown the clustering approach has short training time and accuracy reached 99%. While the performance of prediction approach evaluated using root mean square error (RMSE) and training time, and the results are shown RMSE reached 0.02 and training time approximation of 4.76 s for 4550 samples.

**Keyword** Room temperature and humidity · Web of things · Clustering analysis · Prediction

---

Z. S. Amory (✉) · H. K. Hoomod

Department of Computer Science, College of Education, Mustansiriyah University, Baghdad, Iraq  
e-mail: [efficient2016@gmail.com](mailto:efficient2016@gmail.com)

H. K. Hoomod

e-mail: [drhjnew@gmail.com](mailto:drhjnew@gmail.com)

## 1 Introduction

The idea of the Web as an application-layer for the IoT scenario began to emerge in 2007 [1]. While the internet of things (IoT) is a global network infrastructure, that enables the communication between physical objects to the network layer, the Web of things (WoT) bring interoperability at the application layer which a top layer of IoT. WoT considered numerous platforms and frameworks that integrate heterogeneous of various embedded physical resources (coordinators, sensors, actuators) and real-world services at the Web platforms by using internet network [2]. WoT brings many possibilities to IoT applications that improve human comfort and quality of life. WoT based on application layer for monitoring services, periodic sensor reporting, notification controlling, and triggering action events in the smart system. The application loads dynamic information from devices online mode or offline mode to a friendly user interface in anytime and anywhere. The proposed system is based on the lambda architecture which proposed by Nathan Marz [3] lambda consists of three distinct layers to process data (speed layer, batch layer, service layer) [4], the system involving analytics the environment that help in taking effective decisions and more accurate prediction at right time to provide better services to the citizens where is necessary to use both entire datasets from sources: historical data stored earlier in the batch layer as well as real-time data in speed layer [5]. So, the architecture of the system involved collecting sensor reading and route it to edge computing then processed two times: in speed layer once to cluster the arrived records of temperature with humidity (real-time data analytics) and to raise alarms where the measured value out of ranges to take better decisions at right time, second in the batch layer which considers the master database to build prediction model with time-series algorithm LSTM—RNN on historical data stored per time interval and to perform deeper analytics [6], this layer aggregates the raw data in csv format and trains model. The result of the batch layer is then transferred to a service layer for the application to the surface. While the speed layer allows the web application to access the most current data to ad hoc cater the pending queries [7], these layers are shown in Fig. 1.

Room temperature and humidity monitoring system in real-time for a smart home is necessary for a safety system (fire disasters) and helps to improve managing the property [8]. Monitoring environment system for room temperature and humidity helps to detect heating and ventilation equipment failures at the early stage and reduce costs of operation [9]. Smart home can use a set of sensors to automatically room heating, lighting to reduce the consumed energy by using monitoring of environment system that integrating smart grid web services [10]. Frequently IoT devices generate huge amounts of streaming data per time interval. Streaming data analysis in real-time seemed to be taking a lot of attention to what mattered in many applications and systems [11]. The streaming data clustering analytics proposed with fuzzy ART neural network for temperature and humidity sensors (DHT-22) reading is useful to determine the temperature and humidity pattern (cluster number) which help in making better decisions at right time when the measured value out of ranges and generate the alert through monitoring system based on WoT in real time. And, the

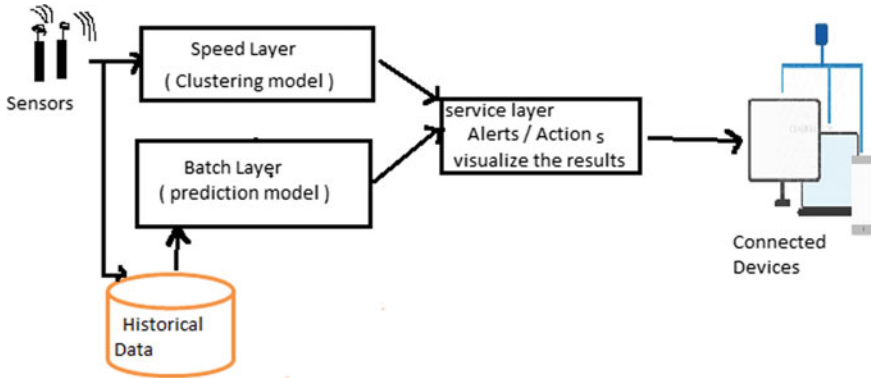


Fig. 1 Modeling sensor data architecture

proposed model of clustering analytics to sensor (DHT-22) data with fuzzy ART is useful to reduce raw data by sending the cluster number to the database server. The clustering analytics approach as an unsupervised learning technique implemented in edge computing (management station)-based WoT-emerging sensor network (ESN) architecture [12].

Predicting potential sensor values over time from previous values so that the next value will be can be known and detecting extreme (abnormal) values by comparing current values with projected values in normal cases and using warning systems such as e-mail or SMS warning to avoid or detect unexpected problems in smart buildings [13] and detect internet intrusion of items intrusion detection of the internet of things (IoT) as a security system in network information [14]. LSTM recurrent neural network is used to construct the predictive technic of sensor readings such as humidity and temperature that is a non-linear process for dynamic data over time and analyzing time-series data. Temperature forecasting for position are produced by artificial neural networks and improved by Jain et al. [15, 16].

## 2 Methodology

### 2.1 Fuzzy ART Neural Network to Clustering Model

Fuzzy ART network is an unsupervised learning technique depends on fuzzy set theory into ART neural network to cluster sensors data into many categories in changing environment, which take input a set of rows and gives the output a set of clusters. Clustering is an important tool in knowledge discovery and data science by grouping the same data together and to analysis temperature and humidity pattern. This method will lead to summarize the data amount into a small number of clusters and easily manage and arrange data. Fuzzy ART method has been used because of

its properties, which are fast, self-organizing [17], small memory requirements, and stability [18], stability refers the ability to create a new category when an input vector does not match prototype vector but the input does not change prototype pattern. Fuzzy ART includes three layers: input layer  $F0$ (complement coding),  $F1$ (comparison layer), output layer  $F2$ (competition layer) without hidden layer and with many neurons in each layer. For cluster input vector ( $P$ ) into pattern  $j$ , fuzzy ART network depends on two criteria, choice and match function. The set of weight values is  $W = \{W_{ji}; i = 1, 2, \dots, 2M; j = 1, 2, \dots, N\}$ , where  $P \equiv (P_1, P_2, \dots, P_N)$ , input vector which represent the real values in  $[0,1]$  interval is normalized by using complement code formula (1) on  $F0$  to prevent cluster proliferation and became  $I$  of  $2N$  elements(two dimensional vector):

Where

$$I \equiv (P, P^c) = (P_1, P_2, \dots, P_N; P_1^c, P_2^c, \dots, P_N^c),$$

$$P_i^c = 1 - P_i \tag{1}$$

where  $P_i$  is the input vector.

To calculate the choice (matching) function for each node using the Eq. (2):

$$T_j = \frac{|I \wedge W_{ji}|}{\alpha + |W_{ji}|} \tag{2}$$

where  $\alpha$  is a choice parameter and  $\alpha > 0$ .

The fuzzy AND operator  $\wedge$  is defined by Eq. (3):

$$I \wedge W_{ji} = \min(I_i, W_{ji}, \min(I_2, W_{j2}) \dots \min(I_{2N}, W_{j2N})) \tag{3}$$

To find similarities between the input vector  $I$  and the weight vector of the winning neurons ( $W_{ji}$ ) by using the following test:  $\frac{|I \wedge W_{ji}|}{I} \geq \rho$  where  $\rho$  is vigilance parameter and value of  $\rho$  between zero and one, the comparison carried on  $F1$  layer.

If the condition is met, then  $j$  is selected and allowed to update the weight vector, otherwise, neurons  $j$  is deactivated for existing input  $I$  and  $T_j$  is set equal to  $-1$ .

The weight vector is updated according to Eq. (4)

$$Z_j^{\text{new}} = \beta(I \wedge Z_j^{\text{old}}) + (1 - \beta)Z_j^{\text{old}} \tag{4}$$

where  $\beta$  represents learning rate and  $\beta = [0,1]$  to fast learning rate when  $\beta$  set equal to 1.

Figure 2 Summarizes the fuzzy ART algorithm

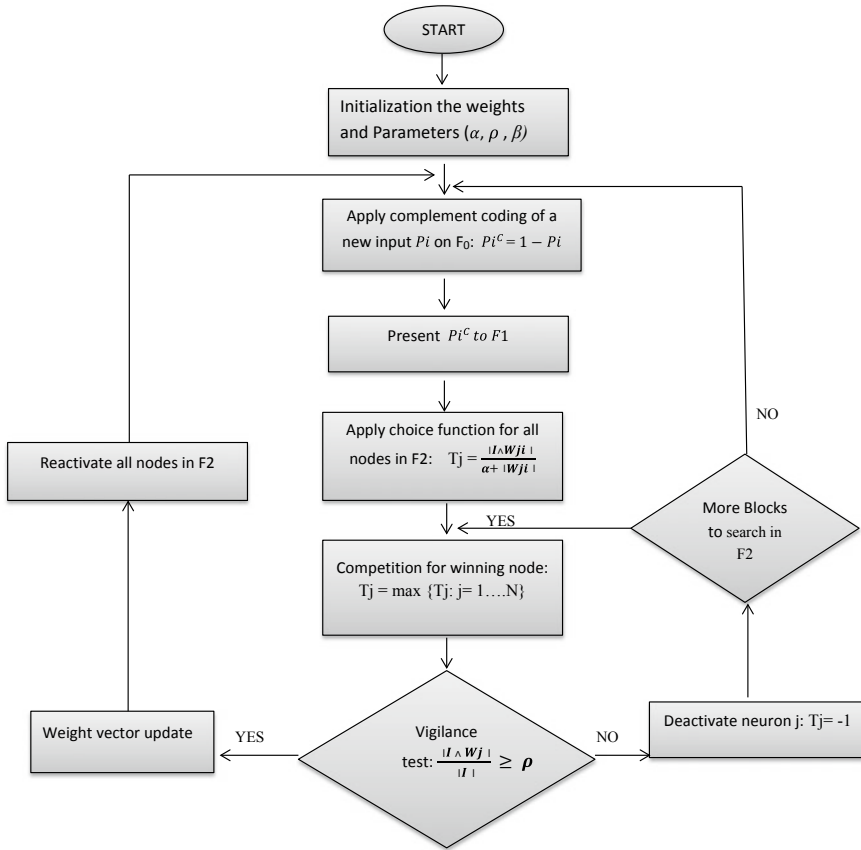


Fig. 2 Flowchart for fuzzy ART neural network algorithm

## 2.2 Long Short Term Memory (LSTM) Neural Network for Prediction

LSTM is a particular type of recurrent neural networks used for sequence data processing and suitable for predictive approach because it has a sequential memory technique designed to capture sequence or time-series data such as a sensor reading [14, 19], handwriting recognition [20], etc. Hochreiter, Schmidhuber [21], proposed LSTM. LSTM capable of remembering long term dependency to prevent gradient problems from vanishing and exploding with using explicit gating mechanism to save sequential data [20]. The concept of using explicit gating architectures is to have the major channel that flows over time and getting gate-controlled and linked modules that decide how much the node will put in the channel. Originally, the cell gates are logistic sigmoid units with a point-wise multiplication that constitute smooth curves in the range 0–1, while one represents that all values passes

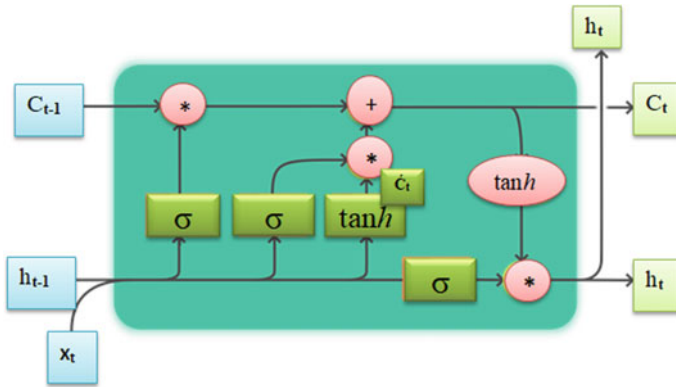


Fig. 3 LSTM memory block architecture

through the gate, but zero represents that nothing passes through the gate. These modules work to delete or add channel content, and the gates decide what to save or discard. LSTM has memory cells that contain (three internal cell gates to provide the details in a sense wide-ranging, these gates are: (1) Input gate controls whether the memory cell is change; (2) Forget gate controls if the memory cell is reset to zero; (3) Output gate decides whether the information of the current cell is made visible. Each gate depends on the previous cell hidden state ( $h_{t-1}$ ) and on the existing input ( $x$ ) as a vector concatenation input applies a sigmoid  $\hat{C}$  range that reflects new values for candidates which can be added to the cell hidden state. Let  $W_f, W_i, W_o$  are independent three-gates weight matrices and  $b_i, b_f, b_o$  represent bias vector for each gate. While  $W_c, b_c$  corresponding the weight vector, the bias to adapt cell state, that is illustrated in Fig. 3.

The input gate is calculated  $I_t$  using Eq. (5)

$$I_t = \sigma (W_i * [h_{t-1}, x_t] + b_i) \tag{5}$$

The forget gate is  $F_t$  which has been obtained as follows:

$$F_t = \sigma (W_f * [h_{t-1}, x_t] + b_f) \tag{6}$$

The Eq. (7) is used for calculating  $O_t$  by output gat

$$O_t = \sigma (W_o * [h_{t-1}, x_t] + b_o) \tag{7}$$

The new hidden state obtained according to formula (8)

$$h_t = O_t * \tanh(C_t) \tag{8}$$

To calculate and update the state of the cell as follows:

$$C_t = (F_t * C_{t-1}) + (I_t * \dot{C}_t) \tag{9}$$

where

$$\dot{C}_t = \tanh (W_\epsilon * [h_{t-1}, x_t] + b_\epsilon) \tag{10}$$

$\sigma$  and  $\tanh$  are used as activation function, where  $\sigma$  is a sigmoid activation which generating numbers between zero and one and defined by formula (11) while  $\tanh$  is hyperbolic tangent function obtained according to formula (12) and includes a set of real values ranging from 1 to  $-1$ .

$$\sigma(x) = \frac{e^x}{1 + e^x} \tag{11}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{12}$$

### 3 Materials

#### 3.1 Network Architecture System

The architecture network of the proposed system is shown in Fig. 4.



**Fig. 4** The network architecture of the proposed system

### 3.2 Data Set Description

The real measured data used in the peoposed system, obtained from a DHT-22 sensor which produces digital values for temperature and humidity in real time. The sensor has been placed in the living room and associated with the raspberry pi system (microcontroller) while raspberry pi connected to the computer through theVNC-viewer application by using IP address, then data collected while the room in regular use and the measured data stored in csv (comma separated values) file which allows data to be stored and retrieved. To boost data quality, the preprocessing stage is used to delete data redundancy and the missing values are filled with the last reported humidity and temperature value. After that normalization phase, it was appropriate to rescue the data in the range of zero to one using the MinMaxScaler, and then the data set divided to two sets the train set and test set with a partition ratio is 0.5 (50% train set and 50%test set).

### 3.3 Designing Fuzzy ART Model for Clustering

The values of temperature and humidity obtained from the sensor and reported as csv file which used to training sets and testing set. To represent the nine labels of temperature AND humidity, three sets for temperature and three sets for humidity (high, medium, low) are determined by using the fuzzy set. As shown in Tables 1, 2 and 3.

**Table 1** The distribution of category index to temperature values

Category index	1 (low)	2 (medium)	3 (high)
Distribution	1–18	19–23	24–30

**Table 2** The distribution of category index to humidity values

Category index	1 (low)	2 (medium)	3 (high)
Distribution	1–20	21–30	31–40

**Table 3** The distribution of category index to temperature values AND humidity values

Category index	1	2	3	4	5	6	7	8	9
Distribution	1 AND 1	1 AND 2	1 AND 3	2 AND 1	2 AND 2	2 AND 3	3 AND 1	3 AND 2	3 AND 3



### 3.4 Designing LSTM Model for Prediction

The data set entered into the model is a time series (sensor readings over time) with window method, so we set 7 for window size and customize the input layer size is one with the sequential model, then add to the model one hidden layer with one dense layer to produce a single output. By using mean squared error (MSE) as a loss function with adaptive moment estimation (Adam) as an optimizer to reduce gradient descent, sigmoid activation function is used in the model. The number of iterations (epoch), number of blocks, and the batch size are determined for the LSTM network in experiments step. Implementation of the LSTM neural network in python language (v.3.7) after importation Keras package (library for Theano and Tensorflow in deep learning framework).

- Evaluation Criteria of the LSTM Model

According to Eq. (13), the root mean squared error (RMSE) is calculated to assess the error metric for numerical prediction, which is an approximation of how badly the network performs at the last propagation that measures the gradients for each node in the network, the gradient aparameter usage to alter the inner weights of the networks allowing the network to know the greater gradient of the bigger changes and vice versa with determining the Epoch. The model of prediction is efficient when RMSE is small.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (13)$$

(Where  $y_i$  is real observed value,  $\hat{y}_i$  is predicted value at a time  $i$ , and  $N$  is the sum number of times for the evaluation).

## 4 Results and Discussion

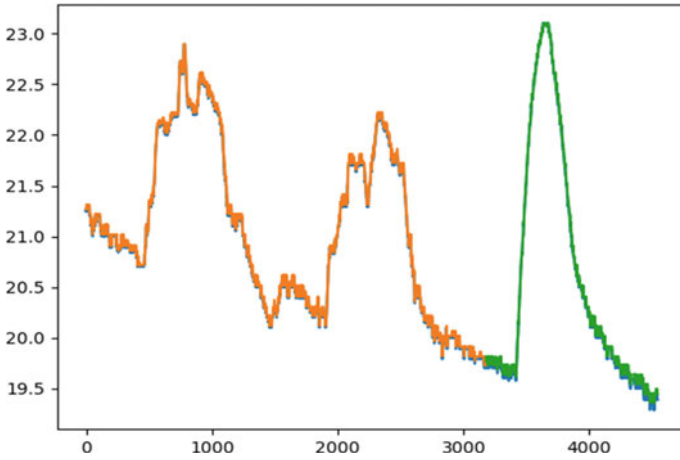
### 4.1 Prediction Model

Prediction model performance depends on execution time and the root mean square error (RMSE) with the mean square error (MSE) that represent a loss function. These measurements were utilized for evaluation the performance of the prediction model.

Initially, we used actual temperature values for four days reported with (4550) test samples; the data set is then divided into (70%) training set and (30%) expected test set, the iteration number (Epoch) set to 100 at first then increase its value by a hundred each time; the number of hidden layers is set to one and batch size set to one too. Table 4 compares the quality results RMSE, loss, training time when increasing the number of iteration at the same data set while Fig. 5 explains the output of the

**Table 4** Quality comparison for the loss, training time, RMSE when increasing epoch number

Data set	Epoch no	Loss	Train time in sec	Training RMSE	Testing RMSE
4550 samples	100	8.8653e-05	493.27	0.04	0.04
4550 samples	200	8.4329e-05	892.13	0.04	0.05
4550 samples	300	8.2607e-05	1922.05	0.04	0.04

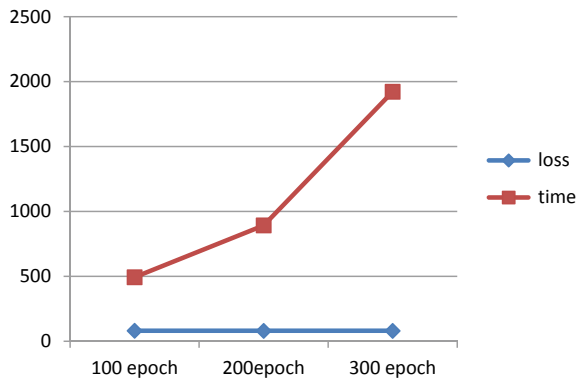


**Fig. 5** prediction of temperature values for four days

LSTM neural network to predict the temperature values for four days with iteration number equal to 200.

The blue color is the original data set, the orange color for the train data set and the green color for the prediction test set, and Fig. 6 explains the number of iterations (Epoch) that impact the training period and the loss (MSE) when adjusting.

**Fig.6** The connection between time and loss of training with a changing epoch



Through Table 4 and Fig. 6, the results show an increase in training time in second when increasing epoch values with a very slight decrease in the value of the loss.

Then, the number of block is adjusted for the LSTM model with using humidity values for one day as dataset and measured the RMSE of training and testing sets and training time; Table 5 presented the results, and Fig. 7 shows the output neural network to prediction humidity for one day with Epoche qual to 200.

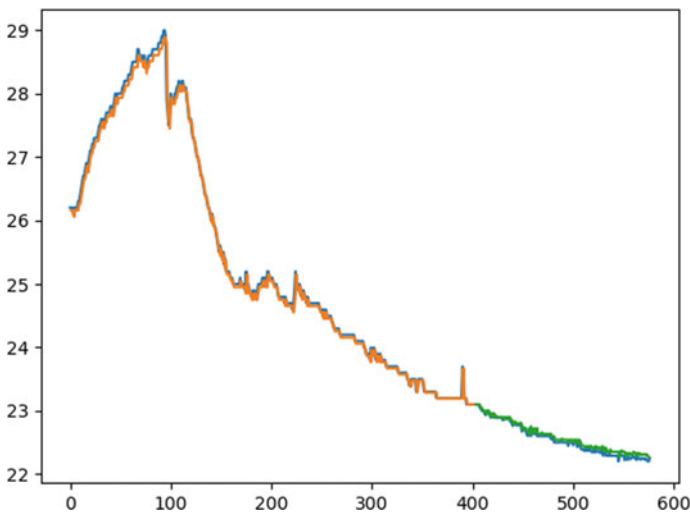
Table 5 compares training time, loss, and RMSE of a dataset-based LSTM prediction model according to a different number of neurons, so increasing in training time is noted when increasing the number of neurons and decrease in values of RMSE for the test set.

Figure 8 explains how the numbers of blocks to effected on the train time of the LSTM network and on the RMSE of the training set and the RMSE of the testing set.

Then, the data entered to the model represents the temperature measurements for four days with 4550 test samples and 70 percent splitting ratio (0.70 training set, 0.30 testing set) and the same prediction model applied for it with changing the batch-size, Table 6 shows the results for RMSE, train time, loss function, where Fig. 9

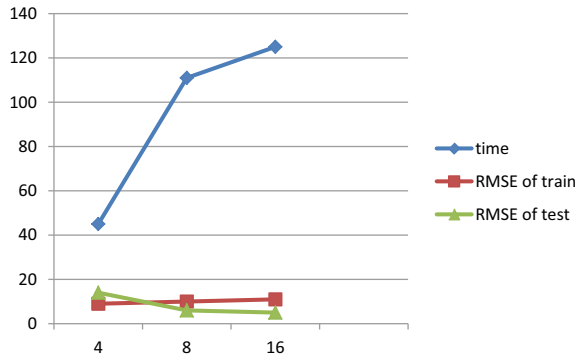
**Table 5** Comparison of RMSE results, training time, and loss when changing block number

Data set	Epoch no.	No. of block	Loss	Training time in sec	Training RMSE	Testing RMSE
580 samples	200	4	2.3707e-04	45.297	0.10	0.14
580 samples	200	8	2.3820e-04	111.087	0.10	0.06
580 samples	200	16	2.4664e-04	125.943	0.11	0.06



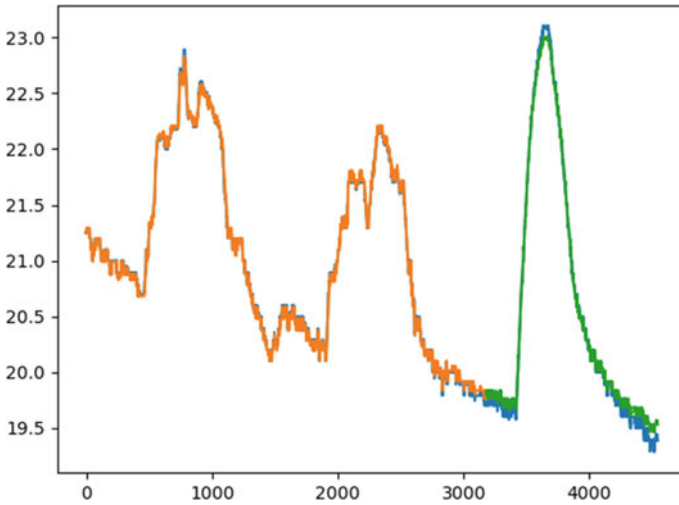
**Fig. 7** Prediction of relative humidity for one day

**Fig. 8** The relationship between training time and RMSE with changing number of blocks



**Table 6** Comparison of results to the loss, train time, RMSE with changing value of the batch-size

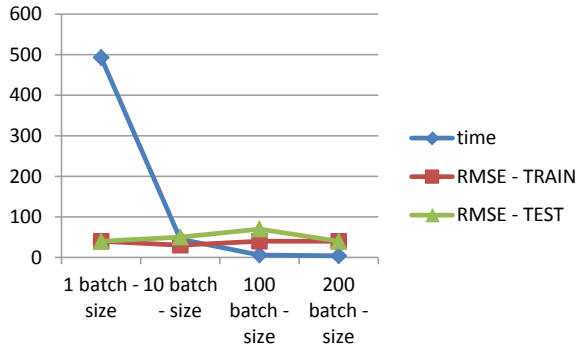
Data set	Batch-size	Loss	Train time in sec	Training RMSE	Testing RMSE
4550 samples	1	8.8653e-05	493.67	0.04	0.04
4550 samples	10	8.1412e-05	45.67	0.03	0.05
4550 samples	100	8.8845e-05	6.26	0.04	0.07
4550 samples	200	1.0189e-04	4.76	0.04	0.09



**Fig. 9** Temperature prediction values for four days with batch-size (100), splitting ratio 70%

presented the output of the LSTM network to temperature prediction for four days with batch-size equal to 100 and Epoch equal to 100 too.

**Fig. 10** The relationship between training time and RMSE with changing the value of batch-size



**Table 7** Comparison the performance of loss, train time, RMSE when changing value of batch-size and splitting ratio 50%

Data set	Batch-size	Loss values	Training time in sec	Training RMSE	Testing RMSE
4550 samples	1	8.5752e-05	302.08	0.03	0.07
4550 samples	10	7.6706e-05	34.66	0.03	0.07
4550 samples	100	7.9951e-05	5.14	0.03	0.10
4550 samples	200	3.7494e-04	3.54	0.07	0.19

Table 6 compares training time, loss, and RMSE of dataset-based LSTM prediction model according to different batch number-size, so a decrease in training time is noted when batch number-size increases and RMSE values for the test set increase.

Figure 10 shows how changing the value of batch-size that effects on training time and the RMSE of the train set and the test set.

The data set of temperature values is then entered to the model for four days( 4550 test samples and splitting ratio 50% for training set and testing set) and the LSTM network applied for it with changing the value of the batch-size, the experiences results are illustraten in Table 7. For RMSE, train time, loss function, Fig. 11 presented the outputs of the neural network to temperature prediction for four days at Epoch equal to100 and the splitting ratio 50%.

Through Table 7 shows the results of loss, training time, and RMSE of the model significantly at the same dataset at different batch-size values, so we can note a clear decrease in training time and in the loss value while an increase in testing RMSE.

### 4.2 Clustering Model

By using the dataset (train set, test set) which include nine categories of temperature and humidity of sensor readings in different cases and time with two attribute and

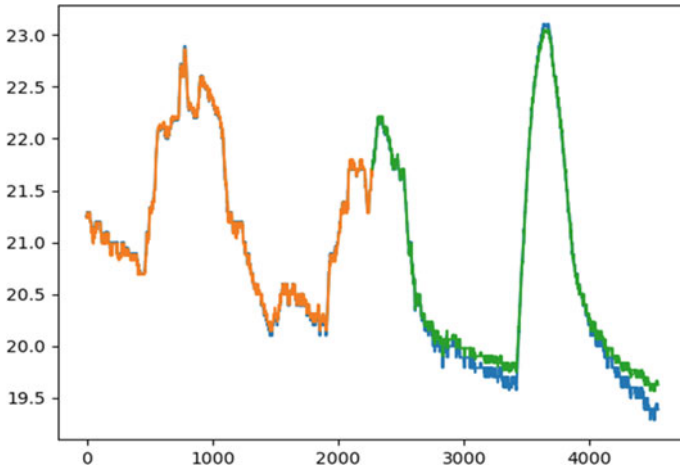


Fig. 11 Predictions of temperature values for four days with batch-size (100), splitting ratio 50%

nine rows. After implementation, the clustering model on training and testing set with determining the parameters are learning rate (beta), vigilance threshold (rho), and choice parameter (alpha), respectively, the values (beta = 1, rho = 1, alpha = 0.00001), Fig. 12 shows the results of clustering training and testing set after carried out the fuzzy ART model, and the accuracy reached 100% and the execution time (CPU time) = 0.0272 s, novel list represent the number of categories that are not found in the memory of model, in this state novel list = 9.

And this model can determine the max, min value in the single-column whether in training and testing set as shown below.

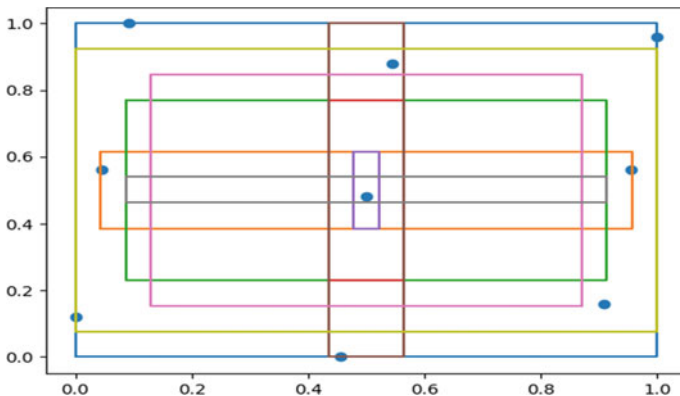


Fig. 12 Clustering training and the testing dataset at rho = 1

```
min = 0.08, max = 0.31
min = 0.15, max = 0.41
min = 0.05, max = 0.27
min = 0.19, max = 0.44
*****
```

when testing the same dataset (training, testing) on the clustering model with changing parameter vigilance threshold ( $\rho$ ) to 0.8, the results are shown that the novel list is empty and the execution Time = 0.0043 s.

when  $\rho = 0.5$ , the novel list = 2, the execution time = 0.0021 s, the output shows in Fig. 13.

when  $\rho = 0.2$ , the novel list = 0, the execution time = 0.0027 s, the output shows in Fig. 14.

Through the above three figures, the effect of the  $\rho$  coefficient on model accuracy and learning time can be seen where the best learning value of  $\rho = 1$ .

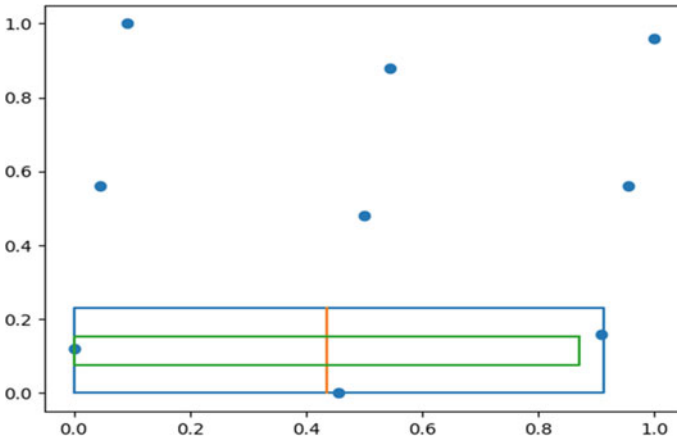


Fig. 13 Clustering training and testing dataset at  $\rho = 0.5$

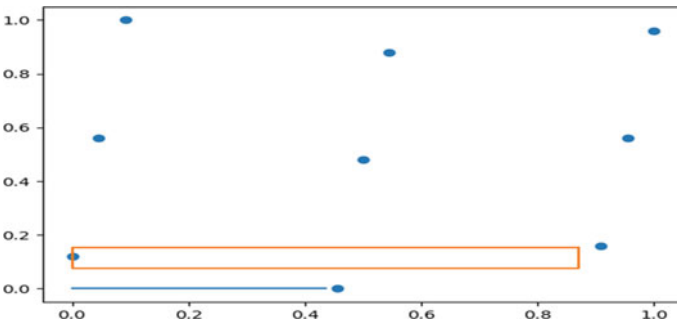
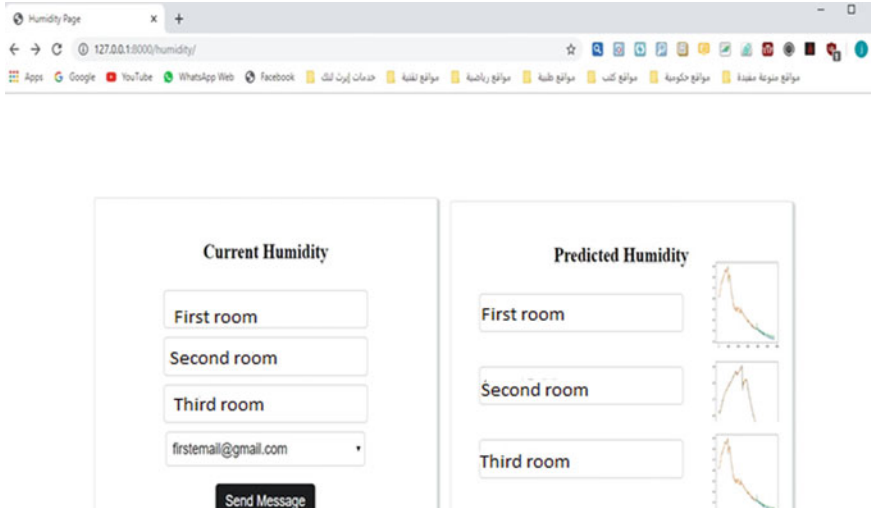


Fig. 14 Clustering training and the testing dataset at  $\rho = 0.2$

**Table 8** Training time and novel list for the first data set when changing the value of parameter rho

Data set	Rho	Training time (s)	Novel list no.
Dataset 1	1	0.0272	9
Dataset 1	0.8	0.0043	0
Dataset 1	0.5	0.0021	2
Dataset 1	0.2	0.0018	0
Dataset 1	1	0.0064	9



**Fig. 15** Humidity page (127.0.0.1:8000/humidity/)

When  $\rho = 1$ , the novel list = 9, the time = 0.0064 s. Where a decrease in the execution time is noted when testing the model in  $\rho = 1$  that difference on the execution time when training the model in  $\rho = 1$ ; Table 8 shows the performance comparison when adjusting the parameter value ( $\rho$ ), and Fig. 15 shows the page of the designed web to monitoring the humidity values.

## 5 Conclusion

WoT is a very effective technology that enables developers to innovate and evolve freely, reducing the need to wait for new components in networks, build new infrastructure, or reinvent how our applications are built. WoT can be monitoring the temperature and humidity of rooms in a smart home with dilution the risk of severe damage and high costs due to predictions and early detection of incidents. The experienced results of the prediction model are illustrated the recurrent neural network



based on LSTM can predict a fixed length of values from the historical data sequence with the same length and presented how the neurons number, batch-size, and number of iterations are effect on training time of LSTM network and on the RMSE of training set and of testing set. The LSTM network for prediction is better accurate where the value of RMSE is smaller, particularly, when using 4550 test samples the training RMSE reached 0.04. The efficient performance is found through the results obtained, and the predicted values are similar to the real measured data with a very low error rate and short training time. While the results of the clustering model are shown, the execution time of clustering model is very small approximately depending on the size of data set but the reduction in the required space is very large, this indicates of effectiveness of the clustering model. The accuracy of the clustering model mainly depends on the parameter vigilance threshold ( $\rho$ ), when the  $\rho$  closer to value 1, the accuracy becomes better and up to 100%.

## Reference

1. Erik W (2007) Putting things to rest. UCB I School Report, 15 Nov 2007
2. Andreas K, Muhammad IA (2016) do web of thing platform truly follow the web of things . conference: iee world forum on internet of things (WF-IoT), At Reston, VA, USA, 978-1-5090-4130-5
3. Nida SK, Sayeed G, Sajjad H (2018) Real-time analysis of a sensor's data for automated decision making in an IoT-based smart home. *Sens J* 18(6)
4. Sangmin P, Soung HP, Lee WP, Sanguk P, Sanghoon L, Tacklim L, Sang HL, Hyeonwoo J, Seung MK, Hangbae C, Sehyun P (2018) Design and implementation of a smart iot based building and town disaster management system in smart city infrastructure. *Appl Sci* 8:2239. <https://doi.org/10.3390/app8112239>
5. Dan H, Cătălin L, Elena A, Valentin C (2018) BigClue: towards a generic IoT cross-domain data processing platform. In: IEEE 14th international conference on intelligent computer communication and processing (ICCP)
6. Gautam P, Gangmin L, Katie A (2018) Multi-agent big-data lambda architecture model for e-commerce analytics. In: IEEE second international conference on data stream mining processing (DSMP), pp 21–25
7. Murphy MKP, Inder M, Jon DS, Singh B (2015) Lambda architecture for cost-effective batch and speed big data processing. In: IEEE international conference on big data
8. Faruque MH (2019) Chapter seven—best management practices , building, energy, roads, bridges, water and sewer systems. Elsevier:419–431
9. Jyoti P, Sonavane SS (2016) Efficient energy management in smart grids based on raspberry pi & web of things. *Int J Sci Technol Eng IJSTE* 2(12)
10. Murad K, Bhagya NS, Kijun H (2017) A web of things-based emerging sensor network architecture for smart control systems. *Sensors J* 17:332
11. Marz N, Warren J (2013) Big data; principles and best practices of scalable real time data systems. Manning Publications
12. Yuvraj K (2020) Lambda architecture—real time data processing. *Journal* 27:19
13. Moehammad A, Prastoyo U, Winarno BH (2019) Server room temperature and humidity monitoring based on internet of thing (IoT). *IOP Conf Ser J Phys Conf Ser* 1306:012030
14. Xuefei L, Chao Z, Pingzeng L, Maoling Y, Baojia W, Jianyong Z, Russell H (2018) Application of temperature prediction based on neural network in intrusion detection of IoT. *Secur Commun Netw* 1635081:10

15. Jain A, McClendon RW, Hoogenboom G, Ramyaa R (2003) Prediction of frost for fruit protection using artificial neural networks. In: American Society of Agricultural Engineers, St. Joseph, MI, ASAE Paper 03-3075
16. Jain A (2003) Frost prediction using artificial neural networks; a temperature prediction approach. M.S. thesis, Artificial Intelligence Center, University of Georgia, Athens, GA
17. Xiao JW, Licheng L, Fengxiang X (2018) Fault diagnosis of rolling bearing based on optimized soft competitive learning fuzzy ART and similarity evaluation technique. *Journal* 38:91–100
18. Cheng CC, Ssu HC (2015) A comparative analysis on artificial neural network-based two-stage clustering. *J Cogent Eng* 2(1)
19. Luo DY, Jin H, Yang DD (2017) LSTM-based temperature prediction for hot-axles of locomotives. *ITM Web Conf* 12:01013
20. George ED, Dong Y, Li D, Alex A (2012) Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *Audio Speech Lang Process IEEE Trans*:30–42
21. Alex G, Marcus L, Horst B, Jürgen S, Santiago FA (2008) Unconstrained on-line handwriting recognition with recurrent neural networks. *Adv Neural Inf Process Syst*:577–584

# Author Index

## A

Achary, Rathnakar, 325  
Adurkar, Rajasi, 483  
Agarwal, Anjali, 99  
Agarwal, Priyanka, 819  
Aggarwal, Vaibhav, 689  
Ahmed, Shah Tuhin, 313  
Aiswarya, S., 663  
Akila, M., 151  
Amara, Indraneel, 725  
Amory, Zahraa Sabeeh, 871  
Anmol, 143  
Anuradha, Sathigari, 523  
Ashturkar, Harsha H., 619  
Awasthi, Saatvik, 41  
Azam, Kazi Sultana Farhana, 313

## B

Bendigeri, Kirankumar Y., 673  
Berberidis, Christos, 1  
Bhalchandra, A. S., 619  
Bharadwaj, Gunjan, 143  
Bharathi, D., 831  
Bharathi, S. H., 119  
Bhatkar, Srushti, 471  
Bhowmick, Kiran, 483  
Bhuvaneshwari, M., 737  
Binu, A., 789

## C

Chacko, Anna Mariam, 211  
Chandran, Geethu, 457  
Chandrashekhar, A. M., 15

Charith, Srihari, 15  
Chauhan, Ajeet, 143  
Chethana, S., 15  
Chitra, E., 291  
Chormunge, Smita, 199  
Chourasia, Mayank, 471

## D

Dandagi, Vidya S., 435  
Daniel, A. K., 553  
Dash, Dattatreya, 233  
David, Merin, 55  
Desai, Shreya, 483  
Dinesh Acharya, U., 499  
Dobhal, Dinesh C., 65  
Doraipandian, Manivannan, 395  
Doriya, Rajesh, 861  
Dushyantha, N. D., 345

## G

Ganesan, T., 523  
Garg, Pardeep, 377  
Gautam, K. S., 151  
Ghosh, Mainak, 335  
Goel, Amit Kumar, 603  
Gokul Krishna, R., 831  
Goswami, Mausumi, 267  
Grace Mary Kanaga, E., 737  
Gupta, Akshat, 377  
Gupta, Jagrati, 447  
Gupta, Kirtika, 335  
Gupta, Srishti, 631

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

J. Hemanth et al. (eds.), *Intelligent Data Communication Technologies and Internet of Things*, Lecture Notes on Data Engineering and Communications Technologies 57, <https://doi.org/10.1007/978-981-15-9509-7>

**H**

Haral, Shriya, 471  
 Harika, Attada, 523  
 Hole, Varsha, 411  
 Hoomod, Haider Kadam, 871

**J**

Jacob, Joel Eliza, 773  
 Jain, Amit, 447  
 Jain, Anurag, 535  
 Jain, Nitin Kumar, 571  
 Jain, Priti Rai, 185  
 Jain, Sarika, 689, 697  
 Jain, Saurabh, 861  
 Jambhale, Tejas, 277  
 Janisha, R. S., 109  
 Janoria, Honey, 643  
 Jashma Suresh, P. P., 499

**K**

Kadakia, Swapnil, 483  
 Kaliappan, Vishnu Kumar, 151  
 Kalluri, Amrutha, 99  
 Kamat, Yogesh, 411  
 Kanade, Vijay A., 513  
 Kanhe, Aniruddha, 89  
 Karthick, S., 761  
 Kasana, Geeta, 301  
 Keerthana, V., 847  
 Khan, Nafis Mahmud, 169  
 Khattar, Anuradha, 245  
 Kohli, Narendra, 749  
 Koleshwar, Ankita S., 89  
 Koukaras, Paraskevas, 1  
 Krishna Varshitha, K., 653  
 Kruthika, G., 345  
 Kudale, Omkar, 631  
 Kuddus, Khushboo, 169  
 Kulkarni, Smita, 471  
 Kumar, Chandan, 689  
 Kumar, Puneet, 143  
 Kumar, Rajneesh, 535  
 Kumar, Tapas, 603  
 Kumbalavati, Santosh B., 673  
 Kuruba, Padmaja, 345

**M**

Madhukar, B. N., 119  
 Madhvesh, Bommanapalli Vijaya, 211  
 Mallapur, Jayashree D., 673  
 Mamatha, H. R., 725

Meenaatchi, S. M., 583  
 Mehrotra, Monica, 223  
 Mehta, Rachana, 199  
 Mehta, Rajesh, 301  
 Menon, Sreedevi K., 663  
 Minj, Jasmine, 643  
 Mishra, Awanish Kumar, 749  
 Mishra, Zishani, 447  
 Mukul, Anand, 831  
 Murugesan, G., 135

**N**

Naik, Manthan S., 325  
 Nalajala, Sunanda, 523  
 Naveen, J., 15  
 Neelima, N., 99  
 Nikitha, Neelisetty, 523  
 Niranjana, D. K., 361  
 Nishanth, 233

**P**

Pancholi, Tirth K., 325  
 Pant, Himanshu, 65  
 Pant, Pawan Kumar, 65  
 Parimala, P., 847  
 Patre, Pooja, 643  
 Pavani, G., 653  
 Pavithra, B. S., 77  
 Phansekar, Shubham, 411  
 Ponmaniraj, S., 603  
 Poornima, A. S., 211  
 Praharsa, 233  
 Pranav, Bhuvanapalli Aditya, 211  
 Prasanna Bharathi, P., 653  
 Prashanth, T., 447  
 Pratap, Ashish, 819  
 Preethi, G., 135  
 Priyanka, J., 29

**Q**

Quadri, S. M. K., 185, 245

**R**

Radhakrishna Rao, K. A., 77  
 Radhesyam, Vaddi, 653  
 Ragul, P., 831  
 Rai, Ashok Kumar, 553  
 Raj, Manish, 261  
 Rajendran, Sujarani, 395  
 Rajeswari, K., 583  
 Rakesh, N., 233, 761

Ramakrishnan, M., 29  
 Ramar, K., 457  
 Rao, Aditi, 631  
 Reddy, Satthi, 233  
 Riya, Farhin Farhad, 313

## S

Sahila, K. M., 805  
 Sai Pranav, K., 725  
 Saini, Shubham, 143  
 Saji, Anvy Elsa, 457  
 Sanjay, N., 447  
 Sarada, V., 291  
 Sarath Chandran, P., 789  
 Sardar, Prasanjit, 697  
 Saritha, S., 773  
 Sekar, K. R., 709  
 Senthil Anand, N., 709  
 Sethi, Rachna, 223  
 Shah, Priya, 483  
 Shanmugam, Raju, 41  
 Shanthini, J., 761  
 Sharma, Archit, 377  
 Sharma, Minakshi, 535  
 Sharma, Raushan Kumar, 689  
 Sheeba, O., 109  
 Sherekar, S. S., 89  
 Shitole, Ajitkumar, 631  
 Shukla, M., 819  
 Sidnal, Nandini, 435  
 Singh, Akhilesh Kumar, 261  
 Singh, Devendra, 65  
 Singh, Vernika, 41  
 Sirisha Devi, J., 593  
 Sminesh, C. N., 421  
 Srinath, Sowmini, 847

Subba Reddy, N. V., 499  
 Subbulakshmi, S., 457  
 Sudha, M., 277  
 Susan, Seba, 335  
 Swaraj, K. P., 55  
 Syed, Fauzan, 847

## T

Tenginakai, Pruthvi, 847  
 Thakare, V. M., 89  
 Thakur, Shabnam, 301  
 Thomas, Bibu, 805  
 Tiwari, Ekansh, 697  
 Tiwari, Prakhar, 377  
 Tjortjis, Christos, 1

## U

Upadhyay, Hemant, 411

## V

Varsha, Viswanathan, 421  
 Venkataraman, V., 709  
 Verma, Ajay, 571  
 Vijaya Bhaskar Reddy, P., 593  
 Vijayakumar, V. A., 761  
 Vishnu, D., 109  
 Vivek Sidhaarthan, S., 831

## Y

Yadav, Nikhil, 631  
 Yamini, S., 135  
 Yogapriya, S., 709